

Molecular World Today and Tomorrow Recent Trends in Biological Sciences

Edited by Wajid Zaman

Printed Edition of the Special Issue Published in International Journal of Molecular Sciences



www.mdpi.com/journal/ijms

Molecular World Today and Tomorrow: Recent Trends in Biological Sciences

Molecular World Today and Tomorrow: Recent Trends in Biological Sciences

Editor

Wajid Zaman

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



Editor Wajid Zaman Department of Life Sciences Yeungnam University Gyeongsan Korea, South

Editorial Office MDPI St. Alban-Anlage 66 4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *International Journal of Molecular Sciences* (ISSN 1422-0067) (available at: www.mdpi.com/journal/ ijms/special_issues/Biological_Sciences).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* Year, *Volume Number*, Page Range.

ISBN 978-3-0365-6997-0 (Hbk) ISBN 978-3-0365-6996-3 (PDF)

© 2023 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

About the Editor
Alexander G. Bobylev, Elmira I. Yakupova, Liya G. Bobyleva, Nikolay V. Molochkov,Alexander A. Timchenko and Maria A. Timchenko et al.Nonspecific Amyloid Aggregation of Chicken Smooth-Muscle Titin: In Vitro InvestigationsReprinted from: Int. J. Mol. Sci. 2023, 24, 1056, doi:10.3390/ijms240210561
Ming Ma, Suya Liu, Zhiwei Wang, Ran Shao, Jianrong Ye and Wei Yan et al.Genome-Wide Identification of the SUN Gene Family in Melon (Cucumis melo) and FunctionalCharacterization of Two CmSUN Genes in Regulating Fruit Shape VariationReprinted from: Int. J. Mol. Sci. 2022, 23, 16047, doi:10.3390/ijms232416047
Junhu Tai, Jaehyung Park, Munsoo Han and Tae Hoon Kim Screening Key Genes and Biological Pathways in Nasopharyngeal Carcinoma by Integrated Bioinformatics Analysis Reprinted from: <i>Int. J. Mol. Sci.</i> 2022, 23, 15701, doi:10.3390/ijms232415701
Denis V. Petrovsky, Vladimir R. Rudnev, Kirill S. Nikolsky, Liudmila I. Kulikova, KristinaM. Malsagova and Arthur T. Kopylov et al.PSSNet—An Accurate Super-Secondary Structure for Protein SegmentationReprinted from: Int. J. Mol. Sci. 2022, 23, 14813, doi:10.3390/ijms23231481353
 Heewon Park and Satoru Miyano Computational Reprinted from: <i>Int. J. Mol. Sci.</i> 2022, 23, 14398, doi:10.3390/ijms232214398 65
 Shuai Liu, Weiming Gong, Lu Liu, Ran Yan, Shukang Wang and Zhongshang Yuan Integrative Analysis of Transcriptome-Wide Association Study and Gene-Based Association Analysis Identifies In Silico Candidate Genes Associated with Juvenile Idiopathic Arthritis Reprinted from: Int. J. Mol. Sci. 2022, 23, 13555, doi:10.3390/ijms232113555
 Intikhab Alam, Khadija Batool, Yuanyuan Huang, Junjie Liu and Liangfa Ge Developing Genetic Engineering Techniques for Control of Seed Size and Yield Reprinted from: Int. J. Mol. Sci. 2022, 23, 13256, doi:10.3390/ijms232113256
Maddalena Di Sanzo, Flora Cozzolino, Anna Martina Battaglia, Ilenia Aversa, Vittoria Monaco and Alessandro Sacco et al. Ferritin Heavy Chain Binds Peroxiredoxin 6 and Inhibits Cell Proliferation and Migration Reprinted from: <i>Int. J. Mol. Sci.</i> 2022 , 23, 12987, doi:10.3390/ijms232112987 129
Vladimir R. Rudnev, Kirill S. Nikolsky, Denis V. Petrovsky, Liudmila I. Kulikova, Anton M. Kargatov and Kristina A. Malsagova et al. 3-Corner Stability by Comparative Molecular Dynamics Simulations Reprinted from: <i>Int. J. Mol. Sci.</i> 2022 , 23, 11674, doi:10.3390/ijms231911674
Camilla Nero, Luca Boldrini, Jacopo Lenkowicz, Maria Teresa Giudice, Alessia Piermattei and Frediano Inzani et al. Deep-Learning to Predict BRCA Mutation and Survival from Digital H&E Slides of Epithelial Ovarian Cancer Reprinted from: <i>Int. J. Mol. Sci.</i> 2022 , <i>23</i> , 11326, doi:10.3390/ijms231911326
Tingting Zhu, Linxuan Li, Huimin Chang, Jiasui Zhan and Maozhi Ren

Maxim Kondratyev, Vladimir R. Rudnev, Kirill S. Nikolsky, Alexander A. Stepanov, Denis V. Petrovsky and Liudmila I. Kulikova et al.
Atomic Simulation of the Binding of JAK1 and JAK2 with the Selective Inhibitor Ruxolitinib Reprinted from: <i>Int. J. Mol. Sci.</i> 2022 , <i>23</i> , 10466, doi:10.3390/ijms231810466 193
Thanapon Charoenwongpaiboon, Amy Oo, Sutita Nasoontorn, Thanyada Rungrotmongkol, Somdej Kanokmedhakul and Panupong MahalapbutrAurisin A Complexed with 2,6-Di-O-methylcyclodextrin Enhances Aqueous Solubility, Thermal Stability, and Antiproliferative Activity against Lung Cancer Cells Reprinted from: Int. J. Mol. Sci. 2022, 23, 9776, doi:10.3390/ijms23179776205
Waqas Rahim, Murtaza Khan, Tiba Nazar Ibrahim Al Azzawi, Anjali Pande, Nusrat JahanMethela and Sajid Ali et al.Exogenously Applied Sodium Nitroprusside Mitigates Lead Toxicity in Rice by RegulatingAntioxidants and Metal Stress-Related TranscriptsReprinted from: Int. J. Mol. Sci. 2022, 23, 9729, doi:10.3390/ijms23179729
Dinesh Kumar Chellappan, Richie R. Bhandare, Afzal B. Shaik, Krishna Prasad, Nurfatihah Azlyna Ahmad Suhaimi and Wei Sheng Yap et al. Vaccine for Diabetes—Where Do We Stand? Reprinted from: <i>Int. J. Mol. Sci.</i> 2022 , <i>23</i> , 9470, doi:10.3390/ijms23169470
Zhao Liang, Qurban Ali, Yujie Wang, Guangyuan Mu, Xuefei Kan and Yajun Ren et al. Toxicity of <i>Bacillus thuringiensis</i> Strains Derived from the Novel Crystal Protein Cry31Aa with High Nematicidal Activity against Rice Parasitic Nematode <i>Aphelenchoides besseyi</i> Reprinted from: <i>Int. J. Mol. Sci.</i> 2022 , <i>23</i> , 8189, doi:10.3390/ijms23158189
Yanhui Zhou, Hakim Manghwar, Weiming Hu and Fen LiuDegradation Mechanism of Autophagy-Related Proteins and Research ProgressReprinted from: Int. J. Mol. Sci. 2022, 23, 7301, doi:10.3390/ijms23137301

About the Editor

Wajid Zaman

Dr. Wajid Zaman completed his Ph.D. at the Institute of Botany, Chinese Academy of Sciences, China. Currently, he is working as an International Research Professor at the Department of Life Sciences, Yeungnam University, Republic of Korea. Previously he worked as an Assistant Researcher/Professor at Lushan Botanical Garden, Chinese Academy of Sciences, China. He is interested in developing a career that combines teaching and research while maintaining his interest in Biological Sciences with advanced technologies.



Article Nonspecific Amyloid Aggregation of Chicken Smooth-Muscle Titin: In Vitro Investigations

Alexander G. Bobylev ^{1,*}, Elmira I. Yakupova ², Liya G. Bobyleva ¹, Nikolay V. Molochkov ¹, Alexander A. Timchenko ³, Maria A. Timchenko ⁴, Hiroshi Kihara ⁵, Alexey D. Nikulin ³, Azat G. Gabdulkhakov ³, Tatiana N. Melnik ³, Nikita V. Penkov ⁶, Michail Y. Lobanov ³, Alexey S. Kazakov ⁴, Miklós Kellermayer ⁷, Zsolt Mártonfalvi ⁷, Oxana V. Galzitskaya ^{1,3} and Ivan M. Vikhlyantsev ^{1,8,*}

- ¹ Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, 142290 Moscow Region, Russia
- ² A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119991 Moscow Region, Russia
- ³ Institute of Protein Research, Russian Academy of Sciences, Pushchino, 142290 Moscow Region, Russia
- ⁴ Institute for Biological Instrumentation, Federal Research Center, Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences, Pushchino, 142290 Moscow Region, Russia
- ⁵ Department of Early Childhood Education, Himeji-Hinomoto College, 890 Koro, Kodera-cho, Himeji 679-2151, Japan
- Institute of Cell Biophysics, FRC PSCBR, Russian Academy of Sciences, Pushchino, 142290 Moscow Region, Russia
- ⁷ Department of Biophysics and Radiation Biology, Semmelweis University, 1085 Budapest, Hungary
- ⁸ Institute of Fundamental Medicine and Biology, Kazan Federal University, 420008 Kazan, Russia
- * Correspondence: bobylev1982@gmail.com (A.G.B.); ivanvikhlyantsev@gmail.com (I.M.V.)

Abstract: A giant multidomain protein of striated and smooth vertebrate muscles, titin, consists of tandems of immunoglobulin (Ig)- and fibronectin type III (FnIII)-like domains representing β -sandwiches, as well as of disordered segments. Chicken smooth muscles express several titin isoforms of ~500–1500 kDa. Using various structural-analysis methods, we investigated in vitro nonspecific amyloid aggregation of the high-molecular-weight isoform of chicken smooth-muscle titin (SMT_{HMW}, ~1500 kDa). As confirmed by X-ray diffraction analysis, under near-physiological conditions, the protein formed amorphous amyloid aggregates with a quaternary cross- β structure within a relatively short time (~60 min). As shown by circular dichroism and Fourier-transform infrared spectroscopy, the quaternary cross- β structure—unlike other amyloidogenic proteins—formed without changes in the SMT_{HMW} secondary structure. SMT_{HMW} aggregates partially disaggregated upon increasing the ionic strength above the physiological level. Based on the data obtained, it is not the complete protein but its particular domains/segments that are likely involved in the formation of intermolecular interactions during SMT_{HMW} amyloid aggregation. The discovered properties of titin position this protein as an object of interest for studying amyloid aggregation in vitro and expanding our views of the fundamentals of amyloidogenesis.

Keywords: smooth muscle titin; protein aggregates; amyloid aggregation; amyloids; cross-β

1. Introduction

It is known that, to normally perform their biological functions, newly synthesized proteins must fold into a certain three-dimensional structure [1,2]. The folded structures of proteins are only moderately stable. Certain factors, such as genetic mutations or disturbances of protein synthesis and degradation, can lead to incorrect protein stacking or misfolding followed by the formation of pathological aggregates [3]. Improper protein stacking is a rather common phenomenon associated most often with the development of diseases such as amyloidoses [4]. With the development of amyloidosis in humans or animals, the protein loses its native conformation to form amyloids, aggregates of

Citation: Bobylev, A.G.; Yakupova, E.I.; Bobyleva, L.G.; Molochkov, N.V.; Timchenko, A.A.; Timchenko, M.A.; Kihara, H.; Nikulin, A.D.; Gabdulkhakov, A.G.; Melnik, T.N.; et al. Nonspecific Amyloid Aggregation of Chicken Smooth-Muscle Titin: In Vitro Investigations. *Int. J. Mol. Sci.* 2023, 24, 1056. https://doi.org/10.3390/ ijms24021056

Academic Editor: Wajid Zaman

Received: 28 November 2022 Revised: 29 December 2022 Accepted: 29 December 2022 Published: 5 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). incorrectly folded forms of protein having a specific structure. Amyloid aggregates, as well as their intermediate forms (mainly oligomers), may lead to cellular death [5–10].

Amyloid aggregates of many proteins have been found in various tissues and organs of amyloidosis humans and animals. These diseases include, in particular, liver amyloidosis, Alzheimer's disease, Parkinson's disease, type II diabetes, and prion diseases, as well as systemic amyloidoses [11–14]. Amyloids have a number of specific characteristics, such as the ability to bind to Congo Red and thioflavin T; they are resistant to proteases and insoluble in most solvents [14]. The main property of amyloids, regardless of the type of aggregates they form, is the presence of a quaternary cross- β structure [15,16].

Besides pathological amyloids, there are functional amyloids—protein aggregates that perform certain functions. Functional amyloids are found in many species of living organisms from plants, fungi, and protozoa to higher animals, including humans [17–27].

Despite the extensive investigation, there is as yet no clear understanding of the molecular mechanisms of amyloid aggregation. The differences behind the mechanisms of the formation of functional versus pathological amyloids have yet to be resolved. This is probably due to the complexity of the object of study. In order to obtain the required information about amyloids, it is necessary to use an array of methods (such as nuclear magnetic resonance spectroscopy, transmission electron microscopy (TEM), cryo-electron microscopy, atomic-force microscopy (AFM), X-ray diffraction, small-angle X-ray scattering (SAXS), Fourier-transform infrared spectroscopy (FTIR), which is often technically difficult to perform, since each of the methods has its own limitations regarding the object of study.

Titin (also called connectin [28,29]) is one of the most difficult objects when studying its aggregation properties. In vitro studies on isolated titin preparations revealed a tendency of this protein to form various sorts of intermolecular interactions that led to the formation of oligomers and aggregates. Thus, the aggregation of titin in vitro was first investigated in 1993 [30]. It was shown that, in solutions of low ionic strength (0.1 M KCl near neutral pH (6.5)), striated muscle titin assembles into higher-order aggregates [30]. In 2003, atomic-force microscopy revealed that, in a solution containing 25 mM imidazole–HCl, 0.2 M KCl, 4 mM MgCl₂, 1 mM EGTA, 0.01% NaN₃, 1 mM dithiothreitol (DTT), 20 μ g/mL leupeptin, 10 μ M E-64, pH 7.4, isolated titin molecules are capable of self-assembly into oligomers within a mere 10 min [31]. The observed titin oligomers could be divided into bimolecular species, as well as consisting of several molecules. In each case, however, a globular head was observed, with which other titin molecules interacted, thus forming an oligomer [31].

In vitro investigations of the specificity of aggregation between titin domains conducted in 2005 concluded that the ability for aggregation increases when the identity of amino acid sequences of the domains is greater than 40% [32]. In 2015, the in vitro aggregation of some titin Ig-domains obtained by a recombinant method was studied [33]. The neighboring identical titin domains were shown to be able to form misfolded structures [33].

The experiments in vitro with molecular simulations carried out in 2015 showed that during the refolding of tandem repeats (I27–I27 and I27–I28 immunoglobulin-like domains of titin), independent of sequence identity, more than half of all molecules transiently formed a wide range of misfolded conformations [33]. Simulations suggested that a large fraction of these misfolds resemble an intramolecular amyloid-like state named "intramolecular amyloids" [33]. These authors also reported that, during the refolding of tandem repeats, the development of a strand-swapped long-lived misfolded state without an amyloid-like structure is possible [33].

In 2016, the ability of a low-molecular-weight (500 kDa) isoform or possibly a truncated form of chicken smooth-muscle titin (SMT_{LMW}) to form in vitro amyloid aggregates was described [34]. In 2018, the ability of that isoform to form in vitro amyloid aggregates with different properties in two solutions differing in ionic strength was found [35].

In the present work, using a number of structural analysis methods, we investigated the formation of an in vitro amyloid structure of a chicken smooth-muscle titin isoform with a molecular weight of ~1500–1600 kDa (SMT_{HMW}). A peculiar feature of this protein and, in particular, this isoform of titin relative to other amyloid proteins studied is its

large size. One titin molecule is comparable to the size of the protofibrils of some amyloid proteins. Thus, it was of interest to elucidate whether this high-molecular-weight isoform of smooth-muscle titin was capable of forming aggregates, their structural features, and the type of aggregation—amyloid/non-amyloid. The results obtained point at the nonspecific amyloid aggregation of the protein.

2. Results

2.1. DLS Analysis of SMT Aggregations

Based on our previous research into SMT_{LMW} [34,35] and SMT_{HMW} [36], we were aware of the conditions required to form amyloid aggregates of these proteins. This work, in order to understand the features of chicken titin aggregation rate, investigated the kinetics of SMT_{HMW} aggregation. We used dynamic light scattering (DLS), the simplest and most convenient method to study the aggregation kinetics of amyloid proteins [34,35,37]. The method also enables analyzing the aggregation reversibility, which has been shown earlier for SMT_{LMW} [35].

Figure 1a shows a change in the autocorrelation function of the scattered light upon the formation of SMT_{HMW} aggregates in a solution containing 0.15 M glycine–KOH at pH 7.0–7.5 over 60 min. During the first 20 min of incubation, the correlation function $g_1(t)$ was observed to decay almost mono-exponentially (Figure 1a), with the subsequent emergence of a shoulder at high correlation times. This is indicative of the formation of large aggregates with smaller diffusion coefficients. After 40 min, the correlation function $g_1(t)$ featured a greater decay time and a more pronounced shoulder (Figure 1a) at high correlation times, which indicated an aggregation increase. A pronounced correlation-function shoulder was also observed during 40, 50, and 60 min of incubation (Figure 1a).

Several peaks reflecting the dimensions of SMT_{HMW} aggregates were obtained by the correlation function analysis. Prior to the formation of aggregates, only two well-resolved peaks with the average R_h of approximately 16 nm (the dominating peak of approximately 73%) and 86 nm (the minor peak of approximately 27%) were observed (Figure 1b, 0 min). The first peak corresponds, most likely, to SMT_{HMW} molecules, whereas the second can be indicative of their insignificant aggregation. During the first 20 min of incubation, we observed a shift of the peaks and an increase in the ratio of the volumes of the larger- and smaller-size fraction, indicating the development of the titin aggregation process (R_h ~53 nm and ~479 nm) (Figure 1b).

After 40 min of the experiment, a stable fraction of larger aggregates with a hydrodynamic radius of ~1175 nm (96.7%) appeared. This fraction was also observed after 50 and 60 min of the experiment (Figure 1b). Another peak, indicating the formation of titin aggregates with the average R_h of approximately 3813 nm emerged after 50 min of incubation and, after over 60 min, R_h was more than 4475 nm. It should be noted that the peak with R_h of approximately 4500 nm was at the limit of the range for this method. Therefore, the formation of larger SMT_{HMW} aggregates cannot be excluded (Figure 1b).

The next task was to elucidate the ability of SMT_{HMW} aggregates to disaggregate upon increasing ionic strength. Figure 1c presents data on the partial disaggregation of SMT_{HMW} aggregates after incubation in a solution containing 0.6 M KCl and 30 mM KH₂PO₄ (pH 7.0). A decrease in the percentage of particles with larger hydrodynamic radii and an increase in the number of particles with smaller hydrodynamic radii (Figure 1c) were observed.



Figure 1. (a) Evolution of field autocorrelation functions $g_1(t)$ of light scattered during SMT_{HMW} amyloid aggregate formation (pH 7.0–7.5, T = 25 °C). Aggregation was observed in a solution containing 0.15 M glycine–KOH, pH 7.0–7.5; (b) size distributions of SMT_{HMW} particles. Generation of large aggregates and their time-dependent growth are shown; (c) distribution of SMT_{HMW} particles after 1 h disaggregation (for clarity, the graph shows monodispersed SMT_{HMW}). The data reflect the distributions corresponding to certain time points. Three independent experiments were carried out.

2.2. Electron and Atomic-Force Microscopy of SMT_{HMW} Aggregates

Based on the DLS data presented above, investigations were conducted using two time intervals (1 h and 24 h). Figure 2 shows AFM images of SMT_{HMW} molecules in a solution of 0.6 M KCl, 30 mM KH₂PO₄, 1 mM DTT, 0.1 M NaN₃, and pH 7.0. It was found that, as in the case of striated-muscle titin [31,38,39], the smooth-muscle titin molecules had a filamentous shape with threads 3–4 nm thick and ~300 nm (on average) long with a globular head on one end (Figure 2a, inside the white squares). Besides individual molecules, oligomers of this protein were also found: several titin molecules interacting with one another, often forming a thickening in the center (Figure 2b). Similar oligomers have been found earlier in titin preparations isolated from rabbit skeletal muscles [31].

Figure 3 shows electron micrographs of negatively stained SMT_{HMW} aggregates formed in a solution containing 0.15 M glycine–KOH, pH 7.0–7.5, at 4 $^\circ$ C. SMT_{HMW} formed

amorphous aggregates after 1 h and 24 h incubation (Figure 3). Aggregates became larger after 24 h of incubation. In a number of cases, filamentous structures of about 4 nm in diameter (represented, most likely, by titin filaments) were observed between amorphous aggregates (Figure 3c).



Figure 2. AFM images of molecular SMT_{HMW} obtained in a solution of 0.6 M KCl, 30 mM KH₂PO₄, 1 mM DTT, 0.1 M NaN₃, and pH 7.0. In the field of view, there are filamentous protein molecules—monomers (marked with white squares) (**a**), which assemble into oligomers forming a central thickening (**b**).



Figure 3. Electron microscopy of SMT_{HMW} aggregates formed in a solution containing 0.15 M glycine–KOH at pH 7.0–7.5 at 4 °C. (a) 1 h aggregation of SMT_{HMW}. (b,c) 24 h aggregation of SMT_{HMW}. Protein filaments of a diameter of about 4 nm between amorphous aggregates (shown by red arrows on the insert) can be seen. The most representative data, obtained as a result of 10 independent experiments, are given. Scale bar, 100 nm.

According to AFM, after 1 h, aggregation SMT_{HMW} aggregates looked like large amorphous structures several micrometers long and up to 350 nm high (Figure 4a). After 24 h of aggregation, they had the form of branching flattened structures of more than 10 μ m in length and 200–250 nm in height (Figure 4b,c).



Figure 4. Atomic-force microscopy of SMT_{HMW} aggregates in a solution containing 0.15 M glycine and pH 7.0–7.5. (a) SMT_{HMW} after 1 h aggregation. (b,c) SMT_{HMW} after 24 h aggregation. (d,g) SMT_{HMW} aggregates after 1 h disaggregation; (e,f) SMT_{HMW} aggregates after 24 h disaggregation. Scale bar in (g), 1 μ m. The most representative data, obtained as a result of 3 independent experiments, are given.

To determine their capability of disaggregation, SMT_{HMW} aggregates formed during 1 and 24 h were transferred to conditions with increased ionic strength (0.6 M KCl, 30 mM KH₂PO₄, 1 mM DTT, 0.1 M NaN₃, and pH 7.0). Figure 4d presents the results of 1 h disaggregation. As can be seen in the figure, SMT_{HMW} became much smaller and spherical amorphous aggregates of about 100–200 nm in diameter and 100–120 nm in height. Apart from amorphous aggregates, individual filaments occurred on the substrate (Figure 4g); judging by their size, they may be bundles of SMT_{HMW} molecules.

Figure 4e,f shows the disaggregation results for SMT_{HMW} aggregates formed for 24 h. Spherical amorphous aggregates up to 23 nm in height can be seen, as well as a network of threads with a height of 5–6 nm.

2.3. Circular Dichroism

Figure 5a illustrates the circular dichroism (CD) spectrum of SMT_{HMW} before and after the formation of aggregates. No changes in the secondary structure were detected upon the formation of SMT_{HMW} aggregates: after chromatography the preparation had $6 \pm 6\%$ α -helices and $41 \pm 6\%$ β -structures, $53 \pm 6\%$ of turns and a disordered structure,

whereas aggregated SMT_{HMW} had a helix and β -structure content of $5 \pm 6\%$ and $42 \pm 6\%$, respectively, $53 \pm 6\%$ of turns and disordered structure. Thus, in both cases, we detected a high content of β -structure and a disordered secondary structure.



Figure 5. Investigation of the structure of SMT_{HMW} aggregates by various methods. (**a**) CD spectrum of SMT_{HMW}; (**b**) FTIR spectra of titin and its aggregates at 20 °C. Protein concentration was 10 mg/mL; (**c**,**d**) thioflavin T (ThT) staining of SMT_{HMW} aggregates formed in a solution containing 0.15 M glycine–KOH at pH 7.0–7.5. (**c**) 1 h formation of aggregates; (**d**) 24 h formation of aggregates; (**e**,**f**) X-ray diffraction of SMT_{HMW} aggregates formed after 1 h (**e**) and 24 h (f).

2.4. Fourier-Transform Infrared Spectroscopy

Figure 5b presents FTIR data obtained at 20 $^{\circ}$ C and corrected for the spectral contribution of water vapor and CO₂. The experimental data were analyzed following the principles described in [40]. The obtained estimates of the content of secondary structure elements in samples of titin and its aggregates are given in Table 1. As the circular dichroism data, the FTIR data indicate that the secondary structure of the protein does not change during SMT_{HMW} aggregation.

Table 1. Secondary structure content in chicken titin samples calculated by the FTIR method.

Preparation	α-Helices, %	β-Folds, %	Turns, %	Disordered Structure, %
Molecular SMT _{HMW}	26 ± 1.0	25 ± 2.0	11.7 ± 0.6	32 ± 2.0
Aggregated SMT _{HMW}	24.6 ± 1.6	26 ± 2.0	10.7 ± 0.4	34 ± 2.0

The results of two experiments are presented. The means and standard deviations are given.

2.5. Association of SMT_{HMW} Aggregates with Thioflavin T

To identify the amyloid nature of SMT_{HMW} aggregates, we investigated their binding to thioflavin T. A significant increase in ThT fluorescence intensity in the presence of SMT_{HMW} aggregates formed over 1 h (Figure 5c) and 24 h (Figure 5d) was recorded and

compared with that in the presence of monodispersed SMT_{HMW} . After disaggregation, the ThT fluorescence in the presence of SMT_{HMW} aggregates was at the same level as that of monodispersed SMT_{HMW} after 1 and 24 h of the experiment.

2.6. X-ray Diffraction of SMT_{HMW} Aggregates

The amyloid cross- β structure of SMT_{HMW} aggregates was revealed by X-ray diffraction (Figure 5e,f). A 1 h aggregation featured a diffuse reflex at ~10 Å and a relatively sharp reflex at ~4.7 Å (Figure 5e). X-ray diffraction of SMT_{HMW} aggregates after a 24 h aggregation revealed a diffuse reflex at ~10 Å and a sharp reflex at ~4.8 Å (Figure 5f). The detected reflections can be ascribed to a cross- β structure. Thus, the presence of a cross- β structure identified by X-ray diffraction analysis confirms that SMT_{HMW} aggregates are amyloids.

2.7. Small-Angle X-ray Scattering (SAXS)

SAXS was used to obtain information about the conformation of molecules in solution. For such high-molecular-weight structures as titin, this method makes it possible to approximately estimate their internal conformation, taking into account the tangent of the slope (tan A) of the log I-log S dependence. It is known that, for a rod-shaped conformation, tan A = 1; that, for a planar conformation, tan A = 2; that, for globular particles tan, A = 4. Using the DAMMIF program [41], it is also possible to visualize the approximate three-dimensional structure of the protein.

From these data (Figure 6), it follows that molecular SMT_{HMW} has a flat shape (plate conformation) (tan A = -2.7) (Figure 6a). The latter is clearly visible in the insert (Figure 6c). The aggregated shape of SMT_{HMW} particles is also close to plate-like (more charged plate conformation) (tan A = 2.65) (Figure 6b).

2.8. Differential Scanning Calorimetry of SMT_{HMW}

The thermal stability of SMT_{HMW} aggregates was elucidated by differential scanning calorimetry (DSC). Figure 7 presents typical temperature dependences of excess heat capacity of SMT_{HMW} in monomeric and aggregated forms. As seen in the figure, the curves have heat absorption peaks that might correspond to cooperative disruption of the structure. Repeated heating of the preparations confirms that the process is irreversible. The heat absorption peak maximum temperature of molecular titin, $T_{\rm m}$, was 317.7 ± 0.1 K; the value of transition calorimetric enthalpy, $\Delta H_{\rm cal} = 870 \pm 90$ kJ mol⁻¹. The heat absorption peak maximum temperature of titin aggregates, $T_{\rm m}$, was 321.7 ± 0.1 K; the value of transition calorimetric enthalpy, $\Delta H_{\rm cal} = 1090 \pm 110$ kJ mol⁻¹.

2.9. SMT Amino Acid Sequence Identity

To reveal segments with large amino acid sequence identity that, according to available data, have an increased tendency for aggregation [33], we calculated the identity in the amino acid sequence between adjacent pairs of domains in smooth-muscle titin. Chicken titin (UniProtKB—A6BM71_CHICK) was chosen for the calculations. Calculations carried out using the BLAST program showed that the average identity in the amino acid sequence between neighboring FnIII domains does not exceed $33 \pm 7\%$, and between neighboring Ig domains, it does not exceed $20 \pm 11\%$, which is a relatively low indicator (Table 2, Supplementary File S1). In two cases, however, domains with an identity higher than 50% were observed.



Figure 6. Investigation of the structure of SMT_{HMW} aggregates by small-angle X-ray scattering. (a) The scattering curve for molecular SMT_{HMW} in logarithmic representation log *I*-log *Q*, where *I* is the scattering intensity in 0.6 M KCl, c = 0.5 mg/mL (non-aggregated protein form), *Q* is the scattering vector, $\tan A = -2.7$; (b) SMT_{HMW} in 0.15 M Gly (aggregated); *Q*, scattering vector; $\tan A = -2.65$; (c) SMT_{HMW} in 0.6 M KCl, c = 0.5 mg/mL; dark line, approximation by the DAMMIF program.



Figure 7. Typical temperature dependence of excess heat capacity of SMT_{HMW} in monomeric (black) and aggregated (blue) forms. Experiments were carried out 2 times for both the monomeric and aggregated SMT_{HMW} forms. The obtained curves coincided in temperature T_m to an accuracy of 0.1 K. The relative error in determining calorimetric enthalpy, Δ Hcal, did not exceed 10%.

1056
24,
2023,
Sci.
Mol.
÷
Int.

r N ₀ /N·100, %	16%	58%	75%	0%0	65%	54%	Chicken titin domaii <i>i</i> th the signature nor tion between the Z-lir
N, Total Numbe of Domain Pairs	4032	384	704	30	99	110	tues in the domain. a pseudodomain w tension and contract the literature.
N ₀ , Number of Domain Pairs with Zero Identity	647	223	526	0	43	59	nt. L—number of resic veen the domains, then , which is involved in e? al titin is not available ii
Side Standard Dis- persion/Deviation of Identity, %	11%	4%	6%	7%	2%	5%	residues in the alignme pre than 50 residues beth celetal muscle sarcomere full-sized chicken skelet
(Id) Mean Identity, %	20%	3%	2%	33%	2%	3%	-number of identical was a distance of mc -band region of the sl ele isoform of titin or
Domain 2 (for Pair Comparison)	Ig-like	FnIII-like	none	FnIII-like	none	none	$\langle N_{id}/(L_1 + L_2)$. $N_{id}-$ M71_CHICK. If there is corresponded to the 1 e for the smooth musc
Average Length of Domain	81.4	81.4	81.4	84.3	84.3	104.2	the formula: Id = 2 > tite: UniProtKB—A6B1 oe-III was created. This The complete sequence
Number of Domains in the Group	64	64	64	9	9	11	titty was calculated by e taken from the webs I-like = Fibronectin tyr the A-I junction [42].
Domain 1 (for Pair Comparison)	Ig-like	lg-like	Ig-like	FnIII-like	FnIII-like	none	Ider werr FnII
Protein			5	Chicken titin			

Table 2. Identity of the chicken-titin amino acid sequence.

2.10. Calculation of Unstructured Areas in the SMT Molecule

The disorder was revealed by analyzing the amino acid sequence UniProtKB—A6BM71 _CHICK using IsUnstruct, a specialized program for predicting the natural disorder of proteins [43,44]. The average disorder of a titin fragment was 90% for PRK12323; 90%, Ig-like; 30%, FnIII-like; 55%, none (Figure 8).



Figure 8. Titin-chicken domain structure schematic aligned with the protein disorder prediction plots by the IsUnstruct program [44]. Probabilities of ≥ 0.5 mean a disorder.

3. Discussion

In the present work, we investigated the amyloidogenic propensity in the smoothmuscle isoform of the giant protein titin. According to our results, three titin isoforms—with molecular weights of ~500, ~1200, and ~1500–1600 kDa—were isolated from chicken gizzard muscle tissue. Previously, we have shown the 500 kDa splice-isoform of titin or its truncated fragment (SMT_{LMW}) to form in vitro aggregates with amyloid properties and structure [34,35]. In this study, we show that the smooth-muscle titin isoform with MW~1500–1600 kDa and an isoform or a proteolytic fragment of this protein with MW~1200 kDa (SMT_{HMW}) also form aggregates in vitro. Detailed research into the SMT_{HMW} aggregation process was conducted to better understand the changes occurring in this protein, which is undoubtedly involved in smooth muscle contraction.

Using DLS, we found that, upon decreasing ionic strength, chicken SMT_{HMW} formed large aggregates with a hydrodynamic radius of ~4500 nm over 1 h (Figure 1a). EM and AFM showed SMT_{HMW} aggregates formed over 1 h to be amorphous (Figures 3 and 4). In general, chicken SMT_{HMW} aggregation was fast, which made it impossible to determine the lag period. The high aggregation rate of SMT_{HMW} (three times as high as that for SMT_{LMW} [34]) was, apparently, due to the oligomeric forms of the protein and its monomers present, as shown by AFM, in the high ionic-strength solution (Figure 2). Our results are consistent with the literature data showing that the presence of oligomeric forms at the initial stage of aggregation accelerates the process [45].

An increase of ThT fluorescence detected at the binding of the dye to SMT_{HMW} aggregates an hour after their formation indicates the amyloid nature of the structures formed (Figure 5c,d). This is also supported by the X-ray diffraction data: the presence of reflexes at 4.7 and 10 Å confirms the amyloid nature of SMT_{HMW} aggregates formed both after 1 and 24 h of incubation (Figure 5e,f). The X-ray diffraction data are indicative of the presence of a quaternary cross- β structure in SMT_{HMW} aggregates. It should be noted that both reflexes are circular and partially blurred. Nevertheless, the reflex representing the distance between the beta segments in the sheet was observed both after 1 h (4.7 Å).

Figure 5e) and after 24 h (4.8 Å, Figure 5f) of aggregation. The 10 Å reflex representing the distance between the beta sheets was blurred but was also present in both cases. According to numerous literature data, such reflexes are characteristic of amyloid or amyloid-like structures [46].

An uncharacteristic feature of the amyloid aggregation of titin is the formation of aggregates without changing the secondary structure, which was confirmed by CD and FTIR independent techniques (Figure 5a,b, Table 1). According to the literature data, in vitro experiments with some proteins have shown that prior to the formation of amyloids, the structure of their molecules must undergo a transformation of the type of α -helix to β -folding or random coil to β -sheet [47–49]. We revealed no such changes in the amyloid aggregation of chicken SMT_{HMW}. Similar results have been obtained earlier for chicken SMT_{LMW} [34,35] and for skeletal myosin binding protein C [46] that also consists of Ig-like and FnIII-like domains. It appears that the ability to form amyloid aggregates without changes in the secondary structure of molecules is a characteristic feature of the above-mentioned multidomain proteins, which distinguishes them from most other amyloid proteins.

Taking into account the EM data on protein filaments between amorphous SMT_{HMW} aggregates (Figure 3c) and the data on the presence of protein "filaments" after disaggregation of SMT_{HMW} aggregates (Figure 4e–g), we suggested that not the entire protein but only its particular domains were involved in intermolecular interactions in the process of amyloid SMT_{HMW} aggregation. The proposed SMT_{HMW} scheme of stacking to form an amyloid structure during the aggregation is given in Figure 9. In particular, segments that have a disordered structure and those whose structure is amyloid are shown.

Are there any confirmations of the assumption we make? In the literature, there are comparative data obtained by cryoelectron microscopy on the structure of functional amyloid aggregates of Orb2 and pathological aggregates of amyloid β -peptide [50]. Those authors have shown that only a small part of its molecule is involved in the formation of the amyloid nucleus in the functional amyloids of Orb2, whereas most of the molecule remains dynamically disordered. Formation of the amyloid nucleus in pathological amyloid β -peptide, on the contrary, involves most of its molecule. Given these data and the revealed properties of SMT_{HMW} aggregates, they can be classified as functional, which is indirectly confirmed by their ability to disaggregate with ionic strength increasing (Figures 1 and 4). It should only be noted that this type of aggregation is a model, since neither functional nor pathological SMT_{HMW} aggregates have been found in living cells.

It should also be understood that, due to the huge size and complex structure of the titin molecule, its complete transition to the amyloid form is hardly possible. However, its particular segments are, most likely, capable of forming an amyloid structure. This assumption can be supported by literature data on the unfolding of individual titin domains [51]. In particular, those authors have shown that the stepwise unfolding/folding of titin immunoglobulin (Ig) domains occurs in the elastic I band region of intact myofibrils at physiological sarcomere lengths and forces of 6–8 pN [52]. For this reason, we consider the formation of amyloid segments to occur exactly between partially opened domains of neighboring titin molecules.

It has also been shown that most proteins have amyloidogenic segments [53,54]. We calculated the number of amyloidogenic segments in randomly selected domains of a chicken titin fragment, predicted using the FoldAmyloid program (three Ig-like and three FnIII-like). From the data obtained (Supplementary File S2), there are at least two amyloidogenic segments in each of the counted domains; furthermore, for example, the Ig-like (I10) domain has five such segments. These data indicate a high potential propensity of titin, inherent in its domains, to form amyloid aggregates.

When discussing the type of SMT_{HMW} aggregates, it is necessary to dwell in more detail on the DSC data, which are indicative of yet another feature of titin's amyloid aggregation. It is known that amyloid fibrils or aggregates of several proteins melt and dissociate at temperatures of the order of 75–100 °C, which manifests itself in the form of characteristic endothermic transitions on the thermograms [55–59]. It has been shown that

amyloid fibrils are more resistant to temperature than the native protein (their transition temperatures differ by about 30–40 K) [59]. At the same time, it has been shown that the enthalpy of the melting of amyloid fibrils is less than that of the native protein. Based on this, it has been concluded that the density of intermolecular interactions in the amyloid structure is lower than in the native protein [59]. Our experiments showed that the melting point of the aggregated form of SMT_{HMW} is slightly higher (by only 4 K) than that of the molecular form of the protein (Figure 7). Herewith, the melting enthalpies of the aggregated and non-aggregated forms of SMT_{HMW} practically did not differ. These data suggest that the total binding energy during the formation of SMT_{HMW} aggregates did not change. Attention should be also paid to the heat absorption peaks themselves. If we compare the data obtained by us for SMT with the data for globular proteins, then the heat absorption peak should be much larger. There is a possibility that the current peak of heat absorption can correspond to the energy spent on the breakdown of the aggregates or oligomers of the molecular and aggregated forms of the protein or even its particular domains. If this is the case, then we failed to register the melting point, and it is above 373 K. Nevertheless, the obtained data are of interest. It can be summed up that, in this temperature range (up to 373 K), the molecular form of the protein and its aggregates have relatively similar stability.



Figure 9. A schematic of the internal structure of aggregates with a cross- β -sheet structure, formed by smooth-muscle titin. (a) The proposed structure of SMT aggregates based on circular dichroism, FTIR and X-ray diffraction data, suggesting the presence of a large amount of disorder. (b) A beta sheet and the distance of 4.7 Å between the beta strands. (c) The beta sheets and the distance of 10 Å between them are shown separately.

Discussing the issue of the type of (functional or pathological) amyloid aggregation, it is worth noting that no cases of the transformation of functional to pathological amyloids have been recorded. Most probably, there are molecular mechanisms of protection against such a transformation that developed in the process of evolution. Thus, it is known that, in multidomain proteins with homologous tandem repeats, the neighboring domains have a low identity in the amino acid sequence. This feature formed, apparently, as a result of evolutionary pressure, prevents incorrect protein stacking and subsequent aggregation [33]. It is generally accepted that a low tendency for aggregation is characteristic of those proteins in which the identity between domains is less than 40% [33]. Data on the identity of individual titin domains in the amino acid sequence are known [33]. We calculated the identity of amino acid sequences for chicken titin (Table 2). Calculations showed this fragment to contain only two domains with an identity greater than 50% (Supplementary File S1). The average identity of the amino acid sequence between pairs of neighboring FnIII-like or Ig-like domains does not exceed 40% (Table 2). Thus, it can be concluded that chicken SMT

has a low tendency towards aggregation. It has been shown, however, that upon sarcomere elongation, titin domains are capable of unfolding in situ [52], herewith opening the hidden hydrophobic sites that may lead to its aggregation.

In conclusion, regarding the possible role of the revealed changes for the muscle in vivo, it is necessary to recall one of the early works in which the authors show, using AFM, that repeated mechanical cycles of the extension/relaxation of multidomain proteins can lead to the formation of misfolded structures formed by two neighboring domains [60]. The misfolding was completely reversible and changed the mechanical topology of the domains, while maintaining the same stability as in the original folded state. The authors conclude that multidomain proteins can assume a new state of incorrect stacking. These data and the data we obtained can be important from the viewpoint of a better understanding of the functioning of multidomain muscle proteins, in particular titin, in the sarcomere and muscle as a whole. It is not to be ruled out that structural changes occurring in this protein during the muscle extension/contraction cycle are involved not only in the fine-tuning of elastic properties but also in changing the contractile response of the muscle. Future in vivo studies of titin structural modifications and misfolding will reveal more subtle nuances of the involvement of this protein in the functioning of muscle cells.

4. Conclusions

In summary, in vitro nonspecific amyloid aggregation of chicken SMT_{HMW} was found. Over relatively short times (within 1 h), the protein formed amorphous amyloid aggregates with a quaternary cross- β structure without undergoing changes in the secondary structure. The thermal stability of SMT_{HMW} aggregates did not practically differ from that observed in protein preparations containing both monomers and oligomers. Amyloid aggregates of SMT_{HMW} disaggregated almost completely at an increase in the ionic strength of the solution. Amyloid aggregates of this type are functional rather than pathological, so a similar aggregation of titin in vivo to perform certain functions cannot be ruled out. However, the involvement of this protein as a model object to study the process of the amyloid aggregation of a non-pathological type. Our data expand the views about the fundamentals of amyloidogenesis. Besides, disclosing the mechanisms of the formation of functional and pathological amyloids and understanding their differences at the structural level can give an idea of how the cell regulates amyloid aggregation for its desired functioning, avoiding the manifestation of the toxicity of pathological amyloids.

5. Materials and Methods

5.1. Purification of Chicken Gizzard SMT

Smooth-muscle titin was prepared from chicken gizzard by the method described in [61] with our modifications described in [34]. SMT_{HMW} was purified by gel filtration on a Sepharose-CL2B column equilibrated in a buffer containing 0.6 M KCl, 30 mM KH₂PO₄, 1 mM DTT, 0.1 M NaN₃, pH 7.0. Protein concentration was determined by a SPECORD UV VIS spectrophotometer using the extinction coefficient ($E_{280}^{1 \text{ mg/mL}}$) of 1.37 for titin [62]. For this research, the protein has been isolated more than 20 times.

5.2. SDS-PAGE and Mass Spectrometry Analysis of Titin

The presence of SMT_{HMW} in the sample was confirmed by sodium dodecyl sulfate– polyacrylamide gel electrophoresis (SDS-PAGE) (Supplementary File S3) and mass spectrometry analysis, the data of which are described in [36]. The molecular weight of SMT_{HMW} was assessed by TotalLab software v1.11 (Supplementary File S3). Two protein bands are visible, which are most likely SMT_{HMW} isoforms, or the lower band is a proteolytic fragment of this protein (Supplementary File S3, gels 1–4). By the densitometry data, the molecular weight of the upper band of the protein is ~1635 ± 245 kDa; its content is ~68.5%. The molecular weight of the lower band is ~1245 ± 189 kDa; its content is ~31.5%. The SDS-PAGE of titin was performed using a separating gel containing 6.5–7% polyacrylamide prepared as described [62]. The gels were stained with Coomassie Brilliant Blue G-250 and R-250 mixed at a 1:1 ratio. For shotgun mass spectrometry analysis, the sample was solubilized into a buffer (4% sodium dodecyl sulfate in 0.1 M Tris-HCl pH 7.6, 0.1 M dithio-threitol) and incubated for 5 min at 95 °C as described [36,63]. The samples were sonicated (4 × 30 s at 20 W; ME220, Covaris, Woburn, MA, USA), centrifuged (5 min, 16,000 × *g*), and the supernatant was collected. The YM-30 filter (Millipore, Ireland) was used for alkylation and trypsinolysis (14 h, 2 µg of trypsin (Trypsin Gold, Promega, Madison, WI, USA)) according to the FASP method [64]. Peptides were desalted using C18 microcolumns and subjected to HPLC–MS/MS analysis using the HPLC Ultimate 3000 RSLCnano system (Thermo Scientific, Waltham, MA, USA) as described [36,63].

5.3. Conditions for the Formation of SMT Aggregates

Purified SMT_{HMW} in a column buffer (0.6 M KCl, 30 mM KH₂PO₄, 1 mM DTT, 0.1 M NaN₃, pH 7.0) was used to form aggregates. SMT aggregates (concentration, 0.2–0.4 mg/mL) were formed by dialysis in Sigma-Aldrich cellulose membrane tubing (size, 25×16 mm) for 1 and 24 h at 4 °C against a solution containing 0.15 M glycine–KOH, pH 7.0–7.5. In disaggregation experiments, SMT aggregates were dialyzed during 1 and 24 h against a column buffer.

5.4. Dynamic Light Scattering Experiments

DLS experiments were conducted according to a protocol described in [34,35]. For the DLS analysis of SMT aggregation, a protein sample in a buffer containing 0.6 M KCl, 30 mM KH₂PO₄, 1 mM DTT, 0.1 M NaNO₃, pH 7.0, at an initial concentration of 1 mg/mL, was transferred into a solution of 0.15 M glycine–KOH, pH 7.0–7.5, by gradual dilution to a final concentration of 0.1 mg/mL to decrease ionic strength. Further steps were as in [35]. The collected autocorrelation functions were converted into particle-size distributions, using the general-purpose algorithm provided with the ZS Zetasizer Nano (Malvern Instruments Ltd., Malvern, UK) used in this experiment. Particle-size distributions obtained from alternative inversion algorithms yielded comparable results. The dynamic viscosity of the protein solutions determined using an SV-10 Sine-wave Vibro Viscometer (A&D Company Ltd., Tokyo, Japan) at 25 °C was 0.92 cP. This value was taken into consideration when measuring the particle dimensions in SMT samples collected 60 min after the dialysis. The analyzed volume of scattering with a beam cross-section of $\sim 100 \,\mu m$, accounting for the protein concentration used, contained about 10 billion SMT protein molecules, the signal from which was measured. The correlation function signal accumulated over 15 cycles of 15 s each (the way it is described in [34,35]). The results were obtained from three independent experiments.

5.5. Transmission Electron Microscopy

A drop of aggregated protein suspension at a concentration of 0.1 mg/mL was applied to a carbon-coated collodion film (2% collodion solution in amyl acetate (Sigma-Aldrich, St. Louis, MO, USA)) on a copper grid (Sigma-Aldrich, St. Louis, MO, USA) and negatively stained with 2% aqueous uranyl acetate (SPI-Chem., West Chester, PA, USA). Samples were examined under a JEM-100B electron microscope (JEOL Ltd., Tokyo, Japan). Samples obtained from five independent protein isolations were analyzed; many different fields of view were analyzed.

5.6. Atomic-Force Microscopy

For AFM measurements, titin aggregates were attached to freshly cleaved mica. An aliquot of titin (10–20 μ L) was pipetted onto the mica surface and incubated at room temperature for 10 min. Unbound protein was washed away by extensive rinsing with distilled water, then by blowing gently with a stream of high-purity N₂ gas. The noncontact mode (alternating current or AC mode) AFM images of titin aggregates bound to the mica surface were acquired with a Cypher ES AFM instrument (Asylum Research, Santa Barbara,

CA, USA). Scanning was performed at high set-point values (0.8–1.2 V) to avoid the binding of the sample to the cantilever tip. Silicon nitride cantilevers (Olympus) were used for scanning in air (AC160TS, resonance frequency~300 kHz). At a typical scanning frequency of 0.7–1.4 Hz, we collected 512 × 512 pixel or 1024 × 1024 pixel height-, amplitude-, and phase-contrast images.

To prepare samples of SMT_{HMW} aggregates for AFM, 2 μ L of the protein was transferred to freshly cleaved mica and incubated for 5 min. The sample was then washed three times in a drop of distilled water deionized by a type I Milli-Q system for 30 s and dried in the air. AFM imaging was performed using an AFM Ntegra-Vita microscope (NT-MDT, Russia) in noncontact (tapping) mode in air. The typical scan rate was 0.5–1 Hz. Measurements were carried out using NSG03 cantilevers with a resonance frequency of 47–150 kHz and ensured a 10 nm tip curvature radius. The processing and presentations of AFM images were performed using Nova software 1.0.26 (NT-MDT, Moscow Region, Russia) and Gwyddion 2.4450 software (http://gwyddion.net/download-old.php accessed on 17 April 2018. Experiments were replicated by independently executing the process of protein preparation, their incubation at 37 °C, and sample analysis three times. AFM images of a buffer solution (0.15 M glycine–KOH, pH 7.0–7.5) containing no titin are presented in Supplementary File S4. Samples obtained from three independent protein isolations were analyzed; many different fields of view were analyzed.

5.7. Circular Dichroism

SMT_{HMW} was dialyzed for 24 h against a buffer containing 0.15 M glycine–KOH, pH 7.0–7.5. The CD spectra prior to and after SMT_{HMW} aggregation were recorded in a Jasco J-815 spectrometer (JASCO Inc., Tokyo, Japan) using 0.1 cm optical path-quartz cells and wavelengths of 250–190 nm.

Three repeats of the spectrum were taken for each investigated sample. Data processing and graphical representation were performed in the SigmaPlot program. Based on the absorption spectrum, the exact protein concentration was calculated using the formula: $C = ABS_{280}/l/K_e$ (where ABS_{280} is the absorption value at a wavelength of 280 nm, *l* is the optical path length in cm, K_e is the extinction coefficient).

Three spectra in the far UV region obtained for the investigated sample were averaged and smoothed in the spectropolarimeter software (Spectra Manager Version 2, Spectra Analysis Version 2.02.06 (Build 1) Spectra Analysis Jasco). A similar procedure was done for the three spectra obtained for the buffer solution. The averaged spectrum of the buffer solution was subtracted from the obtained averaged spectrum of the investigated sample. The value of molar ellipticity $[\Theta]$ was calculated by the formula:

$$[\Theta]_{\lambda} = \Theta_{\lambda} \cdot \text{RMW}/l \cdot c,$$

where Θ_{λ} is the measured value of ellipticity at the wavelength λ , millidegrees; RMW, the average molecular weight of the residue, calculated from the amino acid sequence; *l*, the optical path length, mm; *c*, protein concentration, mg/mL.

The secondary structure was calculated using the CONTIN/LL module of the CDPro program [65]. The mean root square deviation (RMSD) according to the CD_{PRO} program did not exceed 6%.

5.8. Fourier-Transform Infrared Spectroscopy

Measurements were carried out on a Thermo Scientific Nicolet 6700 FT-IR spectrometer, equipped with the Smart Proteus accessory with a Peltier cuvette holder, in transmission mode in a cuvette of crystalline calcium fluoride with an optical pathlength of 4 μ m, using a liquid nitrogen-cooled MCT detector. Scanning in the wavenumber range from 650 to 4000 cm⁻¹; resolution, 1 cm⁻¹; averaging over 256 spectra. The device was calibrated according to the manufacturer's instructions.

The IR spectra of titin preparations' solutions in a corresponding buffer and the spectra of the buffer itself were measured at 20 $^{\circ}$ C. The concentration of the protein was 10 mg/mL.

The optical path length of the CaF₂ cuvette was calculated for each measurement based on the optical density of the test sample at 3404 cm⁻¹, using the water absorption value at an optical pathlength of 1 μ m equal to 0.533 AU, adjusted for the protein concentration in the sample [66]. The optical pathlength of the cuvette was 4.52 \pm 0.04 μ m. The IR spectrum of the protein preparation was measured twice; the buffer spectrum was also registered twice. The IR spectrum of the buffer (0.6 M KCl, 30 mM KH₂PO₄, 1 mM DTT, 0.1 M NaNO₃, pH 7.0, for molecular SMT; and 0.15 M glycine–KOH, pH 7.0–7.5, for the aggregated form of the protein) was subtracted from each protein spectrum, taking into account the difference in the values of the optical path length in the measurements. Each difference spectrum was adjusted for the spectral contribution of water vapor and CO₂, followed by the analysis in the wavenumber range of 1725 to 1481 cm⁻¹ for the content of secondary structure elements in the protein, following the principles described in [40]. A sample obtained from three different isolations of the protein was used. The obtained estimates of secondary structure elements in the protein were averaged by the results of two measurements. The standard deviations of the values of the secondary structure elements in the protein are given.

5.9. Fluorescence Analysis with Thioflavin T

The amyloid nature of the SMT_{HMW} aggregates was estimated by the intensity of thioflavin T (ThT) fluorescence (1 ThT:5 SMT (w/w)). Fluorescence was measured at $\lambda_{ex} = 440$ nm and $\lambda_{em} = 488$ nm using a Cary Eclipse spectrophotometer (Varian, Palo Alto, CA, USA). Four independent series of measurements were carried out. The amyloid nature was assessed from the difference of fluorescence intensity between the non-aggregated and aggregated forms of the protein. This method was used as a control of the amyloid nature of titin aggregates after each isolation of the protein.

5.10. X-ray Diffraction

SMT_{HMW} aggregates for X-ray diffraction analysis were prepared after a 1 h and 24 h incubation at 4 °C in an experimental solution. A sample obtained from three different isolations of the protein was used. Then the aggregates were concentrated up to more than 10 mg/mL by centrifugation at 12,000 rpm for 60 min. Droplets of this preparation were placed between the ends of wax-coated glass capillaries (approximately 1 mm in diameter) separated by a gap of approximately 1.5 mm. Fiber diffraction images were collected using a Microstar X-ray generator with HELIOX optics equipped with a Platinum135 CCD detector (X8 Proteum system, Bruker AXS) at the Institute of Protein Research, Russian Academy of Sciences, Pushchino. Cu K α radiation, $\lambda = 1.54$ Å (1 Å = 0.1 nm), was used. The samples were positioned at a right angle to the X-ray beam using a four-axis kappa goniometer. Different exposures and different oscillation angles were used; the sample itself was irradiated in different orientations.

5.11. Small Angle X-ray Scattering

SAXS experiments were carried out on a small-angle camera of the Photon Factory (Tsukuba, Japan). The protein solution in a thermostatted cuvette with mica windows was irradiated by X-rays of a wavelength 1.503 Å at 23 °C. The distance between the sample and the detector was 2.35 m. The range of the measured scattering vectors is Q = 0.008-0.2 Å⁻¹ ($Q = 4\pi \sin \theta/\lambda$, where λ is the X-ray radiation wavelength and 2 θ is the scattering angle). X-ray scattering was detected by a PILATUS 100K two-dimensional X-ray detector. The shape of the particles was estimated from the tangent of the angle of inclination log *I* on log *Q*, where *I* is the scattering intensity, and *Q* is the scattering vector modulus [67]. Log *I*-log *Q* dependences in SAXS data were approximated by linear regression. Correlation coefficients (R^2) ranged from 0.92 to 0.99.

5.12. Differential Scanning Calorimetry

DSC measurements were made on a SCAL-1 precision scanning microcalorimeter (Scal Co. Ltd., Pushchino, Russia) with 0.33 mL glass cells at a scanning rate of 1 K/min and

under a pressure of 2.5 atm [68]. The experiments were performed in 0.15 M glycine–KOH at pH 7.0–7.5. The protein concentrations were 1.2 mg/mL. The experimental calorimetric traces were adjusted for the calorimetric baseline, and the molar partial heat capacity functions were calculated in a standard manner. The excess heat capacity was evaluated by the subtraction of the linearly extrapolated initial and final heat capacity functions with correction for the difference of these functions by using a sigmoidal baseline [69]. DSC experiments were carried out 2 times for both the monomeric and aggregated SMT_{HMW} forms. The obtained curves coincided in temperature $T_{\rm m}$ to an accuracy of 0.1 K. The relative error in determining calorimetric enthalpy, Δ Hcal did not exceed 10%.

5.13. Calculation of the Identity of the Amino Acid Sequence and Disordered Regions in the SMT_{HMW} Molecule

The SMT_{HMW} amino-acid sequence identity was calculated by the BLAST program. The data were retrieved from the UniProtKB databases: UniProtKB—A6BM71_CHICK.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms24021056/s1.

Author Contributions: A.G.B. conceived and designed the experiments; A.G.B., E.I.Y., L.G.B., N.V.M., A.A.T., M.A.T., H.K., A.D.N., A.G.G., T.N.M., N.V.P., M.Y.L., A.S.K., Z.M., O.V.G. and I.M.V. performed the experiments; A.G.B., E.I.Y., L.G.B., N.V.M., A.A.T., M.A.T., T.N.M., N.V.P., A.S.K., M.K., O.V.G. and I.M.V. analyzed the data; A.G.B., M.K. and I.M.V. wrote the manuscript; A.G.B., E.I.Y., N.V.M., A.A.T., M.A.T., T.N.M. and A.S.K. prepared the figures; A.G.B., M.K., O.V.G. and I.M.V. reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Russian Science Foundation, grant number 22-24-00805, by the Hungarian National Research, Development and Innovation Office (K135360 to M.K.) and by grants from the Hungarian National Research, Development and Innovation Office: FK128956 and 2019-2.1.7-ERA-NET-2020-00013 to Z.M.

Institutional Review Board Statement: The manuscript does not contain clinical studies or patient data.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed in the course of this research (including files of additional information) were incorporated into the article and supplementary files.

Acknowledgments: The research used the equipment of the ITEB RAS Shared Facilities Centre "Structural and Functional Studies of Biosystems" (https://www.ckp-rf.ru/catalog/ckp/3037/ accessed on 29 December 2022), Electron Microscopy Core Facilities of the Pushchino Biological Research Centre (https://www.ckp-rf.ru/catalog/ckp/670266/ accessed on 29 December 2022).

Conflicts of Interest: The authors declare that there is no conflict of interest regarding this work.

Abbreviations

SMT_{HMW}, high-molecular-weight isoform of smooth-muscle titin (~1500 kDa); SMT_{LMW}, lowmolecular-weight isoform of smooth-muscle titin (~500 kDa); Ig, titin immunoglobulin-like domain; FnIII, titin fibronectin-like type III domains; CD, circular dichroism; FTIR, Fourier-transform infrared spectroscopy; DLS, dynamic light scattering; DTT, dithiothreitol; SDS–PAGE, sodium dodecyl sulfate– polyacrylamide gel electrophoresis; ThT, thioflavin T; TEM, transmission electron microscopy, AFM, atomic-force microscopy; SAXS, small-angle X-ray scattering; DSC, differential scanning calorimetry.

References

- 1. Tsytlonok, M.; Craig, P.O.; Sivertsson, E.; Serquera, D.; Perrett, S.; Best, R.B.; Wolynes, P.G.; Itzhaki, L.S. Complex energy landscape of a giant repeat protein. *Structure* **2013**, *21*, 1954–1965. [CrossRef] [PubMed]
- 2. Tian, P.; Best, R.B. Best structural determinants of misfolding in multidomain proteins. *PLoS Comput. Biol.* **2016**, *12*, e1004933. [CrossRef]
- Gershenson, A.; Gierasch, L.M.; Pastore, A.; Radford, S.E. Energy landscapes of functional proteins are inherently risky. *Nat. Chem. Biol.* 2014, 10, 884–891. [CrossRef] [PubMed]

- 4. Dobson, C.M. Protein folding and misfolding. Nature 2003, 426, 884–890. [CrossRef]
- 5. Stefani, M.; Rigacci, S. Protein folding and aggregation into amyloid: The interference by natural phenolic compounds. *Int. J. Mol. Sci.* **2013**, *14*, 12411–12457. [CrossRef]
- Ittner, L.M.; Ke, Y.D.; Delerue, F.; Bi, M.; Gladbach, A.; van Eersel, J.; Wölfing, H.; Chieng, B.C.; Christie, M.J.; Napier, I.A.; et al. Dendritic function of tau mediates amyloid-beta toxicity in Alzheimer's disease mouse models. *Cell* 2010, 142, 387–397. [CrossRef]
- 7. Haass, C.; Mandelkow, E. Fyn-tau-amyloid: A toxic triad. Cell 2010, 142, 356–358. [CrossRef]
- 8. Marshall, K.E.; Marchante, R.; Xue, W.F.; Serpell, L.C. The relationship between amyloid structure and cytotoxicity. *Prion* **2014**, *8*, 192–196. [CrossRef]
- Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C.M.; Stefani, M. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* 2002, 416, 507–511. [CrossRef]
- 10. Sengupta, U.; Nilson, A.N.; Kayed, R. The role of amyloid-β oligomers in toxicity, propagation, and immunotherapy. *EBioMedicine* **2016**, *6*, 42–49. [CrossRef]
- Knowles, T.P.; Vendruscolo, M.; Dobson, C.M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* 2014, 15, 384–396. [CrossRef]
- 12. Dobson, C.M. Experimental investigation of protein folding and misfolding. *Methods* 2004, 34, 4–14. [CrossRef]
- 13. Buxbaum, J.N.; Linke, R.P. A molecular history of the amyloidosis. J. Mol. Biol. 2000, 421, 142–159. [CrossRef] [PubMed]
- 14. Dobson, C.M. Principles of protein folding, misfolding and aggregation. Semin. Cell Dev. Biol. 2004, 15, 3–16. [CrossRef]
- 15. Sunde, M.; Serpell, L.C.; Bartlam, M.; Fraser, P.E.; Pepys, M.B.; Blake, C.C. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* **1997**, *273*, 729–739. [CrossRef]
- 16. Nelson, R.; Eisenberg, D. Recent atomic models of amyloid fibril structure. Curr. Opin. Struct. Biol. 2006, 16, 260–265. [CrossRef]
- Olsen, A.; Jonsson, A.; Normark, S. Fibronectin binding mediated by a novel class of surface organelles on Escherichia coli. *Nature* 1989, 338, 652–655. [CrossRef]
- 18. Römling, U.; Bian, Z.; Hammar, M.; Sierralta, W.D.; Normark, S. Curli fibers are highly conserved between Salmonella typhimurium and Escherichia coli with respect to open structure and regulation. *J. Bacteriol.* **1998**, *180*, 722–731. [CrossRef]
- 19. Otzen, D.; Nielsen, P.H. We find them here, we find them there: Functional bacterial amyloid. *Cell Mol. Life Sci.* **2008**, *65*, 910–927. [CrossRef]
- 20. Claessen, D.; Rink, R.; de Jong, W.; Siebring, J.; de Vreugd, P.; Boersma, F.G.; Dijkhuizen, L.; Wosten, H.A. A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in Streptomyces coelicolor by forming amyloid-like fibrils. *Genes Dev.* **2003**, *17*, 1714–1726. [CrossRef]
- 21. Si, K.; Lindquist, S.L.; Kandel, E.R. A neuronal isoform of the aplysia CPEB has prion-like properties. *Cell* **2003**, *115*, 879–891. [CrossRef]
- 22. Iconomidou, V.A.; Vriend, G.; Hamodrakas, S.J. Amyloids protect the silkmoth oocyte and embryo. *FEBS Lett.* **2000**, 479, 141–145. [CrossRef]
- Iconomidou, V.A.; Chryssikos, G.D.; Giorius, V.; Galarius, A.S.; Cordopatis, P.; Hoenger, A.; Hamodrakas, S.J. Amyloid fibril formation propensity is inherent into the hexapeptide tandemly repeating sequence of the central domain of silkmoth chorion proteins of the A-family. J. Struct. Biol. 2006, 156, 480–488. [CrossRef]
- 24. Podrabsky, J.E.; Carpenter, J.F.; Hand, S.C. Survival of water stress in annual fish embryos: Dehydration avoidance and egg envelope amyloid fibers. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2001**, *280*, R123–R131. [CrossRef]
- 25. Fowler, D.M.; Koulov, A.V.; Alory-Jost, C.; Marks, M.S.; Balch, W.E.; Kelly, J.W. Functional amyloid formation within mammalian tissue. *PLoS Biol.* 2005, *4*, e6. [CrossRef]
- Berson, J.F.; Theos, A.C.; Harper, D.C.; Tenza, D.; Raposo, G.; Marks, M.S. Proprotein convertase cleavage liberates a fibrillogenic fragment of a resident glycoprotein to initiate melanosome biogenesis. J. Cell. Biol. 2003, 161, 521–533. [CrossRef]
- Antonets, K.S.; Belousov, M.V.; Sulatskaya, A.I.; Belousova, M.E.; Kosolapova, A.O.; Sulatsky, M.I.; Andreeva, E.A.; Zykin, P.A.; Malovichko, Y.V.; Shtark, O.Y.; et al. Accumulation of storage proteins in plant seeds is mediated by amyloid formation. *PLoS Biol.* 2020, 18, e3000564. [CrossRef]
- 28. Wang, K.; McClure, J.; Tu, A. Titin: Major myofibrillar components of striated muscle. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 3698–3702. [CrossRef]
- 29. Maruyama, K.; Kimura, S.; Ohashi, K.; Kuwano, Y. Connectin, an elastic protein of muscle. Identification of "titin" with connectin. *J. Biochem.* **1981**, *89*, 701–709. [CrossRef]
- 30. Koretz, J.F.; Irving, T.C.; Wang, K. Filamentous aggregates of native titin and binding of C-protein and AMP-deaminase. *Arch. Biochem. Biophys.* **1993**, *304*, 305–309. [CrossRef]
- Kellermayer, M.S.; Bustamante, C.; Granzier, H.L. Mechanics and structure of titin oligomers explored with atomic force microscopy. *Biochim. Biophys. Acta* 2003, 1604, 105–114. [CrossRef] [PubMed]
- 32. Wright, C.F.; Teichmann, S.A.; Clarke, J.; Dobson, C.M. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* **2005**, *438*, 878–881. [CrossRef] [PubMed]
- 33. Borgia, A.; Kemplen, K.R.; Borgia, M.B.; Soranno, A.; Shammas, S.; Wunderlich, B.; Nettels, D.; Best, R.B.; Clarke, J.; Schuler, B. Transient misfolding dominates multidomain protein folding. *Nat. Commun.* **2015**, *6*, 8861. [CrossRef]

- Bobylev, A.G.; Galzitskaya, O.V.; Fadeev, R.S.; Bobyleva, L.G.; Yurshenas, D.A.; Molochkov, N.V.; Dovidchenko, N.V.; Selivanova, O.M.; Penkov, N.V.; Podlubnaya, Z.A.; et al. Smooth muscle titin forms in vitro amyloid aggregates. *Biosci. Rep.* 2016, 36, e00334. [CrossRef]
- Yakupova, E.I.; Vikhlyantsev, I.M.; Bobyleva, L.G.; Penkov, N.V.; Timchenko, A.A.; Timchenko, M.A.; Enin, G.A.; Khutzian, S.S.; Selivanova, O.M.; Bobylev, A.G. Different amyloid aggregation of smooth muscles titin in vitro. *J. Biomol. Struct. Dyn.* 2018, 36, 2237–2248. [CrossRef] [PubMed]
- 36. Bobylev, A.G.; Fadeev, R.S.; Bobyleva, L.G.; Kobyakova, M.I.; Shlyapnikov, Y.M.; Popov, D.V.; Vikhlyantsev, I.M. Amyloid aggregates of smooth-muscle titin impair cell adhesion. *Int. J. Mol. Sci.* **2021**, *22*, 4579. [CrossRef] [PubMed]
- 37. Hill, S.E.; Robinson, J.; Matthews, G.; Muschol, M. Amyloid protofibrils of lysozyme nucleate and grow via oligomer fusion. *Biophys. J.* **2009**, *96*, 3781–3790. [CrossRef]
- 38. Tskhovrebova, L.; Trinick, J. Titin: Properties and family relationships. Nat. Rev. Mol. Cell Biol. 2003, 4, 679–689. [CrossRef]
- Vikhlyantsev, I.M.; Okuneva, A.D.; Shumilina, U.V.; Salmov, N.N.; Bobylev, A.G.; Molochkov, N.V.; Podlubnaya, Z.A. Method for isolation of intact titin (connectin) molecules from mammalian cardiac muscle. *Biochemistry* 2013, 78, 455–462. [CrossRef]
- 40. Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat. Protoc.* **2015**, *10*, 382–396. [CrossRef] [PubMed]
- Svergun, D.I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* 1999, 76, 2879–2886. [CrossRef]
- 42. Noguchi, H.; Takemori, S.; Kajiwara, J.; Kimura, M.; Maruyama, K.; Kimura, S. Chicken breast muscle connectin: Passive tension and I-band region primary structure. *J. Mol. Biol.* 2007, 370, 213–219. [CrossRef] [PubMed]
- Lobanov, M.Y.; Galzitskaya, O.V. The Ising model for prediction of disordered residues from protein sequence alone. *Phys. Biol.* 2011, *8*, 35004. [CrossRef] [PubMed]
- Lobanov, M.Y.; Sokolovskiy, I.V.; Galzitskaya, O.V. IsUnstruct: Prediction of the residue status to be ordered or disordered in the protein chain by a method based on the Ising model. J. Biomol. Struct. Dyn. 2013, 31, 1034–1043. [CrossRef] [PubMed]
- 45. Kumar, S.; Walter, J. Phosphorylation of amyloid beta (Aβ) peptides—A trigger for formation of toxic aggregates in Alzheimer's disease. *Aging* **2011**, *3*, 803–812. [CrossRef]
- Bobyleva, L.G.; Shumeyko, S.A.; Yakupova, E.I.; Surin, A.K.; Galzitskaya, O.V.; Kihara, H.; Timchenko, A.A.; Timchenko, M.A.; Penkov, N.V.; Nikulin, A.D.; et al. Myosin binding protein-C forms amyloid-like aggregates in vitro. *Int. J. Mol. Sci.* 2021, 22, 731. [CrossRef]
- Pan, K.M.; Baldwin, M.; Nguyen, J.; Gasset, M.; Serban, A.; Groth, D.; Mehlhorn, I.; Huang, Z.; Fletterick, R.J.; Cohen, F.E. Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc. Natl. Acad. Sci. USA* 1993, 90, 10962–10966. [CrossRef]
- Sun, Y.; Kakinen, A.; Xing, Y.; Faridi, P.; Nandakumar, A.; Purcell, A.W.; Davis, T.P.; Ke, P.C.; Ding, F. Amyloid self-assembly of hIAPP8-20 via the accumulation of helical oligomers, α-helix to β-sheet transition, and formation of β-barrel intermediates. *Small* 2019, 15, e1805166. [CrossRef]
- 49. Ji, W.; Yuan, C.; Chakraborty, P.; Gilead, S.; Yan, X.; Gazit, E. Stoichiometry-controlled secondary structure transition of amyloidderived supramolecular dipeptide co-assemblies. *Commun. Chem.* **2019**, *2*, 65. [CrossRef]
- Ulamec, S.M.; Radford, S.E. Spot the Difference: Function versus toxicity in amyloid fibrils. *Trends Biochem. Sci.* 2020, 45, 635–636. [CrossRef]
- 51. Eckels, E.C.; Haldar, S.; Tapia-Rojo, R.; Rivas-Pardo, J.A.; Fernández, J.M. The mechanical power of titin folding. *Cell Rep.* **2019**, 27, 1836–1847.e4. [CrossRef]
- 52. Rivas-Pardo, J.A.; Eckels, E.C.; Popa, I.; Kosuri, P.; Linke, W.A.; Fernández, J.M. Work done by titin protein folding assists muscle contraction. *Cell Rep.* **2016**, *14*, 1339–1347. [CrossRef]
- 53. Rousseau, F.; Serrano, L.; Schymkowitz, J.W. How evolutionary pressure against protein aggregation shaped chaperone specificity. *J. Mol. Biol.* **2006**, *355*, 1037–1047. [CrossRef]
- 54. Goldschmidt, L.; Teng, P.K.; Riek, R.; Eisenberg, D. Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 3487–3492. [CrossRef]
- 55. Azuaga, A.I.; Dobson, C.M.; Mateo, P.L.; Conejero-Lara, F. Unfolding and aggregation during the thermal denaturation of streptokinase. *Eur. J. Biochem.* 2002, 269, 4121–4133. [CrossRef]
- 56. Rezaei, H.; Choiset, Y.; Eghiaian, F.; Treguer, E.; Mentre, P.; Debey, P.; Grosclaude, J.; Haertle, T. Amyloidogenic unfolding intermediates differentiate sheep prion protein variants. *J. Mol. Biol.* **2002**, *322*, 799–814. [CrossRef]
- 57. Baxa, U.; Ross, P.D.; Wickner, R.B.; Steven, A.C. The N-terminal prion domain of Ure2p converts from an unfolded to a thermally resistant conformation upon filament formation. *J. Mol. Biol.* **2004**, *339*, 259–264. [CrossRef]
- 58. Morel, B.; Casares, S.; Conejero-Lara, F. A single mutation induces amyloid aggregation in the α-spectrin SH3 domain: Analysis of the early stages of fibril formation. *J. Mol. Biol.* **2006**, *356*, 453–468. [CrossRef]
- 59. Morel, B.; Varela, L.; Conejero-Lara, F. The thermodynamic stability of amyloid fibrils studied by differential scanning calorimetry. *J. Phys. Chem. B.* **2010**, *114*, 4010–4019. [CrossRef]
- 60. Oberhauser, A.F.; Marszalek, P.E.; Carrion-Vazquez, M.; Fernandez, J.M. Single protein misfolding events captured by atomic force microscopy. *Nat. Struct. Biol.* **1999**, *6*, 1025–1028. [CrossRef]
- 61. Trinick, J.; Knight, P.; Whiting, A. Purification and properties of native titin. J. Mol. Biol. 1984, 180, 331–356. [CrossRef]

- 62. Fritz, J.D.; Swartz, D.R.; Greaser, M.L. Factors affecting polyacrylamide gel electrophoresis and electroblotting of high-molecularweight myofibrillar proteins. *Anal. Biochem.* **1989**, *180*, 205–210. [CrossRef]
- Makhnovskii, P.A.; Zgoda, V.G.; Bokov, R.O.; Shagimardanova, E.I.; Gazizova, G.R.; Gusev, O.A.; Lysenko, E.A.; Kolpakov, F.A.; Vinogradova, O.L.; Popov, D.V. Regulation of proteins in human skeletal muscle: The role of transcription. *Sci. Rep.* 2020, 10, 3514. [CrossRef]
- 64. Wiśniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362. [CrossRef]
- 65. Sreerama, N.; Woody, R.W. Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **2000**, *287*, 252–260. [CrossRef]
- 66. Venyaminov, S.; Prendergast, F.G. Water (H2O and D2O) molar absorptivity in the 1000–4000 cm⁻¹ range and quantitative infrared spectroscopy of aqueous solutions. *Anal. Biochem.* **1997**, 248, 234–245. [CrossRef]
- 67. Feigin, L.A.; Svergun, D.A. *Structure Analysis by Small-angle X-ray and Neutron Scattering*; Springer: Berlin/Heidelberg, Germany; Plenum Press: New York, NY, USA, 1987.
- Senin, A.A.; Potekhin, S.A.; Tiktopulo, E.I.; Filomonov, V.V. Differential scanning microcalorimetr SCAL-1. J. Therm. Anal. Cal. 2000, 62, 153–160. [CrossRef]
- 69. Privalov, P.L.; Potekhin, S.A. Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods Enzymol.* **1986**, *131*, 4–51. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Article



Genome-Wide Identification of the SUN Gene Family in Melon (*Cucumis melo*) and Functional Characterization of Two *CmSUN* Genes in Regulating Fruit Shape Variation

Ming Ma, Suya Liu, Zhiwei Wang, Ran Shao, Jianrong Ye, Wei Yan, Hailing Lv, Agula Hasi * and Gen Che *

Key Laboratory of Herbage & Endemic Crop Biology, Ministry of Education, School of Life Sciences, Inner Mongolia University, Hohhot 010070, China

* Correspondence: hasiagula@imu.edu.cn (A.H.); chegen@imu.edu.cn (G.C.)

Abstract: Melon (*Cucumis melo*) is an important economic crop cultivated worldwide. A unique *SUN* gene family plays a crucial role in regulating plant growth and fruit development, but many *SUN* family genes and their function have not been well-characterized in melon. In the present study, we performed genome-wide identification and bioinformatics analysis and identified 24 *CmSUN* family genes that contain integrated and conserved IQ67 domain in the melon genome. Transcriptome data analysis and qRT-PCR results showed that most *CmSUNs* are specifically enriched in melon, reproductive organs, such as young flowers and ovaries. Through genetic transformation in melons, we found that overexpression of *CmSUN23-24* and *CmSUN25-26-27c* led to an increased fruit shape index, suggesting that they act as essential regulators in melon fruit shape variation. Subcellular localization revealed that the CmSUN23-24 protein is located in the cytoplasmic membrane. A direct interaction between CmSUN23-24 and a Calmodulin protein CmCaM5 was found by yeast two-hybrid assay, which indicated their participation in the calcium signal transduction pathway in regulating plant growth. These findings revealed the molecular characteristics, expression profile, and functional pattern of the *CmSUN* genes, and may provide the theoretical basis for the genetic improvement of melon fruit breeding.

Keywords: melon; CmSUN; IQ67 domain; expression analysis; overexpression phenotype; fruit shape regulation; protein interaction

1. Introduction

Melon (*Cucumis melo* L.) is a globally cultivated horticulture crop, bearing sweet and pleasant fruit with high nutritional value [1–3]. The long-term domestication process and wide distribution have produced multiple melon cultivars with diversified fruit traits, especially the fruit shape and size [4]. Fruit shape index (FSI) refers to the ratio of the fruit's longitudinal diameter to the transverse diameter, and is an important agronomic trait that affects melon consumption [5]. Melon FSI can vary between round, oval, and long, and has always been used as a fruit quality screening standard to classify different market groups and satisfy consumer preferences [6]. In melon, the molecular mechanism underlying the outstanding diversity of FSI needs to be further studied [3]. The establishment of fruit shape is mediated by cell differentiation and cell expansion, which are caused by multiple endogenous gene activity [7]. Considerable research on the tomato has found that the *SUN* gene is one of the pivotal regulators in fruit shape regulation [8,9]. Overexpression of *SISUN/SIIQD12*leads to long slender fruits with an increased number of vertical cells and a reduced number of horizontal cells, and *SISUNs* regulate the expression of genes involved in cell division, cell wall development, and the auxin pathway [10,11].

SUN genes encode the IQD protein, a plant-specific calmodulin-binding protein that consists of 67 amino acid residues, containing three copies of the IQ motif and separated in precise spacing by short sequences of 11 and 15 amino acid residues. IQ67 domain

Citation: Ma, M.; Liu, S.; Wang, Z.; Shao, R.; Ye, J.; Yan, W.; Lv, H.; Hasi, A.; Che, G. Genome-Wide Identification of the *SUN* Gene Family in Melon (*Cucumis melo*) and Functional Characterization of Two *CmSUN* Genes in Regulating Fruit Shape Variation. *Int. J. Mol. Sci.* 2022, 23, 16047. https://doi.org/ 10.3390/ijms232416047

Academic Editor: Wajid Zaman

Received: 4 November 2022 Accepted: 10 December 2022 Published: 16 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). usually contains 1~3 core 'IQ' motifs, 1~3 '1-8-14' motifs, and 1~4 '1-5-10' motifs [12,13]. The conserved motifs can bind to multifunctional calmodulin/calmodulin-like (CaM/CmL) proteins [14,15]. CaM/CmL transmits signals to downstream IQD receptor proteins through different forms of interaction with calcium signals [16–22]. In the *Arabidopsis thaliana*, the *SUN* family genes *AtIQD11*, *AtIQD14*, and *AtIQD16* have also been shown to regulate cell morphology by elongating or distorting cell shape [23].

SUN family genes participate in many aspects of plant development. Arabidopsis AtIQD1 gene was the first to be discovered [12], and can activate the expression of glucosinolate metabolic pathway-related genes, thus increasing the glucosinolate content and improving the disease resistance [13]. Knockout of AtIQD5 leads to variations in cell size, skeleton, and stability of microtubule [24]. AtIQD18 is mediated by auxin signaling, and its overexpression results in a loose cell morphology phenotype [25]. The IQD role in microtubules may be mediated by a functional unknown domain, DUF4005 [26]. SUN genes have been involved in regulating cell shape and cell arrangement, which are closely related to the plant development and fruit shape index.

In recent years, the *SUN* family genes have been identified in many species. There are 33 *SUN* members in *Arabidopsis* [12], 29 members in *Oryza sativa* [12], 33 members in *Solanum lycopersicum* [27], 67 members in *Soybean* [28], 26 members in *Zea mays* [29], and 22 members in *Cucumis sativus* L. [30]. *CsSUN*, a homologous gene of tomato *SISUN*/*SIIQD12*, was the candidate gene for the fruit shape of the cucumber QTL *FS1.2*, and its expression was lower in round-fruit cucumber than in long-fruit cucumber [31]. In watermelon, a 159 bp deletion in *CISUN25-26-27a* underlies the difference between elongated fruit and spherical fruit [32,33]. Pan and Monforte mentioned 21 and 24 melon *SUN* members in the previous study, respectively [30,34]. In *Oryza sativa*, *OsIQD14* can control grain shape by regulating cell shape [35]. In other plants, such as soybean, maize, and Arabidopsis, IQD gene expression can be positively or negatively regulated by hormone treatments and abiotic stress [28,29,36].

In this study, we performed genome-wide identification and bioinformatics exploration of melon *SUN* family genes. The *CmSUNs* expression profile was analyzed by high-throughput sequencing data and verified by qRT-PCR. In the genome-wide association analysis of 297 melon germplasm, *CmSUN23-24* and *CmSUN25-26-27c* were presumed to be underlying the key loci of melon fruit shape [37]. Performing functional analyses through genetic transformation in melon, we found that overexpression of *CmSUN23-24* and *CmSUN25-26-27c* resulted in fruit shape variation. Furthermore, we identified a CaM protein that can directly interact with CmSUN23-24, suggesting that they contribute to fruit shape regulation through a regulatory gene network in melon.

2. Results

2.1. CmSUN Family Gene Identification and Sequence Analyses

A total of 24 CmSUN members were identified by searching the Hidden Markov model PF00612 of the IQ motifs and blasting probe sequences in the melon protein database, including the 21 CmSUNs consistent with the previous study [30,34]. In the CmSUN family, CmSUN9-10a is the shortest protein consisting of 261 amino acids (aa), and the longest protein, CmSUN32b, is 845 aa in length. Physical and chemical analyses showed that the molecular weights of the corresponding proteins range from 29.9 to 93.4 k Dalton (kD), and the theoretical isoelectric points range from 5.59 (CmSUN32b) to 10.87 (CmSUN13-14a) (Table 1). The average theoretical isoelectric point of 24 CmSUN proteins was 9.94, indicating that they contain many basic amino acids.

The 24 *CmSUNs* were unevenly distributed on 12 chromosomes (Figure 1). The gene structure analyses showed that all *CmSUNs* contained exons and introns. The number of exons in *CmSUNs* varies from 2 to 6. *CmSUN19b*, *19c*, *25-26-27a*, and *25-26-27b* have three exons that display similar-length arrangements in structures following a 'medium—short—long' pattern; *CmSUN1-2a*, *1-2b*, *3*, *4*, *6*, *7-8*, *11*, *13-14a*, *13-14b*, and *17-18a* have five exons that display similar-length arrangement in structures following a

'short—medium—medium—short—long' pattern; *CmSUN5*, 30-31, and 32b have six exons that display similar-length arrangements in structures following a 'short—medium—medium—short—long—short' pattern (Figure 2). The strong similarity of the three groups of *CmSUN* genes may indicate their conservation in the gene evolution process and function mode.

Name	Gene ID	Chromosome Distribution	ORF (bp)	Amino Acid Length (aa)	MW (kD)	pI
CmSUN1-2a	MELO3C014258.2.1	Chrom05	1473	490	53.7	9.87
CmSUN1-2b	MELO3C009321.2.1	Chrom04	1491	496	56.4	10.34
CmSUN3	MELO3C025505.2.1	Chrom09	1449	482	53.4	10.42
CmSUN4	MELO3C012442.2.1	Chrom10	1395	464	52.1	10.06
CmSUN5	MELO3C009991.2.1	Chrom02	1302	433	48.5	10.22
CmSUN6	MELO3C016880.2.1	Chrom07	1338	445	49.6	10.47
CmSUN7-8	MELO3C024381.2.1	Chrom01	1221	406	45.8	10.26
CmSUN9-10a	MELO3C016091.2.1	Chrom07	786	261	29.9	10.04
CmSUN9-10b	MELO3C007235.2.1	Chrom08	1107	368	41.9	10.31
CmSUN11	MELO3C005137.2.1	Chrom09	1440	479	54.1	10.02
CmSUN13-14a	MELO3C022423.2.1	Chrom11	1614	537	60.5	10.87
CmSUN13-14b	MELO3C017768.2.1	Chrom07	1650	549	61.4	10.47
CmSUN17-18a	MELO3C022253.2.1	Chrom11	1539	512	57.5	10.42
CmSUN19a	MELO3C014290.2.1	Chrom05	1452	483	53.8	9.63
CmSUN19b	MELO3C006504.2.1	Chrom06	1269	422	46.9	9.81
CmSUN19c	MELO3C004368.2.1	Chrom05	1128	375	43.4	9.98
CmSUN21a	MELO3C008499.2.1	Chrom06	1410	469	52.8	9.60
CmSUN21b	MELO3C005888.2.1	Chrom09	1125	374	41.8	9.96
CmSUN23-24	MELO3C006884.2.1	Chrom06	1413	470	51.5	10.28
CmSUN25-26-27a	MELO3C015418.2.1	Chrom02	1233	410	45.4	10.25
CmSUN25-26-27b	MELO3C024434.2.1	Chrom01	1293	430	48.3	9.66
CmSUN25-26-27c	MELO3C013004.2.1	Chrom04	1161	386	42.8	10.29
CmSUN30-31	MELO3C002201.2.1	Chrom12	1800	599	65.9	9.80
CmSUN32b	MELO3C010997.2.1	Chrom03	2538	845	93.4	5.59
Chrom01 Chr $\begin{bmatrix} 0 \text{ Mb} \\ 4 \text{ Mb} \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$	rom02 Chrom03 Chrom04	Chrom05 Chrom06	Chrom07 Chrom0 දී ි දු ි	8 Chrom09 Ch	rom10 Chrom11	Chrom12

Table 1. Information and physicochemical properties of 24 CmSUN members.



Figure 1. Distribution of 24 CmSUN genes on melon chromosomes.

Through analyzing the CmSUNs protein domains, we found that arginine (R), isoleucine (I), glutamine (Q), and leucine (L) sites are highly conserved in different species (Figure 3A). Multiple sequence alignments of CmSUNs showed that all of the proteins contain a typical IQ67 conserved domain with 1~3 core 'IQ' motifs (IQXXXRGXXXR), or the less-restricted motifs ([ILV]QXXXRXXXR[RK]), 1~3 '1-8-14' motifs ([FILVW]XXXXX[FAILVW] XXXXX[FILV]), and 1~4 '1-5-10' motifs ([FILVW]XXXX[FILV]XXXX[FILVW]). The three IQ motifs are precisely separated by 11 and 15 amino acids [12] (Figure 3B).



Figure 2. Gene structure of 24 *CmSUN* genes. The pink box represents the exon, the black lines represents the intron, blue box represents the UTR.



Figure 3. CmSUNs protein sequence analyses. **(A)** The logos show the conserved sites in IQ67 domain in CmSUN sequences. **(B)** Multiple sequence alignment of 24 CmSUN proteins. The red lines represent the 'IQ' core motif; The blue lines represent the '1-8-14' motif; The green lines represent the '1-5-10' motif; '11 aa' and '15 aa' represent the amino acid interval between the three 'IQ' motifs.

2.2. Phylogenetic Tree Analysis of SUN Family Proteins

The phylogenetic tree was constructed by aligning the 33 AtSUNs in *Arabidopsis* [12], 22 CsSUNs in *Cucumis sativus* [30], 33 SISUNs in *Solanum lycopersicum* [27], and 24 CmSUNs in *Cucumis melo*. The SUN family members can be divided into five clades I~V (Figure 4). CmSUN25-26-27a, 25-26-27b, and 25-26-27c are branched in the same clade with the fruit shape regulator gene CsSUN25-26-27a [31]. CmSUN23-24 and CmSUN25-26-27c are the homolog genes of CsSUN23-24 and CsSUN25-26-27c, which also participate in regulating fruit shape in cucumber [38]. CmSUN11 is homologous to the cell shape-related AtSUN11 gene; CmSUN1-2a and 1-2b are branched in the same clade with AtSUN1, which is the foundational SUN member in Arabidopsis.



Figure 4. Phylogenetic tree analyses of SUN proteins from *Arabidopsis, Solanum Lycopersicon, Cucumis sativus,* and *Cucumis melo.* The blue represents AtSUNs, the green represents CsSUNs, the yellow represents CmSUNs, and the red represents SISUNs. 'I~V' indicates different subfamily of SUN members in four species. The asterisk indicates the functionally studied CmSUN genes.

2.3. Expression Profiling of CmSUN Genes

The heat map of *CmSUN* expression was generated according to the transcriptomic analyses in different melon tissues (Figure 5A). The significant differences in relative expression in the ten different tissues are indicated by salient letter notation. The relatively high transcripts of the *CmSUNs* accumulated in melon stem, ovary, and flower. In the late development stages of the fruit, such as the growing stage (18 days after pollination, DAP), ripening stage (36 DAP), climacteric stage (determined according to breathalyzer), and post-climacteric stage (48 h after climacteric stage), the *CmSUNs* displayed low expression levels. We further verified the expression patterns of *CmSUN* genes by performing qRT-PCR in different melon tissues. The qRT-PCR results were consistent with the transcriptome analyses (Figure 5B). *CmSUN25-26-27a* has the highest expression in female flowers, and may have an active role in regulating gynoecium development; *CmSUN3*, *CmSUN17-18a*, *CmSUN23-24*, and *CmSUN30-31* showed the highest expression in the ovary compared to the other tissues, suggesting their potential role in fruit development regulation.


Figure 5. Expression analysis of *CmSUNs* in melon. (**A**) Heat map of 24 *CmSUN* gene expression in different melon tissues. (**B**) Expression profiles of part *CmSUN* genes in melon. Rt, S, L, FF, MF, O, G, R, C, and P represent root, stem, leaf, female flower, male flower, ovary, growing stage, ripening stage, climacteric stage, and post-climacteric stage, respectively. Three biological replicates and three technique replicates were performed for each qRT-PCR analysis. 'a–g' indicates the significance difference in the tissues.

2.4. Overexpression of CmSUN23-24 and CmSUN25-26-27c Resulted in Melon Fruit Shape Variation

CmSUN23-24 and CmSUN25-26-27c were presumed as candidate genes for melon fruit shape QTL [37]. To verify the role of CmSUN23-24 and CmSUN25-26-27c in fruit regulation, we facilitated the transformation of the genes driven by the 35S promoter into the melon. We obtained six *CmSUN25-26-27c*-overexpressed transgenic lines, and two representative phenotypic lines were chosen for further characterization. Compared to the wild type plants, the vertical diameter of the *CmSUN25-26-27c-Oe* mature fruits was significantly increased, while the horizontal diameter showed no difference (Figure 6A–C). Thus, the melon fruit shape index is more enlarged in the transgenic lines than in WT (Figure 6D). In *CmSUN25-26-27c-Oe* lines, the fruit shape index was about 1.23 (Oe-L1, n > 20) and 1.21 (Oe-L2, n > 20), and the average FSI 1.22 was 11.9% larger than that in WT (1.09, n > 100). We measured the vertical and horizontal diameter of the transgenic fruits at DAP 1~7, which was the crucial time for determining the melon fruit shape. The data showed a similar fruit enlargement curve in the WT and transgenic lines (Figure 7A–C). CmSUN25-26-27c-Oe lines had a higher FSI at early fruit elongation stages than WT, particularly in the fruits at 1 DAP and 3 DAP (Figure 7A–C). Then we examined *CmSUN25-26-27c* expression in the ovaries at 1 DAP and found twofold elevated expression in the overexpression lines (Figure 7D). In the ovaries at 3 DAP, *CmSUN25-26-27c* expression was increased by about 7.2 and 3.3 times in Oe-L1 and Oe-L2 lines compared to the WT, respectively (Figure 7E). A scatterplot analysis was performed to verify the relevant between FSI and CmSUN25-26-27*c* expression (Figure 7F). The correlation coefficient indicates the significant relevance $(R^2 = 0.76)$ between gene expression and FSI (Figure 7F). In the transcriptomic analyses, *CmSUN25-26-27c* transcripts were also highly accumulated in the male flower (Figure 7G). The phenotypic analyses suggested that the *CmSUN25-26-27c* gene positively regulated melon fruit elongation and functions in the early developmental phase.



Figure 6. Phenotypic analysis of *CmSUN25-26-27c*-overexpressed transgenic lines. (**A**) Phenotypic observation of the fruits in transgenic lines. (**B**) Fruit vertical diameter significantly increased in transgenic lines. (**C**) Fruit horizonal diameter comparison between transgenic lines and WT. (**D**) Statistical analyses of fruit shape index in transgenic lines and WT. Scale bars = 2 cm. Oe-L1 and Oe-L2 indicate *CmSUN25-26-27c*-overexpressed transgenic lines. **** *p* value < 0.0001.



Figure 7. Statistical data of *CmSUN25-26-27c*-overexpressed transgenic fruits. (**A**,**B**) Vertical and horizonal diameter comparison of the fruits at 1 DAP (day after pollination), 3 DAP, 5 DAP, and 7 DAP in transgenic lines and WT. (**C**) Statistical analyses of ovary shape index in transgenic lines and WT in DAP 1~7. (**D**,**E**) qRT-PCR analyses of *CmSUN25-26-27c* expression in the ovary at 1 DAP and 3 DAP. Oe-L1 and Oe-L2 indicate *CmSUN25-26-27c*-overexpressed transgenic lines. (**F**) Scatterplot analysis shows the relevant of FSI and *CmSUN25-26-27c* expression. (**G**) Schematic model of relative expression of *CmSUN25-26-27c* in different melon tissues. * *p* value < 0.05; ** *p* value < 0.01; *** *p* value < 0.001.

In *CmSUN23-24-Oe* lines, the altered vertical diameter and unchanged horizontal diameter resulted in the increased fruit shape index (Figure 8A–D). The average FSI (1.18, n > 40) of the mature fruits in *CmSUN23-24-Oe* lines was 8.3% larger than in WT (1.09, n > 100) (Figure 8D). *CmSUN23-24-Oe* lines also had increased vertical diameter and FSI at early developmental stages (Figure 9A–C). *CmSUN23-24* expression was dramatically

increased at 1 DAP and 3 DAP in ovaries in the severe line (Figure 9D,E). The *CmSUN23-24* expression showed a weak relevance ($R^2 = 0.31$) with the FSI (Figure 9F). The transcriptomic analyses in different melon tissues showed that *CmSUN23-24* was highly accumulated in female flowers and ovary (Figure 9G). The phenotype of *CmSUN23-24-Oe* lines is similar to *CmSUN25-26-27c-Oe* lines, indicating their redundant function in mediating melon fruit size development.



Figure 8. Phenotypic analysis of *CmSUN23-24*-overexpressed transgenic lines. (**A**) Phenotypic observation of the fruits in transgenic lines. (**B**) Fruit vertical diameter significantly increased in transgenic lines. (**C**) Fruit horizonal diameter comparison between transgenic lines and WT. (**D**) Statistical analyses of fruit shape index in transgenic lines and WT. Scale bars = 2 cm. Oe-L1 and Oe-L2 indicate *CmSUN23-24*-overexpressed transgenic lines. ** *p* value < 0.01; *** *p* value < 0.001.



Figure 9. Statistical data of *CmSUN23-24*-overexpressed in transgenic fruits. (**A**,**B**) Vertical and horizonal diameter comparison of the fruits at 1 DAP, 3 DAP, 5 DAP, and 7 DAP in transgenic lines and WT. (**C**) Statistical analyses of ovary shape index in transgenic lines and WT in DAP 1~7. (**D**,**E**) qRT-PCR analyses of *CmSUN23-24* expression in the ovary at 1 DAP and 3 DAP. Oe-L1 and Oe-L2 indicate *CmSUN23-24*-overexpressed transgenic lines. (**F**) Scatterplot analysis showed the positive relevant of FSI and *CmSUN23-24* expression. (**G**) Schematic model of relative expression of *CmSUN23-24* in different melon tissues. * *p* value < 0.05; ** *p* value < 0.01; *** *p* value < 0.001.

2.5. Yeast Two-Hybrid and Subcellular Localization

The transcriptional activation activity of CmSUN23-24 and CmSUN25-26-27c proteins were examined. We found that CmSUN23-24 and CmSUN25-26-27c did not contain transcriptional activation activity (Figure 10A). CmCaM5 and CmCmL11 are homologous proteins of AtCaM1, which can interact with several IQD proteins. The Y2H result showed that CmSUN23-24 had a direct interaction with CmCaM5 protein (Figure 10B). However, there was no interaction between CmSUN25-26-27c, CmCaM5 and CmCmL11. We performed subcellular localization in tobacco leaves and found that CmSUN23-24 was localized in the cell membrane (Figure 10C).



Figure 10. CmSUN23-24 has a protein interaction with CmCaM5. (**A**) Transcriptional activation of CmSUN23-24 and CmSUN25-26-27c protein. pGBKT7-53 + pGADT7-T and pGBKT7-lam + pGADT7-T were served as positive and negative control, respectively. The values 10^0 , 10^1 and 10^2 indicate different concentration solutions; -Trp/-His/-Ade indicates DO supplement; X- α -gal was a chromogenic substrate of yeast galactosidase. (**B**) Yeast-two hybrid assays were performed between CmSUN23-24, CmSUN25-26-27c, and CmCaM5. (**C**) Subcellular localization of CmSUN23-24 in tobacco leaves. Bright indicates light field; GFP indicates green fluorescence excitation field; mCherry indicates fluorescence excitation field; Merged indicates the overlapping of the three fields. Scale bar = 20 µm.

3. Discussion

SUN family genes participate in plant disease defense [13], stress resistance, plant development [28,29,36], cell arrangement, and cytoskeleton regulation [8,9,24,25]. Arabidopsis [12] and Solanum lycopersicum [27] have nearly 30 members, and Soybean has 67 more members [28]. In previous studies, Monforte identified 24 CmSUNs in melon by blasting the protein sequences of the known tomato SUN genes into the melon genome [34]; Pan mentioned 21 proteins containing both IQ67 and DUF4005 domains and identified Cm-SUNs by their homology with the Arabidopsis genes [30]. The integrity of the IQ67 domain is crucial in binding with the Ca^{2+} receptors Calmodulin/Calmodulin-like (CaM/CmL) [39]. However, some SUN members were identified incompletely due to the low conservation of the IQ67 domain in sequence and quantity during evolution. In this study, we considered the protein containing an integrated and typical conserved IQ67 domain with three IQ motifs precisely separated by 11 and 15 amino acid residues to be the *CmSUN* genes, following the same criteria as the SUN family gene identification in Arabidopsis and Oryza sativa [12]. We characterized 24 CmSUN members, including 21 consensus genes with the previous work. Moreover, we found two other protein families, Myosin (7 members) and CaMTA (2 members), and several proteins that contained IQ motifs. It is noteworthy that

nine genes of the Myosin and CaMTA family contain 1–3 IQ motifs, but possess unique amino acid spacing patterns. The two gene families have long amino acid sequences (744–1523 aa), high molecular weights (84.7–172.8 kD), and average low isoelectric points (7.6). CmSUNs are relatively short in length, have a high average isoelectric point (9.9), and contain a typical conserved IQ67 domain and a conserved exon-intron structure. During the evolution, the first IQ motif is the most conservative in the sequences, and the second and third IQ are more distinctive [12]. In our preliminary results, MELO3C005949.2.1 and MELO3C005636.2.1, which contain two IQ motifs and the specific intervals, were also annotated as IQD proteins. However, considering the absence of the first IQ motif and low sequence conservation, we did not nominate the two genes as CmSUN members. Taken together, our solid work on identifying CmSUN family members in melon provided a theoretical foundation for further characterization of their function and molecular mechanism.

Expression analyses in melon tissues showed that *CmSUN* transcripts were mainly accumulated in the lateral organs and the young fruits. Most members tended to be highly expressed in flower and ovary, and gradually disappeared in the fruit at the late developmental stage. In cucumber, *CsSUN30-31* had a high expression level in various tissues, which may have been involved in the many aspects of plant growth. The expression of *CsSUN, CsSUN17-18b*, and other three *SUN* genes was correlated with cucumber fruit development. *CsSUN23-24* and *CsSUN25-26-27c* expression were significantly higher in long fruit than short fruit, while *CsSUN21a* and the other two genes had much higher expression in short fruit [38], meaning they had the opposite role in regulating fruit elongation. In tomato, *SISUN1* and *SISUN28* showed high expression in growing fruits; *SISUN5, SISUN11, SISUN12, SISUN21, SISUN22,* and *SISUN27* were expressed actively in vegetative growth [27]. The expression pattern of *SISUNs* in tomato was similar to *CmSUNs* in different melon tissues, which may imply their conservative function in horticulture plant development.

In the functional analyses, we revealed that the overexpression of *CmSUN23-24* and *CmSUN25-26-27c* resulted in elongated fruits and increased fruit shape index, suggesting that they had an important role in melon fruit shape regulation. In cucumber, *CsSUN* (*CsSUN25-26-27a*) is a positive regulator in fruit elongation, and the phenotype has a consistency with its expression in fruits of different lengths [38]. In watermelon, *ClSUN25-26-27a* is a pivotal factor in regulating the round or elongated fruit shape [32,33]. In melon, *CmSUN25-26-27c* is homologous to *CsSUN25-26-27a*, *CsSUN25-26-27c*, and *ClSUN25-26-27a*. Overexpression of *CmSUN23-24* and *CmSUN25-26-27c* altered the fruit shape at the early ovary developmental stage. The increased expression level, enlarged fruit shape index, and significant correlation coefficient in melon transgenic fruits demonstrated that *CmSUN23-24* and *CmSUN25-26-27c* mediate fruit shape index variation.

SUN family genes usually act as downstream targets in the calcium signal transduction pathway, which contributes to various physiological activities in plant development [40–43]. Most *CmSUNs* are rich in basic amino acids and hydrophobic amino acids and provide advantages for them to bind with the Ca²⁺ receptors Calmodulin/Calmodulinlike (CaM/CmL) [23,35]. Through performing the yeast two-hybrid assay, we found that CmSUN23-24 can directly bind to CmCaM5. CmSUN23-24 localized in the cell membrane, which may be regulated by the CaM and Ca²⁺ binding complex in Ca²⁺ signal transduction, and then participated in melon plant development. In Arabidopsis, IQD1 localized in microtubules and interacted with KLCR1 (kinesin light chain-related) protein or CaM2 to recruit them for cell division regulation [39]. AtIQD5 and AtIQD18 can also regulate the cell size and cell shape by protein interaction with microtubulin and then affecting the microtubule dynamics [24,25]. The combination between IQD and microtubulin was mediated by a DUF4005 domain [26]. CmSUN23-24 and CmSUN25-26-27c both contained the DUF4005 domain. Therefore, we speculated that CmSUN23-24 positively regulates melon fruit elongation through the CmCaM5-dependent Ca²⁺ signaling transduction pathway and may have a positive regulation in microtubules. Although CmSUN25-26-27c has the potential CaM binding sites, there was no interaction between CmSUN25-26-27c and

CmCaM5, implying that there may other CmCaMs/CmLs working with CmSUN25-26-27c and then participating in the calcium signal and microtubule regulation.

4. Materials and Methods

4.1. Identification of SUN Gene Family Members

The melon protein sequence file (CM3.6.1_pep) was downloaded from the CuGenDB (http://cucurbitgenomics.org/, accessed on 5 January 2022) online website [44], which refers to the melon genome reported by Garcia-Mas [45]. The Hidden Markov model file (PF00612) of the IQ motif was downloaded from the Pfam (http://pfam.xfam.org/, accessed on 5 January 2022) database [46]. The IQD protein sequences of Arabidopsis, Solanum lycopersicum and Cucumis sativus was downloaded from previous study [12]. We searched the melon IQD proteins in HMM (http://www.hmmer.org/, accessed on 6 January 2022) software [47] using the Hidden Markov model (PF00612) and blasted the melon protein data using IQD sequences from Arabidopsis, tomato and cucumber as the probe in CuGenDB. Redundant sequences were removed and the remaining sequences were aligned through the online websites of Pfam (http://pfam.xfam.org/, accessed on 9 January 2022) and Smart (http://smart.embl-heidelberg.de/, accessed on 9 January 2022) [48] to verify the existence of the IQ motifs. The integrity of IQ67 conserved domain was demonstrated by multi-sequence alignment in MEGA7 software [49]. The protein sequences with three IQ motifs separated by 11 and 15 amino acid residues were identified as CmSUNs.

4.2. Sequence Analysis and Phylogenetic Tree Construction

Open reading frame (ORF) length, molecular weight (MW) and theoretical isoelectric point (pI) of *CmSUN* members were analyzed by ExPASy online website (http://web.expasy.org/protparam/, accessed on 12 January 2022) [50]. Location information of *CmSUN* members on chromosome was obtained from CuGenDB, and the chromosome map was drawn by the MG2C (http://mg2c.iask.in/mg2c_v2.0/, accessed on 16 January 2022) [51]. The CmSUN protein sequences were aligned by the Clustal W program in MEGA7 and visualized by Jalview [52]. Conserved sites of CmSUN protein domain were predicted by online software WEBLOGO (http://weblogo.berkeley.edu/, accessed on 18 January 2022) [53]. The *CmSUN* gene structure information (intron-exon) was obtained from the GFF file (CM3.6.1_gene.gff3) in the CuGenDB, and visualized by GSDS (http://gsds.cbi.pku.edu.cn/, accessed on 20 January 2022) [54]. The neighbor-joining method in MEGA7 software was utilized to construct a phylogenetic tree. The bootstrap value was 1000. SUN members in *Arabidopsis, Cucumis sativus, Solanum lycopersicum*, and *Cucumis melo* were compared. The phylogenetic tree was visualized through iTOL online software (https://itol.embl.de/, accessed on 28 January 2022) [55].

4.3. Plant Materials and Growth Conditions

The melon (*Cucumis melo* cv. Hetao) inbred lines were cultivated in Dengkou county of Inner Mongolia region and used in this study. All plant materials were provided by Hasi lab at Inner Mongolia University, Hohhot, China. The melon seedlings grew in an artificial climate chamber under the conditions of 60% humidity, 16 h light (25 °C), and 8 h dark (18 °C). The female fruits at anthesis day were self-pollinated. The female flower, male flower, ovary, fruits from four different developmental stages, roots, stems, and leaves were collected for expression analyses. Samples from the young lateral root, the second or third section of stem, the 3 to 5 true leaves, the unpollinated female flower, and male flowers at anthesis, and the fruit mesocarp were frozen in liquid nitrogen and stored at -80 °C. The tissues were sampled from different plants at same growth phase.

4.4. Expression Analysis and Quantitative Real-Time PCR

Different tissues from melon were used for the total RNA extraction. Raw reads for RNA-Seq data were quoted from our previous work [56]. The expression level of *CmSUNs*

in the above ten periods was obtained from statistical data analyses. The heat map was drawn by using TBtools software [57], and parameters were set as default.

The relative expression level of *CmSUN23-24* and *CmSUN25-26-27c* in the transgenic melon were verified by qRT-PCR. Primers were designed by Primer 5. The primer information was listed in Supplemental Table S1. RNA was reversed to cDNA by using the PrimeScriptTM RT Reagent Kit and gDNA Eraser Kit (Takara Bio, Shiga, Japan). SYBR[®] Premix Ex TaqTM II (Takara Bio, Shiga, Japan) and 96-well Chromo4 Real-Time PCR System were used for qRT-PCR and $2^{-\Delta\Delta CT}$ method was utilized for expression level analyses. Three biological replicates and three technical replicates were performed for each gene.

4.5. Gene Cloning and Plant Transformation

The whole-length coding sequence of *CmSUN23-24* (MELO3C006884) and *CmSUN25-26-27c* (MELO3C013004) were amplified by gene-specific primers. The primers are listed in Supplemental Table S1. The restriction sites *Xba*I and *Bam*HI were used for constructing the overexpression vector. The target gene was recombined into the overexpression vector pCAMBIA-1305 driven by 35S promoter. To obtain transgenic plants, the recombinant plasmids were diluted to 100 ng/ μ L with 2 × SSC solution (sodium citrate buffer, pH = 7.0) and ddH₂O. After 7 h in artificial pollination, 10 μ L plasmid solution was injected into the melon fruits as described previously [58]. Fruits of the T₁ generation transgenic plants were sampled and detected by PCR with a pair of specific primers designed on pCAMBIA-1305 vector.

4.6. Subcellular Localization

The full-length coding sequence of *CmSUN23-24* without the termination codon was jointed to the vector pCAMBIA-1300 at the upstream of GFP and transformed into agrobacterium GV3101. The primers are listed in Supplemental Table S1. The lower epidermis of tobacco leaves was infected for subcellular localization experiments. A confocal laser-scanning microscope was used for observing the fluorescence signals under 488-nm wave length excitation.

4.7. Yeast Two-Hybrid Assay

Full lengths of *CmSUN23-24* and *CmSUN25-26-27c* were constructed to pGBKT7 vectors. Full lengths of *CmCaM5* (MELO3C014698) and *CmCmL11* (MELO3C006491) were constructed to pGBKT7 and pGADT7 vectors, respectively. The pGBKT7 recombinant plasmids were transformed into the yeast strain AH109 (TaKaRa). SD/-Trp/-His/-Ade solid medium with 4 mg/mL X- α -gal and without X- α -gal was used for the activation test. Concentrations of 10⁰, 10⁻¹, and 10⁻² of bacterial concentrations were used for selecting interacted combination. pGBKT7-53+pGADT7-T served as positive control; pGBKT7-lam+pGADT7-T was used as negative control. CmSUN23-24-pGBKT7 or CmSUN25-26-27c-pGBKT7 were co-transformed with CmCaM5-pGADT7 or CmCmL11-pGADT7.

5. Conclusions

We identified 24 *SUN* family genes in melon, and performed a comprehensive analysis. Twenty-four CmSUN members all contained typical IQ67 motifs precisely separated by 11 and 15 amino acids. Most *CmSUNs* transcripts were accumulated in the melon stem, ovary, and flower. The biological function of *CmSUN23-24* and *CmSUN25-26-27c* in regulating melon fruit shape was revealed by the gene transformation analysis. Protein interaction between CmSUN23-24 and CmCaM5 suggested their participation in the Ca²⁺ signaling transduction pathway. Future studies on the function and molecular mechanism of *CmSUNs* would be valuable for uncovering the gene regulation in fruit development in melon.

Supplementary Materials: The supporting information can be downloaded at: https://www.mdpi. com/article/10.3390/ijms232416047/s1.

Author Contributions: Conceptualization, A.H. and G.C.; data curation, J.Y., W.Y. and H.L.; formal analysis, M.M. and S.L.; funding acquisition, A.H. and G.C.; investigation, M.M.; methodology, S.L., Z.W. and R.S.; resources, A.H. and G.C.; supervision, A.H. and G.C.; validation, M.M. and Z.W.; visualization, M.M. and S.L.; writing—original draft, M.M.; writing—review & editing, M.M. and G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (31860563 and 32202513), Applied Technology Research and Development Foundation of Inner Mongolia Autonomous Region (2021PT0001), Natural Science Foundation of Inner Mongolia Autonomous Region (2021BS03002), and the Inner Mongolia University High-Level Talent Research Program (10000-21311201/056).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data and plant materials in relation to this work can be obtained through contacting with the corresponding author Gen Che (chegen@imu.edu.cn).

Acknowledgments: We thanked all the colleagues in our laboratory for providing useful discussions and technical assistance. We are very grateful to the editor and reviewers for critically evaluating the manuscript and providing constructive comments for its improvement.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Rolim, P.M.; Fidelis, G.P.; Padilha, C.E.A.; Santos, E.S.; Rocha, H.A.O.; Macedo, G.R. Phenolic profile and antioxidant activity from peels and seeds of melon (Cucumis melo L. var. reticulatus) and their antiproliferative effect in cancer cells. *Braz. J. Med. Biol. Res.* 2018, *51*, e6069. [CrossRef] [PubMed]
- Fleshman, M.K.; Lester, G.E.; Riedl, K.M.; Kopec, R.E.; Narayanasamy, S.; Curley, R.W.; Schwartz, S.J.; Harrison, E.H. Carotene and novel apocarotenoid concentrations in orange-fleshed Cucumis melo melons: Determinations of β-carotene bioaccessibility and bioavailability. J. Agric. Food Chem. 2011, 59, 4448. [CrossRef]
- 3. Amanullah, S.; Liu, S.; Gao, P.; Zhu, Z.; Zhu, Q.; Fan, C.; Luan, F. QTL mapping for melon (*Cucumis melo* L.) fruit traits by assembling and utilization of novel SNPs based CAPS markers. *Sci. Hortic.* **2018**, 236, 18–29. [CrossRef]
- Nunez-Palenius, H.G.; Gomez-Lim, M.; Ochoa-Alejo, N.; Grumet, R.; Lester, G.; Cantliffe, D.J. Melon fruits: Genetic diversity, physiology, and biotechnology features. *Crit. Rev. Biotechnol.* 2008, 28, 13–55. [CrossRef] [PubMed]
- 5. Qiao, J.; Liu, F.; Chen, Y.; Lian, Y. Research Progress on Inheritance of Fruit Shape in Horticultural Crops. *Acta Hortic. Sin.* 2011, *38*, 1385–1396.
- 6. Zhang, T.; Hong, Y.; Zhang, X.; Yuan, X.; Chen, S. Relationship between Key Environmental Factors and the Architecture of Fruit Shape and Size in Near-Isogenic Lines of Cucumber (*Cucumis sativus* L.). *Int. J. Mol. Sci.* **2022**, *23*, 14033. [CrossRef] [PubMed]
- Ando, K.; Grumet, R. Transcriptional Profiling of Rapidly Growing Cucumber Fruit by 454-Pyrosequencing Analysis. J. Am. Soc. Hortic. Sci. 2010, 135, 291–302. [CrossRef]
- Xiao, H.; Jiang, N.; Schaffner, E.; Stockinger, E.J.; van der Knaap, E. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 2008, 319, 1527–1530. [CrossRef]
- 9. Wu, S.; Xiao, H.; Cabrera, A.; Meulia, T.; van der Knaap, E. SUN regulates vegetative and reproductive organ shape by changing cell division patterns. *Plant Physiol.* **2011**, *157*, 1175–1186. [CrossRef] [PubMed]
- Clevenger, J.P.; Van Houten, J.; Blackwood, M.; Rodríguez, G.R.; Jikumaru, Y.; Kamiya, Y.; Kusano, M.; Saito, K.; Visa, S.; Esther, V.D.K. Network Analyses Reveal Shifts in Transcript Profiles and Metabolites That Accompany the Expression of SUN and an Elongated Tomato Fruit. *Plant Physiol.* 2015, *168*, 1164–1178. [CrossRef]
- 11. Wang, Y.; Clevenger, J.P.; Illa-Berenguer, E.; Meulia, T.; van der Knaap, E.; Sun, L. A Comparison of sun, ovate, fs8.1 and Auxin Application on Tomato Fruit Shape and Gene Expression. *Plant Cell Physiol.* **2019**, *60*, 1067–1081. [CrossRef] [PubMed]
- 12. Abel, S.; Savchenko, T.; Levy, M. Genome-wide comparative analysis of the IQD gene families in Arabidopsis thaliana and Oryza sativa. *BMC Evol. Biol.* **2005**, *5*, 72. [CrossRef] [PubMed]
- 13. Levy, M.; Wang, Q.; Kaspi, R.; Parrella, M.P.; Abel, S. Arabidopsis IQD1, a novel calmodulin-binding nuclear protein, stimulates glucosinolate accumulation and plant defense. *Plant J.* 2005, 43, 79–96. [CrossRef]
- 14. Snedden, W.A.; Fromm, H. Calmodulin as a versatile calcium signal transducer in plants. *New Phytol.* **2010**, *151*, 35–66. [CrossRef] [PubMed]
- 15. Perochon, A.; Aldon, D.; Galaud, J.P.; Ranty, B. Calmodulin and calmodulin-like proteins in plant calcium signaling. *Biochimie* **2011**, *93*, 2048–2053. [CrossRef] [PubMed]

- 16. McCormack, E.; Tsai, Y.C.; Braam, J. Handling calcium signaling: Arabidopsis CaMs and CMLs. *Trends Plant Sci.* 2005, 10, 383–389. [CrossRef] [PubMed]
- 17. Reddy, A.S.; Ali, G.S.; Celesnik, H.; Day, I.S. Coping with stresses: Roles of calcium- and calcium/calmodulin-regulated gene expression. *Plant Cell* **2011**, *23*, 2010–2032. [CrossRef]
- 18. Miao, Z.; Abrams, C.; Wang, L.; Gizzi, A.; He, L.; Lin, R.; Yuan, C.; Loll, P.; Pascal, J.; Zhang, J.F. Structural basis for calmodulin as a dynamic calcium sensor. *Structure* **2012**, *20*, 911–923.
- 19. Ranty, B.; Aldon, D.; Galaud, J.P. Plant calmodulins and calmodulin-related proteins: Multifaceted relays to decode calcium signals. *Plant Signal Behav.* **2006**, *1*, 96–104. [CrossRef]
- 20. Rhoads, A.R.; Friedberg, F. Sequence motifs for calmodulin recognition. FASEB J. 1997, 11, 331–340. [CrossRef]
- 21. Bahler, M.; Rhoads, A. Calmodulin signaling via the IQ motif. FEBS Lett. 2002, 513, 107–113. [CrossRef] [PubMed]
- Hoeflich, K.P.; Ikura, M. Calmodulin in action: Diversity in target recognition and activation mechanisms. *Cell* 2002, 108, 739–742. [CrossRef] [PubMed]
- Burstenbinder, K.; Moller, B.; Plotner, R.; Stamm, G.; Hause, G.; Mitra, D.; Abel, S. The IQD Family of Calmodulin-Binding Proteins Links Calcium Signaling to Microtubules, Membrane Subdomains, and the Nucleus. *Plant Physiol.* 2017, 173, 1692–1708. [CrossRef] [PubMed]
- 24. Liang, H.; Yi, Z.; Pablo, M.; Carolyn, R.; Xu, T.; Yang, Z. The microtubule-associated protein IQ67 DOMAIN5 modulates microtubule dynamics and pavement cell shape. *Plant Physiol.* **2018**, *177*, 1555–1568. [CrossRef]
- 25. Wendrich, J.R.; Yang, B.J.; Mijnhout, P.; Xue, H.W.; Weijers, D. IQD proteins integrate auxin and calcium signaling to regulate microtubule dynamics during Arabidopsis development. *Cold Spring Harb. Lab.* **2018**, 275560. [CrossRef]
- Li, Y.; Huang, Y.; Wen, Y.; Wang, D.; Liu, H.; Li, Y.; Zhao, J.; An, L.; Yu, F.; Liu, X. The domain of unknown function 4005 (DUF4005) in an Arabidopsis IQD protein functions in microtubule binding. *J. Biol. Chem.* 2021, 297, 100849. [CrossRef]
- Huang, Z.; Houten, J.V.; Gonzalez, G.; Xiao, H.; Knaap, E. Genome-wide identification, phylogeny and expression analysis of SUN, OFP and YABBY gene family in tomato. *Mol. Genet. Genom.* 2013, 288, 111–129. [CrossRef]
- 28. Feng, L.; Chen, Z.; Ma, H.; Chen, X.; Li, Y.; Wang, Y.; Xiang, Y. The IQD gene family in soybean: Structure, phylogeny, evolution and expression. *PLoS ONE* **2014**, *9*, e110896. [CrossRef]
- 29. Cai, R.; Zhang, C.; Zhao, Y.; Zhu, K.; Wang, Y.; Jiang, H.; Xiang, Y.; Cheng, B. Genome-wide analysis of the IQD gene family in maize. *Mol. Genet. Genom.* 2016, 291, 543–558. [CrossRef]
- 30. Pan, Y.; Wang, Y.; Mcgregor, C.; Liu, S.; Weng, Y. Genetic architecture of fruit size and shape variation in cucurbits: A comparative perspective. *Theor. Appl. Genet.* **2020**, *133*, 1–21. [CrossRef]
- Pan, Y.P.; Liang, X.J.; Gao, M.L.; Liu, H.Q.; Meng, H.W. Round fruit shape in WI7239 cucumber is controlled by two interacting quantitative trait loci with one putatively encoding a tomato SUN homolog. *Theor. Appl. Genet.* 2017, 130, 573–586. [CrossRef] [PubMed]
- 32. Dou, J.; Zhao, S.; Lu, X.; He, N.; Zhang, L.; Ali, A.; Liu, W. Genetic mapping reveals a candidate gene (CIFS1) for fruit shape in watermelon (*Citrullus lanatus L.*). *Theor. Appl. Genet. Int. J. Breed. Res. Cell Genet.* **2018**, 131, 947–958. [CrossRef] [PubMed]
- 33. Legendre, R.; Kuzy, J.; Mcgregor, C. Markers for selection of three alleles of CISUN25-26-27a (Cla011257) associated with fruit shape in watermelon. *Mol. Breed.* 2020, 40, 19. [CrossRef]
- 34. Monforte, A.J.; Diaz, A.; Caño-Delgado, A.; Van Der Knaap, E. The genetic basis of fruit morphology in horticultural crops: Lessons from tomato and melon. *J. Exp. Bot.* **2014**, *65*, 4625–4637. [CrossRef]
- 35. Yang, B.; Wendrich, J.; Rybel, B.D.; Weijers, D.; Xue, H. Rice microtubule-associated protein IQ67-DOMAIN14 regulates grain shape by modulating microtubule cytoskeleton dynamics. *Plant Biotechnol. J.* **2020**, *18*, 1141–1152. [CrossRef]
- Zentella, R.; Zhang, Z.L.; Park, M.; Thomas, S.G.; Endo, A.; Murase, K.; Fleet, C.M.; Jikumaru, Y.; Nambara, E.; Kamiya, Y.; et al. Global analysis of della direct targets in early gibberellin signaling in Arabidopsis. *Plant Cell* 2007, 19, 3037–3057. [CrossRef] [PubMed]
- 37. Liu, S.; Gao, P.; Zhu, Q.; Zhu, Z.; Liu, H.; Wang, X.; Weng, Y.; Gao, M.; Luan, F. Resequencing of 297 melon accessions reveals the genomic history of improvement and loci related to fruit traits in melon. *Plant Biotechnol. J.* **2020**, *18*, 2545–2558. [CrossRef]
- Jiang, L.; Yan, S.; Yang, W.; Li, Y.; Xia, M.; Chen, Z.; Wang, Q.; Yan, L.; Song, X.; Liu, R.; et al. Transcriptomic analysis reveals the roles of microtubule-related genes and transcription factors in fruit length regulation in cucumber (*Cucumis sativus* L.). Sci. Rep. 2015, 5, 8031. [CrossRef]
- Bürstenbinder, K.; Savchenko, T.; Müller, J.; Adamson, A.W.; Stamm, G.; Kwong, R.; Zipp, B.J.; Dinesh, D.C.; Abel, S. Arabidopsis Calmodulin-binding Protein IQ67-Domain 1 Localizes to Microtubules and Interacts with Kinesin Light Chain-related Protein-1. J. Biol. Chem. 2013, 288, 1822–1871. [CrossRef]
- 40. Batistič, O.; Kudla, J. Analysis of calcium signaling pathways in plants. *Biochim. Et Biophys. Acta* (*BBA*)-*Gen. Subj.* **2012**, 1820, 1283–1293. [CrossRef]
- 41. Alistair, M.H.; Colin, B. The generation of Ca²⁺ signals in plants. Annu. Rev. Plant Biol. 2004, 55, 401–427.
- 42. Dodd, A.N.; Kudla, J.; Sanders, D. The language of calcium signaling. *Annu. Rev. Plant Biol.* **2010**, *61*, 593–620. [CrossRef] [PubMed]
- DeFalco, T.A.; Bender, K.W.; Snedden, W.A. Breaking the code: Ca²⁺ sensors in plant signalling. *Biochem. J.* 2009, 425, 27–40. [CrossRef] [PubMed]

- Zheng, Y.; Wu, S.; Bai, Y.; Sun, H.; Jiao, C.; Guo, S.; Zhao, K.; Blanca, J.; Zhang, Z.; Huang, S.; et al. Cucurbit Genomics Database (CuGenDB): A central portal for comparative and functional genomics of cucurbit crops. *Nucleic Acids Res.* 2019, 47, D1128–D1136. [CrossRef]
- 45. Garcia-Mas, J.; Benjak, A.; Sanseverino, W.; Bourgeois, M.; Baulcombe, D.C. The genome of melon (*Cucumis melo* L.). *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11872–11877. [CrossRef]
- 46. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef]
- 47. Johnson, L.S.; Eddy, S.R.; Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinform.* **2010**, *11*, 431. [CrossRef]
- 48. Letunic, I.; Khedkar, S.; Bork, P. SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Res.* 2021, 49, D458–D460. [CrossRef]
- 49. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]
- 50. Gasteiger, E. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook;* Springer: Berlin/Heidelberg, Germany, 2005; pp. 571–607.
- 51. Chao, J.; Li, Z.; Sun, Y.; Aluko, O.O.; Wu, X.; Wang, Q.; Liu, G. MG2C: A user-friendly online tool for drawing genetic maps. *Mol. Hortic.* **2021**, *1*, 16. [CrossRef]
- 52. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009, 25, 1189–1191. [CrossRef] [PubMed]
- 53. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef] [PubMed]
- 54. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* 2015, 31, 1296–1297. [CrossRef]
- 55. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]
- 56. Tian, Z.; Han, J.; Che, G.; Hasi, A. Genome-wide characterization and expression analysis of SAUR gene family in Melon (*Cucumis melo* L.). *Planta* **2022**, 255, 123. [CrossRef]
- 57. Chen, C.; Rui, X.; Hao, C.; He, Y. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *Cold Spring Harb. Lab.* **2018**, *289660*, 289660.
- 58. Bai, S.; Tian, Y.; Tan, C.; Bai, S.; Hasi, A. Genome-wide identification of microRNAs involved in the regulation of fruit ripening and climacteric stages in melon (*Cucumis melo*). *Hortic. Res.* **2020**, *7*, 13. [CrossRef]



Article Screening Key Genes and Biological Pathways in Nasopharyngeal Carcinoma by Integrated Bioinformatics Analysis

Junhu Tai ¹, Jaehyung Park ¹, Munsoo Han ^{1,2}, and Tae Hoon Kim ^{1,2,*}

- ¹ Department of Otorhinolaryngology-Head & Neck Surgery, College of Medicine, Korea University, Seoul 02841, Republic of Korea
- ² Mucosal Immunology Institute, College of Medicine, Korea University, Seoul 02841, Republic of Korea
- * Correspondence: doctorthk@gmail.com; Tel.: +82-02-920-5486

Abstract: The purpose of this study was to identify the hub genes and biological pathways of nasopharyngeal carcinoma (NPC) through bioinformatics analysis and potential new therapeutic targets. In this study, three datasets were downloaded from the Gene Expression Omnibus (GEO), and differentially expressed genes (DEGs) between NPC and normal tissues were analyzed using the GEO2R online tool. Volcano and heat maps of the DEGs were visualized using the hiplot database. Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the upregulated and downregulated DEGs were performed using the DAVID database. Finally, we established a protein-protein interaction (PPI) network using the STRING database and showed the differential expression of hub genes between the normal and tumor tissues. In all, 109,371,221 upregulated DEGs and 139,226,520 downregulated DEGs were obtained in datasets GSE40290, GSE61218, and GSE53819, respectively, and 18 common differential genes, named co-DEGs, were screened in the three datasets. The most abundant biological GO terms of the co-DEGs were inflammatory response et al. The KEGG pathway enrichment analysis showed that co-DEGs mainly participated in the interleukin (IL)-17 signaling pathway et al. Finally, we identified four hub genes using PPI analysis and observed that three of them were highly expressed in tumor tissues. In this study, the hub genes of NPC, such as PTGS2, and pathways such as IL-17 signaling, were identified through bioinformatics analysis, which may be potential new therapeutic targets for NPC.

Keywords: nasopharyngeal carcinoma; bioinformatics; genes

1. Introduction

Nasopharyngeal carcinoma (NPC) is an aggressive head and neck cancer that forms in the tissues of the nasopharynx with high malignancy, which often occurs in the pharyngeal recess, and is relatively rare compared with other cancers [1]. In 2018, approximately 130,000 new NPC cases and 73,000 related deaths were reported in Southeast Asia [2]. The occurrence and development of NPC are related to various factors, including Epstein-Barr virus (EBV) infection [3]. EBV-associated NPC is extremely sensitive to radiation therapy, while squamous histological subtypes are much less sensitive. There are also significant differences in clinical manifestations and treatment responses between undifferentiated and squamous variants [4]. Concurrent chemoradiotherapy is one of the main treatment methods for NPC; however, the emergence of chemotherapy resistance and high incidence of adverse events limit its application [5].

Bioinformatics is an analytical method that uses mathematical, statistical, and computational methods to process and analyze biological data, which differs from traditional laboratory work [6]. For example, Song et al. analyzed the key genes of NPC using bioinformatics [7], and Yue et al. expounded on the differentially expressed genes (DEGs) in NPC tissues and their correlation with the recurrence and metastasis of NPC [8]. Although

Citation: Tai, J.; Park, J.; Han, M.; Kim, T.H. Screening Key Genes and Biological Pathways in Nasopharyngeal Carcinoma by Integrated Bioinformatics Analysis. *Int. J. Mol. Sci.* **2022**, *23*, 15701. https://doi.org/10.3390/ ijms232415701

Academic Editors: Weg M. Ongkeko and Wajid Zaman

Received: 7 November 2022 Accepted: 9 December 2022 Published: 11 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). many studies have focused on identifying biomarkers related to NPC, a more comprehensive analysis is needed to explore better molecular targets to treat NPC and clarify its biological pathways.

The purpose of this study was to screen DEGs and co-DEGs in multiple datasets by analyzing NPC-related datasets in the Gene Expression Omnibus (GEO) database [9], and to conduct gene ontology (GO) terminology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Finally, hub genes were obtained by constructing a protein-protein interaction (PPI) network, and the expression of hub genes in NPC was studied.

2. Results

2.1. Identification of DEGs in the Three GEO Data Sets

DEGs were defined as follows: when p < 0.05 and fold change > 2, DEGs were upregulated; when p < 0.05 and fold change < 2, DEGs were downregulated. After online analysis using GEO2R, 109,371,221 upregulated DEGs and 139,226,520 downregulated DEGs were obtained in datasets GSE40290, GSE61218, and GSE53819, respectively. The corresponding volcano maps of GSE40290 (Figure 1a), GSE61218 (Figure 1b), and GSE53819 (Figure 1c) are shown. Heat maps of the top 20 DEGs in GSE40290 (Figure 2a), GSE61218 (Figure 2b), and GSE53819 (Figure 2c) were generated. By cross-analyzing the three datasets, eight co-upregulated genes (Figure 3a) and 10 co-downregulated genes (Figure 3b) were identified, and a scale Wayne diagram was used to visualize the data.



Figure 1. Volcano maps of DEGs in GSE40290 (a), GSE61218 (b), and GSE53819 (c).

2.2. GO and KEGG Pathway Enrichment Analysis of Upregulated DEGs

We used GO and KEGG pathway enrichment analyses to characterize the functional effects of each dataset and co-DEGs in the three datasets. The most abundant GO terms in the upregulated DEGs of GSE40290 included signal transduction, cell adhesion, extracellular matrix organization, and so on (Figure 4a). The most abundant GO terms in the upregulated DEGs of GSE61218 included cell division, signal transduction, cell cycle, and so on (Figure 4b). The most abundant GO terms in the upregulated signal transduction, positive regulation of transcription from RNA polymerase II promoter, cell adhesion, and so on (Figure 4c). The most abundant GO terms in the co-upregulated DEGs in the three datasets included inflammatory response, extracellular matrix organization, cell-cell signaling, and so on (Figure 4d).

KEGG analysis results showed that the upregulated DEGs of GSE40290 were significantly enriched in pathways in cancer, neuroactive ligand-receptor interaction, protein digestion and absorption, and so on (Figure 5a); upregulated DEGs of GSE61218 were significantly enriched in the cell cycle, pathways in cancer, PI3K-Akt signaling pathway, and so on (Figure 5b); upregulated DEGs of GSE53819 were significantly enriched in cytokine-cytokine receptor interaction, IL-17 signaling pathway, human papillomavirus infection,

and so on (Figure 5c); and co-upregulated DEGs in the three datasets were significantly enriched in the IL-17 signaling pathway, rheumatoid arthritis, viral protein interaction with cytokine and cytokine receptor, and so on (Figure 5d).



GSE40290

Figure 2. Cont.



GSE53819

Figure 2. Heat maps of the top 20 DEGs in GSE40290 (a), GSE61218 (b), and GSE53819 (c).



Figure 3. Wayne diagram of the upregulated DEGs (a) and downregulated DEGs (b).

2.3. GO and KEGG Pathway Enrichment Analyses of Downregulated DEGs

The most abundant GO terms in the downregulated DEGs of GSE40290 included immune response, innate immune response, cell adhesion, and so on (Figure 6a). The most abundant GO terms in the downregulated DEGs of GSE61218 included cilium movement, cilium assembly, spermatogenesis, and so on (Figure 6b). The most abundant GO terms in the downregulated DEGs of GSE53819 included cell adhesion, cilium assembly, cilium movement, and so on (Figure 6c). The most abundant GO terms in the co-downregulated DEGs in the three datasets included immune response, activation of GTPase activity, and cilium assembly (Figure 6d).



Figure 4. GO analyses of upregulated DEGs in GSE40290 (a), GSE61218 (b), GSE53819 (c), and co-upregulated DEGs (d) in the three datasets.



Figure 5. KEGG analyses of the upregulated DEGs in GSE40290 (**a**), GSE61218 (**b**), GSE53819 (**c**), and co-upregulated DEGs (**d**) in the three datasets.



Figure 6. GO analyses of downregulated DEGs in GSE40290 (**a**), GSE61218 (**b**), GSE53819 (**c**), and co-downregulated DEGs (**d**) in the three datasets.

KEGG analysis results showed that the downregulated DEGs of GSE40290 were significantly enriched in amyotrophic lateral sclerosis, drug metabolism-cytochrome P450, hematopoietic cell lineage, and so on (Figure 7a); downregulated DEGs of GSE61218 were significantly enriched in cytokine-cytokine receptor interaction, pathways of neurodegeneration-multiple diseases, drug metabolism-cytochrome P450, and so on (Figure 7b); downregulated DEGs of GSE53819 were significantly enriched in hematopoietic cell lineage, chemokine signaling pathway, cytokine-cytokine receptor interaction, and so on (Figure 7c); and co-downregulated DEGs in the three datasets were most significantly enriched in the B-cell receptor signaling pathway and hematopoietic cell lineage (Figure 7d).



Figure 7. KEGG analyses of downregulated DEGs in GSE40290 (**a**), GSE61218 (**b**), GSE53819 (**c**), and co-downregulated DEGs (**d**) in the three datasets.

2.4. PPI Network Construction of Co-DEGs

We submitted eight co-upregulated DEGs and 10 co-downregulated DEGs in GSE40290, GSE61218, and GSE53819 to the STRING database for PPI analysis to identify hub genes. DEGs with connectivity greater than eight were selected as the hub genes, and four hub genes were identified, of which prostaglandin-endoperoxide synthase 2 (PTGS2) had a connectivity of 10; the degree of chemokine (C-C motif) ligand 21 (CCL21) was 9, degree of matrix metalloproteinase (MMP) 1 was 8, and degree of MMP3 was 8 (Figure 8).

2.5. Expression of Selected Hub Genes in Tumor Tissues

The UALCAN database was used for the analysis of the TCGA head and neck squamous cancer (HNSC) dataset. We found that the PTGS2 showed higher expression in tumor tissues (Figure 9a). CCL21 did not show differential expression in normal and tumor tissues, and the expression of MMP1 (Figure 9c) and MMP3 (Figure 9d) showed higher expression in tumor tissues. Finally, we confirmed the expression of PTGS2 (Figure 10a) and MMP3 (Figure 10c) in various cancer tissues and high expression of PTGS2 (Figure 10b) and MMP3 (Figure 10d) in HNSC through Human Protein Atlas (HPA) database.



Figure 8. PPI network for co-DEGs in the three datasets.



Figure 9. Cont.



Figure 9. Different expressions of PTGS2 (a), CCL21 (b), MMP1 (c), and MMP3 (d) between normal and tumor tissues, ***: p < 0.001.



Figure 10. Expressions of PTGS2 in various cancer tissues (**a**), PTGS2 shows high staining in HSNC (**b**), expressions of MMP3 in various cancer tissues (**c**), and MMP3 also shows strong staining in HSNC (**d**).

3. Discussion

Various cancers, including NPC, are among the major causes of human death worldwide. Globalization and an increase in various risk factors may further aggravate this situation. Bioinformatics methods are rapidly being developed for the analysis of biological data, especially in the analysis of large datasets, and have become an area of interest for researchers. Transforming biological data into knowledge through bioinformatics methods for study and analysis is more time-saving, efficient, and cost-effective than traditional methods [10]. In this study, we analyzed three datasets (GSE40290, GSE61218, and GSE53819) related to NPC using microarray data and identified 18 co-DEGs.

GO enrichment analysis showed that the most abundant GO terms in the co-upregulated DEGs in the three datasets included inflammatory response, extracellular matrix organization, and cell-cell signaling. Li et al. showed that the EBV M81 strain isolated from NPC-induced chronic inflammation in its target cells resulted in an increase in virus production. They explained the relationship between M81 virus replication and chemokines involved in inflammation and carcinogenesis [11]. Consistent with our results, inflammatory response was significantly enriched in the co-upregulated DEGs in their study. A clinical study that included 30 patients with NPC and 20 controls found that the process of tumor invasion and metastasis can be effectively reduced by controlling the activity of MMPs and extracellular matrix components [12]. This indicates that extracellular matrix components may play a promoting role in the development of NPC, which is consistent with our results. Exosomal microscopic RNAs from cancer cells play a key role in mediating cell-cell signaling and tumor microenvironment crosstalk. Lu et al. identified the inhibitory role of tumor-derived, exosome-related miR-9 in NPC tumorigenesis. In our results, the upregulated GO terms included cell-cell signaling, which is also similar to the result obtained by Lu et al. [13]. KEGG analysis results showed that the co-upregulated DEGs in the three datasets were significantly enriched in the IL-17 signaling pathway, rheumatoid arthritis and so on. The results of Wang et al. strongly revealed that IL-17 could activate the p38-NF-KB signaling pathway and promote the migration and invasion of NPC cells [14], which is consistent with our finding that co-upregulated DEGs were most enriched in the IL-17 signaling pathway. It has been pointed out that in rheumatoid arthritis, the response of antibodies to EBV induced cell antigens is significantly higher than that of healthy individuals [15], which is also consistent with our finding.

The most abundant GO terms in the co-downregulated DEGs in the three datasets were immune response, activation of GTPase activity, and cilium assembly. NPC tumorigenesis is significantly associated with genetic susceptibility. Recent epidemiological and large-scale genome-wide association studies have demonstrated an association between HLA class I genes and the risk of NPC. HLA class I gene coding is used to initiate the host immune response against malignant cells, but studies have found that high-risk people with several specific HLA haplotypes have low efficiency in the immune response to persistent EBV infection [16], which is consistent with our results. Jiang et al. showed that low GTPase expression is related to an increase in signal transduction, cell movement, and metastatic behavior of NPC cells [17]. This is consistent with the finding of our study that GO terms enriched in the co-downregulated DEGs included activation of GTPase activity. A previous study analyzing gene expression data from NPC and non-NPC nasopharyngeal tissues through a comprehensive pathway showed that the loss of function of the axonemal dynein complex in patients with NPC leads to impaired ciliary function, which in turn leads to poor mucociliary clearance and respiratory tract infection [18], which is also consistent with our results. The co-downregulated DEGs in the three datasets were most significantly enriched in the B-cell receptor signaling pathway and hematopoietic cell lineage. Morrison et al. showed that LMP2A expressed in most EBV-related tumors, including NPC, maintains virus latency by blocking the activation and signaling of B-cell receptors [19], which is consistent with our finding that co-downregulated DEGs were most enriched in the B-cell receptor signaling pathway. An article studying new aberrant methylation, differentially expressed genes and pathways in NPC pointed out that the hypermethylation/low-expression genes

significantly enriched in hematopoietic cell lineage [20], which is also consistent with our finding.

Based on the PPI network, we screened four genes with the highest node degrees, including PTGS2, CCL21, MMP1, and MMP3. HNSC develops from the mucous lining of upper respiratory and digestive tract, including nasal cavity, paranasal sinus, oropharynx, larynx, and so on. However, NPC is a specific entity different from HNSC, the disease behavior of NPC is different from HNSC, and the treatment strategy is also different [21]. Since there is no separate NPC dataset in the TCGA database, we can only use the HNSC dataset for analysis. Our analysis using the HNSC dataset in the TCGA database revealed that the expression of PTGS2, MMP1, and MMP3 in tumor tissues was significantly higher. Through HPA database, we also found that PTGS2 and MMP3 are expressed in various cancers, including head and neck cancers. The results of antibody staining showed that PTGS2 and MMP3 were strongly expressed in HSNC. A recent meta-analysis identified the upregulation of PTGS2, MMP1, and MMP3 in NPC tissues, shows that the maladjustment of nasal epithelial barrier and maladjusted immune response are the key components in the pathogenesis of NPC [22]. Another previous meta-analysis found that the overexpression of PTGS2 was significantly associated with a low survival rate in patients with NPC [23]. A study detected high expression of PTGS2 in patients with NPC and distant metastasis and showed that PTGS2 was related to the migration and invasion of NPC cells, in addition to the low survival rate of patients with NPC [24]. A study involving 56 normal people and 114 patients with NPC was conducted to explore the correlation between PTGS2 gene polymorphism and the occurrence of NPC [25]. It was found that PTGS2 gene polymorphism was related to the susceptibility of NPC, and both smoking and EBV infection, which are the main risk factors of NPC, can affect PTGS2 gene polymorphism. Some studies have confirmed that the upregulation of MMP1 is related to lymph node metastasis in NPC [26]. Additionally, studies have shown that MMP1 is significantly associated with the risk of NPC [27]. Song et al. confirmed the upregulation of MMP1 in NPC tissues and cell lines by RT-qPCR and western blotting and found that knockdown of the MMP1 gene significantly inhibited cell proliferation and enhanced apoptosis [28]. A study that detected the mRNA and protein levels of MMP3 in NPC tissues and cells found that the concentration and enzymatic activity of MMP3 in the NPC group were much higher [29]. Another study showed that the overexpression of MMP3 in NPC epithelial cells increased EBV-induced epithelial cell migration and invasion in an in vitro cell model [30]. One study is aimed at analyzing the co-deregulated genes and their transcriptional regulators in lung cancer [31]. They used a Connectivity Map to find putative repurposing drugs for selected hub genes. Although we also tried to discover the putative repurposing drugs by the same method, the results were not satisfactory. In future research, through more data and deeper research, we plan to complete the unfinished research of discovering the putative repurposing drugs.

4. Materials and Methods

4.1. Microarray Data

The NCBI-GEO is a public database (https://www.ncbi.nlm.nih.gov/geo/) (accessed on 28 September 2022). By searching for keywords, such as nasopharyngeal carcinogen and Homo sapiens, we obtained three datasets, GSE40290 dataset including 25 NPC tissues and 8 normal tissues, GSE61218 dataset including 10 NPC tissues and 6 normal tissues, and GSE53819 dataset including 18 NPC tissues and 18 normal tissues, which could be downloaded and analyzed by GEO2R.

4.2. Identification of DEGs and Data Visualization

We analyzed DEGs between NPC and normal tissues in the GSE40290, GSE61218, and GSE53819 datasets using the GEO2R tool. Volcano and heat maps drawn in each dataset were obtained from the hiplot database (https://hiplot-academic.com/) (accessed on 29 September 2022) [32].

4.3. GO and KEGG Pathway Enrichment Analysis of Up- and Downregulated DEGs

DAVID [33] is a web server for gene lists, functional enrichment analysis, and functional annotation (https://david.ncifcrf.gov/) (accessed on 29 September 2022). We used the latest version of the DAVID database (version 7.0) for GO and KEGG pathway enrichment analyses of the upregulated and downregulated DEGs.

4.4. PPI Network Construction of Up- and Downregulated DEGs

STRING [34] is an online resource database used to obtain protein association networks (https://string-db.org/) (accessed on 29 September 2022). The content in the database is precomputed and can be downloaded separately by users. We used the 11.5 version of STRING for PPI analysis.

4.5. Analyzing the Expression of Hub Genes in Tumor

UALCAN [35] data portal is an interactive network resource (http://ualcan.path.uab. edu/) (accessed on 29 September 2022). The Cancer Genome Atlas (TCGA) transcriptome and clinical patient data were used to study the differential expression of hub genes in normal and tumor tissues through this data portal. And through HPA (https://www.proteinatlas.org/) (accessed on 29 November 2022), protein related data were obtained.

5. Conclusions

In conclusion, our analysis identified hub genes and signaling pathways associated with NPC. This provides information for exploring the pathogenesis, identifying molecular targets, and clarifying the biological pathways of NPC. However, further experiments are needed to verify and explore the functions of these genes. Authors should discuss the results and how they can be interpreted from the perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

Author Contributions: Writing-original draft preparation; corrections after review: J.T.; data curation: J.P.; methodology: M.H.; review and editing: T.H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program of the National Research Foundation of Korea, funded by the Ministry of Science and Technology and the Ministry of Science, ICT and Future Planning (2017R1A2B2003575 and NRF-2020R1A2C1006398); the Ministry of Science and ICT, Korea, under the ICT Creative Consilience program (IITP-2023-2020-0-01819), supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP); and the Korea Health Technology R&D Project (HI17C0387, HR22C1302) through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare. This research was supported by a Korean University grant and a grant from the Korea University Medical Center and Anam Hospital in Seoul, Republic of Korea.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in GEO datasets.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Chen, Y.P.; Chan, A.T.C.; Le, Q.T.; Blanchard, P.; Sun, Y.; Ma, J. Nasopharyngeal carcinoma. *Lancet* 2019, 394, 64–80. [CrossRef] [PubMed]
- Lee, H.M.; Okuda, K.S.; González, F.E.; Patel, V. Current Perspectives on Nasopharyngeal Carcinoma. Adv. Exp. Med. Biol. 2019, 1164, 11–34. [CrossRef] [PubMed]
- 3. Chua, M.L.K.; Wee, J.T.S.; Hui, E.P.; Chan, A.T.C. Nasopharyngeal carcinoma. Lancet 2016, 387, 1012–1024. [CrossRef]
- Chua, M.L.; Sun, Y.; Supiot, S. Advances in nasopharyngeal carcinoma-"West meets East". Br. J. Radiol. 2019, 92, 20199004. [CrossRef] [PubMed]

- 5. Guan, S.; Wei, J.; Huang, L.; Wu, L. Chemotherapy and chemo-resistance in nasopharyngeal carcinoma. *Eur. J. Med. Chem.* 2020, 207, 112758. [CrossRef]
- 6. Tao, Z.; Shi, A.; Li, R.; Wang, Y.; Wang, X.; Zhao, J. Microarray bioinformatics in cancer—A review. J. Buon. 2017, 22, 838–843.
- Song, Y.; Feng, T.; Cao, W.; Yu, H.; Zhang, Z. Identification of Key Genes in Nasopharyngeal Carcinoma Based on Bioinformatics Analysis. *Comput. Intell. Neurosci.* 2022, 2022, 9022700. [CrossRef]
- Yue, H.; Zhu, H.; Luo, D.; Du, Q.; Xie, Y.; Huang, S.; Liu, W. Differentially Expressed Genes in Nasopharyngeal Carcinoma Tissues and Their Correlation with Recurrence and Metastasis of Nasopharyngeal Carcinoma. *Comput. Math. Methods Med.* 2022, 2022, 1941412. [CrossRef]
- 9. Clough, E.; Barrett, T. The Gene Expression Omnibus Database. Methods Mol. Biol. 2016, 1418, 93–110. [CrossRef]
- 10. Fu, Y.; Ling, Z.; Arabnia, H.; Deng, Y. Current trend and development in bioinformatics research. *BMC Bioinform.* **2020**, *21*, 538. [CrossRef]
- Li, Z.; Tsai, M.H.; Shumilov, A.; Baccianti, F.; Tsao, S.W.; Poirey, R.; Delecluse, H.J. Epstein-Barr virus ncRNA from a nasopharyngeal carcinoma induces an inflammatory response that promotes virus production. *Nat. Microbiol.* 2019, *4*, 2475–2486. [CrossRef] [PubMed]
- 12. El Badry, A.A.; El-Fadle, A.A.; El-Balshy, A.L. Tissue inhibitor of matrix metalloproteinase-2 in nasopharyngeal carcinoma. *MedGenMed* **2007**, *9*, 3. [PubMed]
- Lu, J.; Liu, Q.H.; Wang, F.; Tan, J.J.; Deng, Y.Q.; Peng, X.H.; Liu, X.; Zhang, B.; Xu, X.; Li, X.P. Exosomal miR-9 inhibits angiogenesis by targeting MDK and regulating PDK/AKT pathway in nasopharyngeal carcinoma. *J. Exp. Clin. Cancer Res.* 2018, 37, 147. [CrossRef]
- 14. Wang, L.; Ma, R.; Kang, Z.; Zhang, Y.; Ding, H.; Guo, W.; Gao, Q.; Xu, M. Effect of IL-17A on the migration and invasion of NPC cells and related mechanisms. *PLoS ONE* **2014**, *9*, e108060. [CrossRef]
- 15. Hasler, F. Recent aspects of Epstein-Barr virus infections. Schweiz. Med. Wochenschr. 1984, 114, 520–523. [PubMed]
- Tsang, C.M.; Lui, V.W.Y.; Bruce, J.P.; Pugh, T.J.; Lo, K.W. Translational genomics of nasopharyngeal cancer. *Semin. Cancer Biol.* 2020, *61*, 84–100. [CrossRef] [PubMed]
- Jiang, C.; Li, L.; Xiang, Y.Q.; Lung, M.L.; Zeng, T.; Lu, J.; Tsao, S.W.; Zeng, M.S.; Yun, J.P.; Kwong, D.L.W.; et al. Epstein-Barr Virus miRNA BART2-5p Promotes Metastasis of Nasopharyngeal Carcinoma by Suppressing RND3. *Cancer Res.* 2020, *80*, 1957–1969. [CrossRef] [PubMed]
- Chin, Y.M.; Tan, L.P.; Abdul Aziz, N.; Mushiroda, T.; Kubo, M.; Mohd Kornain, N.K.; Tan, G.W.; Khoo, A.S.; Krishnan, G.; Pua, K.C.; et al. Integrated pathway analysis of nasopharyngeal carcinoma implicates the axonemal dynein complex in the Malaysian cohort. *Int. J. Cancer* 2016, 139, 1731–1739. [CrossRef]
- 19. Morrison, J.A.; Raab-Traub, N. Roles of the ITAM and PY motifs of Epstein-Barr virus latent membrane protein 2A in the inhibition of epithelial cell differentiation and activation of {beta}-catenin signaling. *J. Virol.* **2005**, *79*, 2375–2382. [CrossRef]
- 20. Han, B.; Yang, X.; Zhang, P.; Zhang, Y.; Tu, Y.; He, Z.; Li, Y.; Yuan, J.; Dong, Y.; Hosseini, D.K.; et al. DNA methylation biomarkers for nasopharyngeal carcinoma. *PLoS ONE* **2020**, *15*, e0230524. [CrossRef]
- Lee, T.F.; Liou, M.H.; Huang, Y.J.; Chao, P.J.; Ting, H.M.; Lee, H.Y.; Fang, F.M. LASSO NTCP predictors for the incidence of xerostomia in patients with head and neck squamous cell carcinoma and nasopharyngeal carcinoma. *Sci. Rep.* 2014, *4*, 6217. [CrossRef] [PubMed]
- Allen, D.Z.; Aljabban, J.; Silverman, D.; McDermott, S.; Wanner, R.A.; Rohr, M.; Hadley, D.; Panahiazar, M. Meta-Analysis illustrates possible role of lipopolysaccharide (LPS)-induced tissue injury in nasopharyngeal carcinoma (NPC) pathogenesis. *PLoS ONE* 2021, 16, e0258187. [CrossRef] [PubMed]
- Sim, C.C.; Sim, E.U. Over-expression of cyclo-oxygenase-2 predicts poor survival of patients with nasopharyngeal carcinoma: A meta-analysis. J. Laryngol. Otol. 2020, 134, 338–343. [CrossRef]
- Zhu, Y.; Shi, C.; Zeng, L.; Liu, G.; Jiang, W.; Zhang, X.; Chen, S.; Guo, J.; Jian, X.; Ouyang, J.; et al. High COX-2 expression in cancer-associated fibiroblasts contributes to poor survival and promotes migration and invasiveness in nasopharyngeal carcinoma. *Mol. Carcinog.* 2020, *59*, 265–280. [CrossRef] [PubMed]
- 25. Sun, N.; Wang, X.Q. Correlation between COX-2 gene polymorphism and susceptibility to nasopharyngeal carcinoma. *Eur. Rev. Med. Pharmacol. Sci.* **2019**, 23, 5770–5778. [CrossRef] [PubMed]
- 26. Ben Nasr, H.; Chahed, K.; Remadi, S.; Zakhama, A.; Chouchane, L. Expression and clinical significance of latent membrane protein-1, matrix metalloproteinase-1 and Ets-1 transcription factor in tunisian nasopharyngeal carcinoma patients. *Arch. Med. Res.* **2009**, *40*, 196–203. [CrossRef]
- Tsai, C.W.; Chang, W.S.; Gong, C.L.; Shih, L.C.; Chen, L.Y.; Lin, E.Y.; Li, H.T.; Yen, S.T.; Wu, C.N.; Bau, D.T. Contribution of Matrix Metallopeptidase-1 Genotypes, Smoking, Alcohol Drinking and Areca Chewing to Nasopharyngeal Carcinoma Susceptibility. *Anticancer Res.* 2016, *36*, 3335–3340.
- 28. Song, L.; Liu, H.; Liu, Q. Matrix metalloproteinase 1 promotes tumorigenesis and inhibits the sensitivity to 5-fluorouracil of nasopharyngeal carcinoma. *Biomed. Pharmacother.* **2019**, *118*, 109120. [CrossRef]
- 29. Li, Y.; Feng, Z.; Xing, S.; Liu, W.; Zhang, G. Combination of serum matrix metalloproteinase-3 activity and EBV antibodies improves the diagnostic performance of nasopharyngeal carcinoma. *J. Cancer* **2020**, *11*, 6009–6018. [CrossRef]
- Lan, Y.Y.; Yeh, T.H.; Lin, W.H.; Wu, S.Y.; Lai, H.C.; Chang, F.H.; Takada, K.; Chang, Y. Epstein-Barr virus Zta upregulates matrix metalloproteinases 3 and 9 that synergistically promote cell invasion in vitro. *PLoS ONE* 2013, 8, e56121. [CrossRef]

- 31. Chatziantoniou, A.; Zaravinos, A. Signatures of Co-Deregulated Genes and Their Transcriptional Regulators in Lung Cancer. *Int. J. Mol. Sci.* **2022**, *23*, 10933. [CrossRef] [PubMed]
- 32. Li, J.; Miao, B.; Wang, S.; Dong, W.; Xu, H.; Si, C.; Wang, W.; Duan, S.; Lou, J.; Bao, Z.; et al. Hiplot: A comprehensive and easy-to-use web service for boosting publication-ready biomedical data visualization. *Brief. Bioinform.* **2022**, *23*, bbac261. [CrossRef]
- Sherman, B.T.; Hao, M.; Qiu, J.; Jiao, X.; Baseler, M.W.; Lane, H.C.; Imamichi, T.; Chang, W. DAVID: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids. Res.* 2022, 50, W216–W221. [CrossRef] [PubMed]
- 34. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids. Res.* 2021, *49*, D605–D612. [CrossRef] [PubMed]
- 35. Chandrashekar, D.S.; Bashel, B.; Balasubramanya, S.A.H.; Creighton, C.J.; Ponce-Rodriguez, I.; Chakravarthi, B.; Varambally, S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 2017, *19*, 649–658. [CrossRef] [PubMed]





Communication PSSNet—An Accurate Super-Secondary Structure for Protein Segmentation

Denis V. Petrovsky, Vladimir R. Rudnev, Kirill S. Nikolsky, Liudmila I. Kulikova, Kristina M. Malsagova *D, Arthur T. Kopylov D and Anna L. Kaysheva D

Biobanking Group, Branch of Institute of Biomedical Chemistry "Scientific and Education Center", 109028 Moscow, Russia

* Correspondence: kristina.malsagova86@gmail.com; Tel.: +7-499-764-98-78

Abstract: A super-secondary structure (SSS) is a spatially unique ensemble of secondary structural elements that determine the three-dimensional shape of a protein and its function, rendering SSSs attractive as folding cores. Understanding known types of SSSs is important for developing a deeper understanding of the mechanisms of protein folding. Here, we propose a universal PSSNet machine-learning method for SSS recognition and segmentation. For various types of SSS segmentation, this method uses key characteristics of SSS geometry, including the lengths of secondary structural elements and the distances between them, torsion angles, spatial positions of C α atoms, and primary sequences. Using four types of SSSs ($\beta\alpha\beta$ -unit, α -hairpin, β -hairpin, $\alpha\alpha$ -corner), we showed that extensive SSS sets could be reliably selected from the Protein Data Bank and AlphaFold 2.0 database of protein structures.

Keywords: super-secondary structure; data bank; AlphaFold 2.0; graph neural network; machine learning; protein features

1. Introduction

Protein folding mechanisms have fascinated scientists for a half of a century [1–3]. According to the "nucleation–condensation" model of protein folding, self-folding proteins, such as molecular chaperones [4], that do not participate in the protein machinery, become unstructured tangles immediately after translation. Folding nuclei (a time-limiting stage) are formed and condensed in coils, and the process is completed by spontaneous packing into a native three-dimensional structure [5–7]. In relation to this concept, attention has been focused toward simple motifs such as super-secondary structures (SSSs) that comprise several secondary structure elements with unique and compact folding of a polypeptide chain. Super-secondary structures serve as a bridge between the secondary and tertiary structure of a protein and probably are autonomously stable (i.e., stable outside the protein globule) [8,9].

The use of SSSs to solve biomedical problems is rather desirable, as the alpha-helical and beta-hairpin types of SSSs can serve as initial unique structures for the construction of protein epitope mimetics (PEMs) [10,11]. These PEMs mimic the structural and conformational properties of their target epitopes (SSS), as well as their biological activity (protein–protein and protein–nucleic acid interactions). It is possible to optimize biological activity to maintain antimicrobial activity, for example, by transferring an epitope from a recombinant to a synthetic scaffold [11].

In previous studies, we reported the possibility of studying SSSs in aberrant protein forms caused by post-translational modifications (PTMs). We observed that PTMs that have been detected in patients with various types of cancer are frequently localized in the SSS (alpha-structural motifs, beta-hairpins) [12]. So far, it is obvious that a comprehensive study of the known SSS types is essential for deeper insights of protein folding mechanisms and to solve some challenges in biomedical research [9].

Citation: Petrovsky, D.V.; Rudnev, V.R.; Nikolsky, K.S.; Kulikova, L.I.; Malsagova, K.M.; Kopylov, A.T.; Kaysheva, A.L. PSSNet—An Accurate Super-Secondary Structure for Protein Segmentation. *Int. J. Mol. Sci.* 2022, 23, 14813. https://doi.org/ 10.3390/ijms232314813

Academic Editor: Wajid Zaman

Received: 28 October 2022 Accepted: 24 November 2022 Published: 26 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Predicting the secondary and supersecondary structures of proteins by their 3Dstructures (PDB, AlfaFold) is becoming a top priority in structural biology research. Numerous approaches for prediction are currently known but the most commonly used are based on (1) probabilistic models, such as kernel density estimation (KDE) [13,14] and naïve Bayes [15]; (2) linear classifiers, such as support vector machines (SVM) [16–18]; and (3) machine learning methods [9,19]. The performance of the first two approaches is limited by the huge amounts of data with relatively low classification and semantic segmentation accuracy (60–75%) [16,20].

Neural networks (NN) have recently been applied to the problem of structures classification and segmentation. Neural networks are typically designed to classify and/or predict one or two types of SSSs, though several NN-utilizing methods are now capable of predicting β -hairpin and $\beta \alpha \beta$ -units (StackSSSPred) [21–23]. The following main groups of machine learning models are most widely used:

- 1. Models based on the sequence-to-sequence architecture, where the protein structure is considered as a sequence of amino acids with the main characteristics of their localization, i.e., contact map. Featured sequences are processed using a group of recurrent layers [24].
- 2. Models based on 3D-CloudSegmentation, in which each atom of a molecule is represented as a point in a 3D space. PointNet, PointNet++, and dynamic graph CNN (DGCNN) architectures [25] are used to segment and classify structures.
- 3. Models based on the representation of a protein molecule as a 3D volumetric object (protein voxelization) with subsequent processing by 3D-Convolution family networks [26].
- 4. Models based on the representation of a protein molecule as a graph with subsequent processing by graph neural networks (CGNs) [27].

Here, we present a new approach to classify different types of SSSs, specifically $\beta \alpha \beta$ -unit, α -hairpin, β -hairpin, and $\alpha \alpha$ -corner, and the approach was tested on standard format files extracted from the public Protein Data Bank (PDB) [9]. The neural network PSSNet (Protein Secondary Structure Segmentation) was realized on a new deep learning architecture that uses the integrative synergy of CGN, convolutional neural networks (CNN), and (bidirectional) recurrent neural network (RNN) predictions. The proposed architecture achieves an accuracy of 84% and endorses a wide range of valuable annotations for over 1.9 million SSSs available in the open-access knowledge base at https://psskb.org/ (accessed on 28 October 2022). In addition to the secondary structure prediction, PSSNet can also be applied for the prediction of free energy, solvent availability, contact maps, and searching for stable protein structures.

2. Results and Discussion

2.1. SSS Segmentation Using the PSSNet Model

This model was the basis for the filling of gaps in the open knowledge base of SSSs (available at https://psskb.org/, accessed on 28 October 2022). After training the model, we applied it to complete dataset maintained in the PDB and AlphaFold. The results were selectively assessed by expert researchers, entered into a database, and finalized as publicly available. The number of SSSs defined by the model are listed in Table 1.

Table 1. Number of SSSs recognized in open knowledge databases (PDB and AlphaFold).

_					
	SSS	PDB (185,469 Structures)	AlphaFold (2021) (360,000 Structures)		
	βαβ-unit	461,336	233,882		
	α-hairpin	390,965	563,946		
	β-hairpin	360,845	280,181		
	αα-corner	5977	8153		

The model was built according to the proposed architecture and combined high accuracy and performance. We assessed the performance of the model engine by comparing its quality with that of networks with distinct architectures, i.e., CurveNet [28] and DGCNN36 networks (Table 2). Such networks are among the top 10 utilized in "3D point cloud" classification and segmentation [29]. Training and evaluation of the results were carried out on the same datasets. Plots of the loss function and IOU versus iterations are provided in Supplementary Materials (model evaluation metrics section).

SSS	PSSNet		CurveNet		DGCNN	
	Train	Val	Train	Val	Train	Val
βαβ-unit	0.928	0.894	0.742	0.697	0.691	0.656
α-hairpin	0.964	0.957	0.814	0.795	0.731	0.688
β-hairpin	0.998	0.983	0.845	0.833	0.749	0.711
αα-corner	0.933	0.991	0.781	0.732	0.621	0.571

Table 2. Comparison between our proposed and other models (mean IOU).

The predictive power of a machine learning model is mainly determined by its feature representation and feature extraction algorithms. The models considered in our comparative experiment operate only with the 3D coordinates of protein atoms, however these features were barely enough to provide acceptable recognition accuracy (Table 2). In contrast, our model operates with an extended set of specific structural features, encompassing torsion angles, the spatial positions of atoms in the amino acid sequence, and the primary protein sequence. The convolutional filters of our neural network blocks capture and generalize the local geometric features of the protein sequence well, and the subsequent blocks of Bi-GRUs capture the features of the global feature context.

2.2. Practical Evaluation of the Model: Key Issues

The proposed model was empirically tested to assess its accuracy and discover key objections that might arise from structure segmentation and classification. A random sample of 5000 SSS structures from PDB and AlphaFold was consolidated, and topology compliance with the studied motifs was examined.

The main issues that arose from the classification and segmentation of structural motifs were as follows (Figure 1):

- (a) Capture of excess structure sections;
- (b) Breakages in structure element links;
- (c) Incorrect definition of corners (typical for the $\alpha\alpha$ -corner motif (70°–90°)).

Despite the results of the model validation being quite satisfying (IoU = 0.92), the estimated accuracy ranged between 0.83-0.85. Thus, we established several problems specific to the topology of a specific type of SSS.

- a. Capturing extra sections is a prerequisite for SSS with α -helix elements, and it usually manifests if the distance between the last and first C α atoms of the first and the second structures, respectively, is 9.7 Å. However, the network at the output from the GRU layer generates a feature map that captures both structures (Figure 1e). This issue was fixed by reducing the number of neighbors in the knn-graph, when generating features, or by producing a sufficiently larger sample of such structures and subsequently retraining the model. The identification and extraction of such elements from the PSSKB database are currently ongoing.
- b. Breakages in structures generally occurred in low-resolution (>4.0 Å) PDB files, but the proportion was insignificant and relatively narrow compared to the total size of the consolidated databank.
- c. The network identified curved helices with a large angle of inflection as two elements with incorrectly defined angles for $\alpha\alpha$ -corner structures. Rigorous analysis revealed that the issue can be effectively resolved only if we performed retraining of the model



on a meaningfully larger representative sample that covers all such elements; the retraining and sample collecting are currently in progress.

e) An example of a heat map for the problem of capturing extra structure elements



Figure 1. Evaluation of the model on empirical consolidated datasets (random sampling of 5000 elements for each type of SSS) of (a) $\beta \alpha \beta$ -unit, (b) β -hairpin, (c) α -hairpin and (d) $\alpha \alpha$ -corner, (e) heat map for the problem of extra α -helix.

Despite the difference in accuracy between the actual and validation datasets, the model managed SSS segmentation and classification tasks well. The difference in accuracy suggests relatively high folding variability among the structures downloaded from PDB and AlphaFold. Hence, the training dataset must be sufficiently extended, especially in terms of negative examples, to improve the accuracy of segmentation and classification.

The composition of super-secondary structures is simple combinations of α -helices, β -hairpins, and short loops with a well-defined hydrophobic core involved in SSS stabilization. The loop–helix, β -hairpin, and Greek key motifs are prominent representatives of SSSs [30–32]. Characterization of such structures allows us to collect a catalog of autonomously stable protein motifs and archetypes [30,31]. These structures also serve as promising objects of study in protein physics (the study of folding), bioengineering (the development of peptide mimetics), and biomedicine (the study of conformational changes in aberrant forms of proteins compared with intact forms).

Medical proteomic research is mainly focused on the extensive study of the molecular basis of a disease associated with the arrival of aberrant forms of proteins that are regularly not found under normal (healthy) conditions. Aberrant proteins are frequently caused by genetic polymorphisms, alternative splicing, and PTMs [33,34]. Such structural changes associated with the disease can be localized in different types of SSSs. Numerous aberrant forms of proteins are fraught with dire structural changes, including isomers of beta-amyloid in Alzheimer's disease [35], splice isoforms of osteopontins b and c in prostate cancer [36], amino acid substitutions in protein C7 in type II diabetes [37], and PTMs of proteins in oncological diseases [12].

Here, we present a new approach to frame the problem of SSS recognition and segmentation based on the geometric characteristics of structures and spatial relationships within a protein sequence. The main advantage of our method is a low requirement for computing sources. We used a standard personal desktop computer with a typical GeForce GTX 1650—4Gb video card for training and data processing in the PDB and AlphaFold2.0 databases. We also operated the PSSNet model with high recognition accuracy (mean IoU > 0.84; F1 > 0.08) and annotation capability of >1.9 million SSSs of $\beta \alpha \beta$ -unit hairpin, β -hairpin, and $\alpha \alpha$ -corner. This opens up wide margins for investigation on the PSSKB resource. The model does not require a large training set, since sets of 2000 specimens were used to train the model. Plots of GPU memory usage as a function of protein sequence length can be found in Supplementary Materials (GPU memory usage section).

A distinctive feature of our model is its ability to recognize and segment the SSSs within a protein sequence of arbitrary length, i.e., regardless of the sequence length. The model can operate directly with any file in PDB format, including those with low data quality, poor resolution, and sparse protein sequences. The only limitation of the model is the amount of graphics processing unit (GPU) memory. Likewise, most current models focus on recognition of only one certain type of structure and work only with a few prepared datasets, resulting in relatively low numbers of recognized structures within the range of thousands to tens of thousands.

The architecture of the proposed method is powered by a comprehensive combination of CGNs, CNNs, Bi-GRUs, self-attention, and multi-head attention mechanisms, which encourage network flexibility and easy adaptation to solve problems in structural biology and bioinformatics. Primary examinations have shown that a model with minimal modifications can predict the structural alphabet based on geometric characteristics for differentiable molecular modeling problems. A subsequent investigation will target this and other issues.

3. Materials and Methods

3.1. Data Preparation

Training, test, and validation datasets downloaded from the Protein Data Bank were represented by the following types of SSSs: a $\beta \alpha \beta$ -motif (beta-alpha-beta motif), a β -hairpin, an α -hairpin, and an $\alpha \alpha$ -corner (70°–90°).

The datasets were generated using STRIDE, which takes a PDB file as input and returns secondary structure assignments. Thereafter, data were manually curated by a team of experts to ensure compliance with the declared types of SSS. Eventually, the sets of positive and negative examples included almost 2000 and 4000 elements of SSS of each type.

Training and test model datasets are available at https://doi.org/10.6084/m9.figshare.2152 9812.v1 (open access, accessed on 28 October 2022) [38].

The balance between positive and negative examples in packets supplied to the network input was regulated by the software implementation. Before entering the network, the coordinates of atoms x, y, and z were augmented (rotation around the x, y, and z axes at a random angle and the y-axis with random jitter for each point using Gaussian noise with zero-mean value and standard deviation of 0.08). Data augmentation was executed dynamically during the training time for 40% of input structures.

Before entering the network, elements of amino acid sequence (AA codes and 3D coordinates of the corresponding group of atoms (N, C_{α} , C, and O)) were extracted from PDB files.

Ultimately, an array of 3D coordinates was generated to describe the 3D structure of the protein. The final array of coordinates was applied to generate a graph, with each vertex representing a C_{α} -atom in the main protein chain, connected by edges to the 32 nearest C_{α} -atoms (KNN-graph, k = 32). Each edge and vertex of the graph contained scalar and vector features describing the 3D geometry of the protein structure. The method for determining the optimal value of k is described in the Supplementary Materials (determining the optimal value of k (nearest C_{α} atoms)).

3.2. Feature Extraction and Input Encoding

3.2.1. Node-Level Features

The signs of a graph node are described by the following elements (Figure 2):



Output of the Gaussian radial basis function for one edge

Figure 2. Feature extractions from protein sequences. The graph shows the protein structure.

- {sin, cos} \circ { φ , ψ , ω }, where φ , ψ , and ω are the torsion angles calculated for C_{i-1} , N_i , $C\alpha_i$, and C_i and N_{i+1} ;
- Unit vectors of the directions to the $C\alpha$ -atoms in the main chain $(V_{f1} = (C\alpha_{i+1} C\alpha_i))$ and $\vec{V_{r1}} = (C\alpha_{i-1} - C\alpha_i)$;
- Unit direction vectors to the C-atom in the main chain $(V_{f2} = (C_{i+1} C_i))$ and $V_{r2} = (C_{i-1} C_i)$;

- Cosines of the angles between vectors V_{f2} , V_{r2} ;
- The distance between the C-atom in the chain $||C_{i+1} C_{i-1}||$;
- A unit vector that determines the conditional direction of the side chain (direction of the C β atom), $\overrightarrow{Vo}=C\beta_i-C\alpha_i$. This vector is calculated from the tetrahedral representation of the geometry of the N, C α , and C atoms as follows:

$$\overrightarrow{Vo} = \sqrt{\frac{1}{3}} \frac{(a \times b)}{\|a \times b\|} - \sqrt{\frac{2}{3}} \frac{(a+b)}{\|a+b\|}$$

where vectors *a* and *b* are defined as $a = N_i - C\alpha_i b = C_i - C\alpha_i$. This vector, together with the forward and reverse vectors (V_{f1}, V_{r1}) determines the orientation of the amino acid residue in 3D Euclidean space.

• The amino acid sequence is encoded as a sequence of numbers (0–21).

3.2.2. Edge-Level Features

Graph edge features are described by the following elements:

- a unit vector defining the direction between neighboring vertices, $C\alpha_i C\alpha_i$;
- the distance between the vertices of the graph is encoded using Gaussian radial basis functions:

$$\varphi(r) = e^{-(\varepsilon r)^2} \ (r = \|x - x_i\|)$$

For each edge, the distance was encoded with 32 Gaussian functions, with centers uniformly spaced in the range of 0-24 Å. The edge position code (*i*, *j*) was obtained using a sinusoidal encoder, which is widely used in transformer models. This approach to the positional encoding of sequences has been previously described in detail [39].

3.3. Network Architecture

The architecture of our model is based on a combination of the geometric vector perceptron (GVP), graph neural network (GNN), and multi-layer gated recurrent unit (GRU) methods (Figure 3). The network architecture is based on the encoder–decoder principle, which is widely used in classification and segmentation problems. The encoder generates a feature map based on the input data (node position in the graph, local topology, vector, and scalar attributes of the node itself and its neighbors). The decoder extracts information from the feature map and generates classification labels for the graph nodes. The model was implemented using a binary classifier, and a separate model was trained for each SSS in the training set. As data are being accumulated, a multiclass model that works with a variety of structures has to be used in the future.

The GVP elements of the model architecture extract invariant and equivariant features from a combination of scalar and vector representations of geometric features. In addition, the GVP can approximate any continuous rotation and reflection invariant scalar function. The architecture (GVP–GNN) uses GVP modules for feature extraction and the graph convolutional network (GCN) mechanism for message, which passes between graph nodes (messaging), feature aggregation of neighboring nodes and edges, and updates node attachments during a propagation operation [40]. The GVP architecture has been previously described in detail [41].



Figure 3. The architecture of the proposed PSSNet.

A GVP-based neural network was used to predict amino acid sequences based on the geometric characteristics of a protein and PPI (protein–protein interaction). Because proteins are connected in sequential structures, we supplemented the model with bidirectionally controlled recurrent units (Bi-GRUs) to highlight relationships between geometric characteristics [41]. Adding GRU layers to the model significantly increased the predictive accuracy and reduced the amount of time required to train the model. Table 3 shows the architecture of the model and a brief description of the functions of the blocks.

Block	Layer	Description	
	Embedding	Words in AA-sequence using a dense vector form.	
	GVP	Module for learning vector- and scalar-valued functions over geometric vectors and scalars.	
	Norm	Layer normalization for vector features (L2-normalization).	
Encoder	GVPConv (5-layers)	Implements GVP transforms and uses message-passing mechanisms from neighboring nodes and edges to aggregate a function of hidder states and update node embedding at each graph propagation step.	
	Bi-GRU (2-layer module)	Recurrent unit with input and forget gates. The Bi-GRU considers tw separate sequences: from right to left and vice-versa. We considered the sequence of the hidden states of the node features of the graph.	
	Self-attention	This mechanism allows the discovery of connections between elements of the input sequence and the selection of those required for future generations [42]. We considered the sequence of the hidden states of the node features of the graph.	
	GVPConv + Bi-GRU (5-layers)	Decoder block to reconstruct and obtain the graph structure from the encoder's hidden state.	
Decoder	GVP	Last GVP module.	
	Multi-head graph attention	This module has a one-way scalar sigmoid output to predict node labels [43].	

Table 3. Model architecture and implementation details.

The Adam optimizer was applied with a reduced learning rate when the accuracy metric stopped the improvement process (start with a 1×10^{-3} and reduce factor of 0.5). The Dice BCE loss was selected as the loss function, as it combines the Dice loss with the standard binary cross-entropy (BCE) loss, which is generally the default for segmentation models. Combining the two methods allowed for moderate diversity in the loss while improving the stability of the BCE.

3.4. Training and Performance Evaluation

The model training process lasted 24 epochs for each SSS and models were assessed on the validation datasets. During the learning process, the learning rate was changed from 1×10^{-3} to less than 1×10^{-4} in order to reduce the learning rate on the plateau. We used the intersection over union (IoU; also known as the Jaccard index) as the main metric for assessing the quality of model predictions. Values from 0–1 show the extent to which positions of two objects (reference [ground-truth]) predicted by the model coincide according to the following equation:

$$IoU = \frac{Area of Overlap}{Area of Union}$$

We considered the position of SSS in the reference and predicted structures and evaluated the coincidence of their positions. The harmonic mean of the recall and precision metrics (F1) were also evaluated (Table 4).

Table 4. Performance metrics: IoU and F1.

SSS	Mean IoU (Training)	Mean IoU (Validation)	F1
βαβ-unit	0.928	0.894	0.978
α-hairpin	0.964	0.957	0.984
β-hairpin	0.998	0.983	0.991
αα-corner	0.933	0.991	0.994

4. Conclusions

Super-secondary structures are blocks of protein molecules with unique and compact spatial arrangements. Such structures are stable outside the protein globule due to pronounced hydrophobic cores. Structural biology considers SSSs as the nuclei of protein folding and as starting structures when looking for the possible folding pattern of polypeptide chains while modeling protein structures. Our model combines GNN, CNN, and RNN methods and suggests the following advantages:

- small datasets for rapid, efficient learning, and retaining;
- ability to generalize features on a relatively small training set;
- good recognition accuracy (mean IoU > 0.83);
- huge amounts of information (such as that in the PDB and AlphaFold) can be assessed within a reasonable timeframe.

Our model can classify more than 2.3 million SSSs for all protein structures available in the PDB and AlphaFold databases. The reliability and accuracy of the model were demonstrated on four types of SSSs taken from the public Structural Elements Database (PSSKB, https://psskb.org/, accessed on 28 October 2022); however, the model is generic and can be applied to a wider set of SSS types. The assembled set of SSS structures opens up new options for studying the uniqueness and compactness of protein spatial packing and folding nuclei, and can also act as starting structures for searching for possible polypeptide chain folding while modeling protein structures.

Future efforts will target the diversity of SSS types (Greek key, Rossmann fold, etc.) in the segmentation model and replenishing the database. We will also focus on improving annotations and ensuring the quality of SSS presentations. Furthermore, we will generate sufficient information for users with extensive experience in structural biology and new entrants into that. We will also tailor the database to meet the needs of the research community and provide accurate SSS information for future updates.

Supplementary Materials: The supporting information can be downloaded at: https://www.mdpi. com/article/10.3390/ijms232314813/s1. Reference [44] is cited in the supplementary materials.

Author Contributions: D.V.P., V.R.R., A.L.K. and L.I.K., conceived the project; D.V.P., conducted experiments; L.I.K., K.S.N. and V.R.R., generated sets of positive and negative examples; D.V.P., K.M.M., A.L.K. and A.T.K., wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study proceeded within the framework of the Russian Federation Fundamental Research Program for the long-term period of 2021–2030 (Nº 122092200056-9).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The code source designed for super-secondary structures classification (PSSNet) has been deposed to the open-access GitHub resource and is available at the following link: https://github.com/Denis21800/PSSNet.

Acknowledgments: The authors are grateful to A.V. Efimov for the helpful discussions. Equipment at the shared research facilities of HPC computing resources at Lomonosov Moscow State University and the Joint Supercomputer Center of the Russian Academy of Sciences was used for simulations.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Wetlaufer, D.B. Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins. *Proc. Natl. Acad. Sci. USA* **1973**, 70, 697–701. [CrossRef]
- 2. Karplus, M.; Weaver, D.L. Protein-Folding Dynamics. *Nature* **1976**, *260*, 404–406. [CrossRef]
- 3. Anfinsen, C.B. Principles That Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223–230. [CrossRef]
- 4. Hartl, F.U. Molecular Chaperones in Cellular Protein Folding. *Nature* **1996**, *381*, 571–580. [CrossRef]
- 5. Dobson, C.M. Protein Folding and Misfolding. Nature 2003, 426, 884–890. [CrossRef]
- 6. Abkevich, V.I.; Gutin, A.M.; Shakhnovich, E.I. Specific Nucleus as the Transition State for Protein Folding: Evidence from the Lattice Model. *Biochemistry* **1994**, *33*, 10026–10036. [CrossRef]
- 7. Fersht, A.R. Nucleation Mechanisms in Protein Folding. Curr. Opin. Struct. Biol. 1997, 7, 3–9. [CrossRef]
- MacCarthy, E.; Perry, D.; KC, D.B. Advances in Protein Super-Secondary Structure Prediction and Application to Protein Structure Prediction. In *Protein Supersecondary Structures: Methods and Protocols*; Kister, A.E., Ed.; Methods in Molecular Biology; Springer: New York, NY, USA, 2019; pp. 15–45. [CrossRef]
- 9. Rudnev, V.R.; Kulikova, L.I.; Nikolsky, K.S.; Malsagova, K.A.; Kopylov, A.T.; Kaysheva, A.L. Current Approaches in Supersecondary Structures Investigation. *Int. J. Mol. Sci.* 2021, 22, 11879. [CrossRef]
- Robinson, J.A. The Design, Synthesis and Conformation of Some New β-Hairpin Mimetics: Novel Reagents for Drug and Vaccine Discovery. Synlett 2000, 2000, 429–441. [CrossRef]
- Robinson, J.A. β-Hairpin Peptidomimetics: Design, Structures and Biological Activities. Acc. Chem. Res. 2008, 41, 1278–1288. [CrossRef]
- Tikhonov, D.; Kulikova, L.; Kopylov, A.T.; Rudnev, V.; Stepanov, A.; Malsagova, K.; Izotov, A.; Kulikov, D.; Zulkarnaev, A.; Enikeev, D.; et al. Proteomic and Molecular Dynamic Investigations of PTM-Induced Structural Fluctuations in Breast and Ovarian Cancer. *Sci. Rep.* 2021, *11*, 19318. [CrossRef]
- 13. Brownlee, J. A Gentle Introduction to Probability Density Estimation. Machine Learning Mastery. Available online: https://machinelearningmastery.com/probability-density-estimation/ (accessed on 14 February 2022).
- 14. Niranjan Pramanik, N.P. Kernel Density Estimation— Kernel Construction and Bandwidth Optimization using Maximum Likelihood Cross Validation. Analytics Vidhya. Available online: https://medium.com/analytics-vidhya/kernel-density-estimation-kernel-construction-and-bandwidth-optimization-using-maximum-b1dfce127073 (accessed on 14 February 2022).
- 15. Schmidler, S.C.; Liu, J.S.; Brutlag, D.L. Bayesian Segmentation of Protein Secondary Structure. J. Comput. Biol. 2000, 7, 233–248. [CrossRef] [PubMed]
- 16. Sun, L.; Hu, X.; Li, S.; Jiang, Z.; Li, K. Prediction of Complex Super-Secondary Structure Bαβ Motifs Based on Combined Features. *Saudi J. Biol. Sci.* **2016**, *23*, 66–71. [CrossRef]
- 17. Kumar, M.; Bhasin, M.; Natt, N.K.; Raghava, G.P.S. BhairPred: Prediction of Beta-Hairpins in a Protein from Multiple Alignment Information Using ANN and SVM Techniques. *Nucleic Acids Res.* **2005**, *33*, W154–W159. [CrossRef] [PubMed]

- 18. Xia, X.; Longo, L.M.; Sutherland, M.A.; Blaber, M. Evolution of a Protein Folding Nucleus. *Protein Sci.* 2016, 25, 1227–1240. [CrossRef]
- 19. AlQuraishi, M. Machine Learning in Protein Structure Prediction. Curr. Opin. Chem. Biol. 2021, 65, 1–8. [CrossRef]
- 20. Melvin, I.; Ie, E.; Kuang, R.; Weston, J.; Noble, W.S.; Leslie, C. SVM-Fold: A Tool for Discriminative Multi-Class Protein Fold and Superfamily Recognition. *BMC Bioinform.* **2007**, *8*, S2. [CrossRef]
- Flot, M.; Mishra, A.; Kuchi, A.S.; Hoque, M.T. StackSSSPred: A Stacking-Based Prediction of Supersecondary Structure from Sequence. In *Protein Supersecondary Structures: Methods and Protocols*; Kister, A.E., Ed.; Methods in Molecular Biology; Springer: New York, NY, USA, 2019; pp. 101–122. [CrossRef]
- 22. Kuhn, M.; Meiler, J.; Baker, D. Strand-Loop-Strand Motifs: Prediction of Hairpins and Diverging Turns in Proteins. *Proteins* **2004**, 54, 282–288. [CrossRef]
- 23. Cruz, L.; Rao, J.S.; Teplow, D.B.; Urbanc, B. Dynamics of Metastable β-Hairpin Structures in the Folding Nucleus of Amyloid β-Protein. *J. Phys. Chem. B* 2012, *116*, 6311–6325. [CrossRef]
- 24. Li, Z.; Yu, Y. Protein Secondary Structure Prediction Using Cascaded Convolutional and Recurrent Neural Networks. *arXiv* 2016, arXiv:1604.07176[cs, q-bio].
- 25. Kalimeris, A.G.; Emiris, I. Deep Learning on Point Clouds for 3D Protein Classification Based on Secondary Structure. Available online: https://pergamos.lib.uoa.gr/uoa/dl/object/2880834/file.pdf (accessed on 14 February 2022).
- Stepniewska-Dziubinska, M.; Zielenkiewicz, P.; Siedlecki, P. Detection of Protein-Ligand Binding Sites with 3D Segmentation. 2019. Available online: https://www.researchgate.net/publication/332438981_Detection_of_protein-ligand_binding_sites_with_ 3D_segmentation (accessed on 14 February 2022).
- Gligorijević, V.; Renfrew, P.D.; Kosciolek, T.; Leman, J.K.; Berenberg, D.; Vatanen, T.; Chandler, C.; Taylor, B.C.; Fisk, I.M.; Vlamakis, H.; et al. Structure-Based Protein Function Prediction Using Graph Convolutional Networks. *Nat. Commun.* 2021, *12*, 3168. [CrossRef]
- 28. Xiang, T.; Zhang, C.; Song, Y.; Yu, J.; Cai, W. Walk in the Cloud: Learning Curves for Point Clouds Shape Analysis. *arXiv* 2021, arXiv:2105.01288[cs].
- 29. Papers with Code—ModelNet40 Benchmark (3D Point Cloud Classification). Available online: https://paperswithcode.com/ sota/3d-point-cloud-classification-on-modelnet40 (accessed on 14 February 2022).
- Sborgi, L.; Verma, A.; Sadqi, M.; de Alba, E.; Muñoz, V. Protein Folding at Atomic Resolution: Analysis of Autonomously Folding Supersecondary Structure Motifs by Nuclear Magnetic Resonance. In *Protein Supersecondary Structures*; Kister, A.E., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2013; pp. 205–218. [CrossRef]
- 31. Kubelka, J.; Hofrichter, J.; Eaton, W.A. The Protein Folding "Speed Limit". Curr. Opin. Struct. Biol. 2004, 14, 76-88. [CrossRef]
- 32. Muñoz, V. Conformational Dynamics and Ensembles in Protein Folding. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 395–412. [CrossRef]
- Shafi, S.; Singh, A.; Gupta, P.; Chawla, P.A.; Fayaz, F.; Sharma, A.; Pottoo, F.H. Deciphering the Role of Aberrant Protein Post-Translational Modification in the Pathology of Neurodegeneration. CNS Neurol. Disord. Drug Targets 2021, 20, 54–67. [CrossRef]
- 34. Venables, J.P. Aberrant and Alternative Splicing in Cancer. Cancer Res. 2004, 64, 7647–7654. [CrossRef]
- 35. Indeykina, M.I.; Popov, I.A.; Kozin, S.A.; Kononikhin, A.S.; Kharybin, O.N.; Tsvetkov, P.O.; Makarov, A.A.; Nikolaev, E.N. Capabilities of MS for Analytical Quantitative Determination of the Ratio of α- and BAsp7 Isoforms of the Amyloid-β Peptide in Binary Mixtures. *Anal. Chem.* 2011, *83*, 3205–3210. [CrossRef]
- Tilli, T.M.; Mello, K.D.; Ferreira, L.B.; Matos, A.R.; Accioly, M.T.S.; Faria, P.A.S.; Bellahcène, A.; Castronovo, V.; Gimba, E.R. Both Osteopontin-c and Osteopontin-b Splicing Isoforms Exert pro-tumorigenic Roles in Prostate Cancer Cells. *Prostate* 2012, 72, 1688–1699. [CrossRef]
- 37. Su, Z.-D.; Sun, L.; Yu, D.-X.; Li, R.-X.; Li, H.-X.; Yu, Z.-J.; Sheng, Q.-H.; Lin, X.; Zeng, R.; Wu, J.-R. Quantitative Detection of Single Amino Acid Polymorphisms by Targeted Proteomics. *J. Mol. Cell Biol.* **2011**, *3*, 309–315. [CrossRef]
- 38. Petrovskiy, D. Supersecondary_Structures_Dataset.zip. figshare. Dataset. 2022. Available online: https://figshare.com/articles/ dataset/supersecondary_structures_dataset_zip/21529812/1 (accessed on 18 November 2022). [CrossRef]
- 39. Dufter, P.; Schmitt, M.; Schütze, H. Position Information in Transformers: An Overview. *arXiv* 2021, arXiv:2102.11090[cs]. [CrossRef]
- 40. Mayachita, I. Understanding Graph Convolutional Networks for Node Classification. Towards Data Science. Available online: https://towardsdatascience.com/understanding-graph-convolutional-networks-for-node-classification-a2bfdb7aba7b (accessed on 14 February 2022).
- 41. Jing, B.; Eismann, S.; Suriana, P.; Townshend, R.J.L.; Dror, R. Learning from Protein Structure with Geometric Vector Perceptrons. 2020. Available online: https://arxiv.org/abs/2009.01411 (accessed on 18 November 2022).
- 42. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* 2017, arXiv:1706.03762[cs].
- 43. Zhang, H.; Li, M.; Wang, M.; Zhang, Z. Understand Graph Attention Network—DGL 0.6.1 Documentation. Available online: https://docs.dgl.ai/en/0.6.x/tutorials/models/1_gnn/9_gat.html (accessed on 14 February 2022).
- 44. Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R.J.; Milles, L.F.; Wicky, B.I.M.; Courbet, A.; de Haas, R.J.; Bethel, N.; et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **2022**, *378*, 49–56. [CrossRef] [PubMed]





Computational Tactics for Precision Cancer Network Biology

Heewon Park ^{1,*} and Satoru Miyano ^{1,2}

- ¹ M&D Data Science Center, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan
- ² Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan
- * Correspondence: hwpark.dsc@tmd.ac.jp

Abstract: Network biology has garnered tremendous attention in understanding complex systems of cancer, because the mechanisms underlying cancer involve the perturbations in the specific function of molecular networks, rather than a disorder of a single gene. In this article, we review the various computational tactics for gene regulatory network analysis, focused especially on personalized anticancer therapy. This paper covers three major topics: (1) cell line's (or patient's) cancer characteristics specific gene regulatory network estimation, which enables us to reveal molecular interplays under varying conditions of cancer characteristics of cell lines (or patient); (2) computational approaches to interpret the multitudinous and massive networks; (3) network-based application to uncover molecular mechanisms of cancer and related marker identification. We expect that this review will help readers understand personalized computational network biology that plays a significant role in precision cancer medicine.

Keywords: gene regulatory network; computational cancer biology; precision medicine; oxaliplatin and capecitabine (XELOX)

1. Introduction

Gene regulatory network describes functional interactions between genes, where the network is presented by a graph whose nodes present the genes, and the edges between nodes represent the regulatory interactions between genes [1,2]. Heterogeneous gene regulatory system is a useful tool to analyse and visualize biological activities and is crucial to understanding complex biological processes of cancer, because the molecular mechanisms underlying diseases reflect the perturbations in a specific function of molecules in the complex cellular network, rather than a consequence of an abnormality in a single gene [3].

The molecular interplays between genes involved in cellular processes and pathways can be represented by statistical and mathematical models. The computational strategies to estimate large-scale gene networks from gene expression levels have drawn a large amount of attention. The Gaussian graphical model (GGM), that is the probability model, has often been used to infer the conditional dependence structure of a set of genes. The GGM represents which genes (variables) predict one another and allows for sparse modeling of covariance structures, highlighting potential causal relationships between genes [4]. The Bayesian network (BN) is also a probabilistic graphical model describing a directed acyclic graph. BN has been used to uncover cancer mechanisms, i.e., unique cancer molecular mechanisms of clone cancer [5], causal networks of breast metastasis to bone, brain, or lung [6], assessing the risk of breast cancer [7], etc. Boolean networks are discrete models and one of the most widely used techniques to estimate the gene regulatory system. In the model, gene expression levels are discretized, and each gene takes on two values, i.e., if the gene expression is above a threshold value, then 1, otherwise 0, and the interactions between genes are described by standard logic (Boolean) functions [8]. Various cancer

Citation: Park, H.; Miyano, S. Computational Tactics for Precision Cancer Network Biology. *Int. J. Mol. Sci.* 2022, 23, 14398. https:// doi.org/10.3390/ijms232214398

Academic Editor: Wajid Zaman

Received: 16 October 2022 Accepted: 17 November 2022 Published: 19 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). research has been based on Boolean networks for cancer drug discovery [9], identifying lung cancer diagnostic and prognostic biomarkers [10], uncovering the mechanisms of tumorigenesis and possible treatment responses of prostate cancer [11], etc. Additionally, various computational models and strategies (e.g., differential equation-based Model, artificial neural network (ANN) approaches, correlation network, information theory, etc.) have been developed and applied to cancer research. Furthermore, the effectiveness of the networks-based analysis has been proven in various fields of research, e.g., cancer prediction, drug combinations identification, and protein-protein interaction [12–14].

Although many computational tactics for gene regulatory network estimation have been developed and numerous studies have been conducted to uncover cancer mechanisms based on the estimated gene networks, the existing studies were conducted by an averaged gene network for all cell lines. Thus, we cannot effectively identify crucial information for precision cancer medicine.

In this article, we reviewed the computational strategies for the cell line's (or patient's) cancer characteristics specific gene network analysis. Especially, we reviewed machine learning approaches for varying coefficient models, where the varying coefficients describe the strength of the interaction between genes for a specific characteristic of each cell line. That is, the model enables us to construct a gene regulatory network for a specific status related to cancer of the cell line. The cell line characteristic specific gene networks estimation provides hundreds of networks for hundreds of cell lines, where each network is given as a matrix form with about 20,000 columns for target genes, 2000 rows for regulator genes, and the elements of the matrix indicate the strength of interaction between the regulator and target genes. The analysis and interpretation of the multiple and massive networks are quite difficult tasks and have remained a serious challenge in computational biology. In this article, we also review some computational tactics for comprehensive analysis and interpretation of the large-scale networks.

The remainder of this paper is organized as follows. In the gene regulatory network estimation section, the regression framework to gene regulatory network estimation is represented. The computation tactics estimate the cell line characteristic specific gene regulatory network in the sample-specific gene network estimation section. In the section of gene network analysis in multi-dimensional cell line space, the machine learning and Artificial Intelligence (AI) approaches to comprehensive analysis of the estimated multiple and massive gene regulatory network are represented. In the Applications section, the application result of the reviewed computation strategies for network-based anti-cancer drug prediction and related markers identification is introduced. Conclusions are provided in the Discussion section.

2. Gene Regulatory Network Estimation

Suppose $X = (x_1, ..., x_n)^T \in \mathbb{R}^{n \times p}$ is an $n \times p$ data matrix describing the expression of p regulators that may control the transcription of ℓ^{th} target gene $y_{\ell} \in \mathbb{R}^n$, $\ell = 1, ..., q$. Consider the linear regression model,

$$\mathbf{y}_{\ell} = \sum_{j=1}^{p} \beta_{j\ell} \mathbf{x}_{j} + \boldsymbol{\epsilon}_{\ell}, \quad \ell = 1, \dots, q,$$
 (1)

where $\beta_{j\ell}$ describes the effect of the j^{th} regulator gene on the ℓ^{th} target gene, and ε_{ℓ} is a random error vector $\varepsilon_{\ell} = (\varepsilon_{\ell 1}, \dots, \varepsilon_{\ell n})^T$ that is assumed to be independently and identically distributed with mean 0 and variance σ^2 . To estimate the gene regulatory network, the following L_1 -type regularization methods were used successfully,

$$L(\boldsymbol{\beta}_{\ell}) = \arg\min_{\boldsymbol{\beta}_{\ell}} \{ \frac{1}{2} \sum_{i=1}^{n} (y_{i\ell} - \sum_{j=1}^{p} \beta_{j\ell} x_{ij})^{2} + P(\boldsymbol{\beta}_{\ell}) \},$$
(2)

where

- ridge [15]: $P(\boldsymbol{\beta}_{\ell}) = \lambda \sum_{j=1}^{p} \beta_{j\ell}^2$
- lasso [16]: $P(\boldsymbol{\beta}_{\ell}) = \lambda \sum_{j=1}^{p} |\boldsymbol{\beta}_{j\ell}|$
- elastic net [17]: $P(\boldsymbol{\beta}_{\ell}) = \lambda \sum_{j=1}^{p} \{\gamma \beta_{j\ell}^2 + (1-\gamma) |\beta_{j\ell}|\}$

etc

and $\lambda, \gamma > 0$ are the regularization parameters, where λ controls model complexity, and γ is a mixing parameter between the lasso and ridge penalties. The L_1 -type regularization methods enable us to simultaneously identify crucial regulators and estimate their effect on a target gene. In particular, the methods effectively perform analysis of the high dimensional genomic alterations dataset.

Although the methods successfully perform edge selection and network estimation, the approaches provide an averaged network for all *n* cell-lines. Thus, we cannot estimate cell line (or patient) characteristic-specific models (i.e., molecular interplay). In other words, the methods are not enough to extract useful information for precision medicine.

3. Sample-Specific Gene Network Estimation

To effectively extract crucial information for precision medicine, cell line (or patient) characteristic-specific identification is a crucial issue. We reviewed computational approaches for cell-line characteristic-specific modelling, especially cell line characteristic-specific gene regulatory network estimation. The following varying coefficient model was used for cell-line characteristic-specific modelling [18],

$$\boldsymbol{y}_{\ell} = \sum_{j=1}^{p} \beta_{j\ell}(m_{\alpha}) \cdot \boldsymbol{x}_{j} + \boldsymbol{\epsilon}_{\ell}, \quad \ell = 1, \dots, q, \quad \alpha = 1, \dots, n,$$
(3)

where $\beta_{j\ell}(m_{\alpha})$ describes the effect of the *j*th regulator gene on the ℓ^{th} target gene in the α^{th} target cell line. m_{α} is a cancer related characteristic of the α^{th} cell lines, such as drug sensitivity and survival risk of cell lines. The model enables us to describe cell-line characteristic- $(M = m_{\alpha})$ specific molecular interplay between genes, i.e., $\beta_{j\ell}(m_{\alpha})$.

3.1. NetworkProfiler

The varying coefficient $\beta_{j\ell}(m_{\alpha})$ describing cell-line characteristic-specific strength of the relationship between the *j*th regulator and the ℓ^{th} target genes in the α^{th} cell line can be estimated by the following kernel-based L_1 -type regularization method, called a NetworkProfiler [19],

$$L(\boldsymbol{\beta}_{\ell\alpha}|b_{\ell}) = \frac{1}{2} \sum_{i=1}^{n} \{y_{i\ell} - \sum_{j=1}^{p} \beta_{j\ell}(m_{\alpha})x_{ij}\}^2 G(m_i - m_{\alpha}|b_{\ell}) + P(\boldsymbol{\beta}_{\ell\alpha}),$$
(4)

where

$$G(m_i - m_{\alpha} | b_{\ell}) = \exp\{\frac{-(m_i - m_{\alpha})^2}{b_{\ell}}\},$$
(5)

is a Gaussian kernel function to control the weight of cell lines when modelling the α^{th} target cell line. The NetworkProfiler groups cell lines, according to the similarity of the specific characteristics of cell lines (i.e., modulator m_i for i = 1, ..., n), and performs modelling for α^{th} cell lines, based only on the cell-lines in the neighbourhood around the α^{th} cell line. That is, the modelling for the α^{th} cell line is based only on the cell lines having similar modulator characteristics to with the target sample's modulator value m_{α} . This implies that the NetworkProfiler can estimate cell line characteristic -specific gene regulatory networks.

The cancer-related characteristics of cell lines are not usually uniformly distributed. Figure 1 shows the anti-cancer drug sensitivity of cell lines, where the eight drugs are randomly selected from Genomics of Drug Sensitivity in Cancer (GDSC) and Cancer



Dependency Map (DepMap) projects. As shown in Figure 1, sensitivities of some anticancer drugs (characteristics of cell lines: modulator) are non-uniformly distributed, i.e., there are cell lines having rare cancer characteristics.

Figure 1. Drug sensitivities of GDSC and DepMap databases: each of the four drugs are randomly selected from GDSC and DepMap datsets.

Limitation: The NetworkProfiler cannot perform well when the modulator is not uniformly distributed, especially when modelling the target cell line with a rare characteristic located in a sparse region of its distribution, because the method is based on the constant bandwidth (b_ℓ) of Gaussian kernel function. In the NetworkProfiler, the bandwidth specifies the length-scale of the kernel function and controls the weights of cell lines. It implies that the NetworkProfiler based on the constant bandwidth performs cell line characteristic-specific modelling without consideration of the distribution of the modulator and location of the modulator value m_{α} of the target sample in the distribution. Thus, the NetworkProfiler imposes a small amount of weight to almost all samples for modelling a target sample in a sparse region. Figure 2 shows the values of the Gaussian kernel function (i.e., weight for cell lines) with a constant bandwidth b_{ℓ} for a target sample in both sparse and dense regions, where *y*-axis and *x*-axis indicate weights and modulator values of cell-lines, respectively. As shown in Figure 2, the Gaussian kernel function based on the constant bandwidth imposes the non-zero weight on only a few samples for the modelling of the target sample in a sparse region. It leads to extremely high dimensional data situations; thus, gene regulatory network estimation (i.e., edges selection and edge size estimation) cannot be appropriately performed.



Figure 2. Gaussian kernel function to impose weight on cell lines, where *y*-axis and *x*-axis indicate weight and modulator values of cell lines.

3.2. Adaptive NetworkProfiler

To settle on the issue, Park et al. [20] developed a novel strategy, called an adaptive NetworkProfiler, based on the adaptive bandwidth of the Gaussian function. The adaptive NetworkProfiler computes the weight of cell lines by using the adaptive Gaussian kernel function, where the bandwidth is based on the *k*-nearest neighbour (KNN) rule, called an adaptive bandwidth [21]. The adaptive bandwidth for an α^{th} target cell line is based on the Euclidean distance between the modulator value of α^{th} cell line (m_{α}) and its k^{th} nearest neighbour. By using not a constant but the adaptive bandwidth based on Euclidean distance, the KNN-Gaussian kernel function has a relatively wide width of the kernel for modelling a target sample, having a rare modulator value located in the sparse region, because the k^{th} nearest neighbourhood of α^{th} target cell line is also far from the m_{α} . Thus, the KNN-Gaussian kernel function can overcome the drawback of the ordinary NetworkProfiler for modelling a target sample in the sparse region.

The adaptive NetworkProfiler was developed based on the adaptive kennel function, with an additional parameter incorporating dispersion of modulators (i.e., range of a modulator: r(M)) as follows,

$$L(\boldsymbol{\beta}_{\ell\alpha}|b_{\ell}^{\text{KNN}}, r(M)) = \frac{1}{2} \sum_{i=1}^{n} \{y_{i\ell} - \sum_{j=1}^{p} \beta_{j\ell}(m_{\alpha}) x_{ij}\}^{2} K(m_{i} - m_{\alpha}|b_{\ell\alpha, r(M)}^{\text{KNN}}) + P(\boldsymbol{\beta}_{\ell\alpha}), \quad (6)$$

where

$$K(m_i - m_\alpha | b_{\ell\alpha}^{\text{KNN}}, r(M)) = \exp\left(\frac{-(m_i - m_\alpha)^2}{b_{\ell\alpha}^{\text{KNN}} \cdot r(M)}\right),\tag{7}$$

and $b_{\ell\alpha}^{\text{KNN}}$ is the Euclidean distance between m_{α} , with its k^{th} nearest neighbour $m_{\alpha}^{k^{th}}$,

$$b_{\ell\alpha}^{\text{KNN}} = \sqrt{(m_{\alpha} - m_{\alpha}^{k^{th}})^2} \quad \text{for} \quad \alpha = 1, 2, \dots, n,$$
 (8)

and r(M) is the hyperparameter, incorporating dispersion of the modulator *M*.

The adaptive bandwidth in the Gaussian kernel function with the additional parameter incorporates distribution of cell line characteristics (M) and location of the characteristic value (m_{α}) in their distribution. Thus, the adaptive NetworkProfiler can overcome only a small number of samples that have non-zero weight and can effectively perform cell line characteristic-specific gene network estimation for modelling the target sample in not only dense regions but also sparse regions.

Limitation: The NetworkProfiler and Adaptive NetworkProfiler construct the cancer characteristics-specific gene network based on a specific cancer characteristic. That is, the methods consider a characteristic and measure similarity of

cell lines in one-dimensional cell line characteristic space based only on one characteristic. Thus, the cancer characteristic-specific gene networks estimated by the methods cannot described gene regulatory system under varying conditions of various cancer characteristics because the methods are based on a characteristic.

3.3. Gene Network Analysis in Multi-Dimensional Cell Line Space

In order to incorporate various cancer-related characteristics of cell lines and extract more precise cell-line specific molecular interplays, the cell line characteristic specific gene network estimation is extended to the multi-dimensional cell line space [22].

For *h* characteristics of cell line $M = (m_1, ..., m_h) \in \mathbb{R}^{n \times h}$, the varying coefficient model in (3) is given as follows,

$$\boldsymbol{y}_{\ell} = \sum_{j=1}^{p} \beta_{j\ell}(\boldsymbol{m}_{\alpha}) \cdot \boldsymbol{x}_{j} + \boldsymbol{\epsilon}_{\ell}, \quad \ell = 1, \dots, q, \quad \alpha = 1, \dots, n,$$
(9)

where $m_{\alpha} = (m_{\alpha 1}, \dots, m_{\alpha h})$. In the *h*-dimensional cell line space, the similarity between cell lines is measured by the following multivariate Gaussian kernel function,

$$K(\boldsymbol{m}_{i} - \boldsymbol{m}_{\alpha} | \boldsymbol{H}_{\ell}) = |\boldsymbol{H}_{\ell}|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{m}_{i} - \boldsymbol{m}_{\alpha})^{T} \boldsymbol{H}_{\ell}^{-1}(\boldsymbol{m}_{i} - \boldsymbol{m}_{\alpha})\}$$
(10)

where H_{ℓ} is the bandwidth matrix (e.g., covariance matrix). Then, the multi-dimensional cell line characteristic specific gene network is estimated by the following multivariate kernel-based L_1 -type regularization method,

$$L(\boldsymbol{\beta}_{\ell\alpha}|\boldsymbol{H}_{\ell}) = \frac{1}{2} \sum_{i=1}^{n} \{y_{i\ell} - \sum_{j=1}^{p} \beta_{j\ell}(m_{\alpha}) x_{ij}\}^{2} K(\boldsymbol{m}_{i} - \boldsymbol{m}_{\alpha}|\boldsymbol{H}_{\ell}) + P(\boldsymbol{\beta}_{\ell\alpha}).$$
(11)

The multi-dimensional cell line characteristic specific analysis enables us to extract more precise characterization of cell-lines, and thus we can effectively estimate precision cancer gene regulatory networks.

Limitation: The precision cancer gene networks estimation provides hundreds of matrices with more than 2000 rows for regulator genes and more than 10,000 columns for target genes. Although various computational tactics have been developed and successfully applied to gene regulatory network estimation, the interpretation of the large-scale gene networks remains a challenge. The existing studies on the cell line characteristic-specific gene networks focused only on the known markers and then interpreted the massive networks based on the neighbourhoods of the known markers, i.e., only narrow interpretation was performed. However, comprehensive analysis of the multiple massive networks is essential to understand the complex mechanism of cancer. The interpretation of the multi-layer massive network was the bottle network of the existing studies on the precision cancer gene networks analysis.

4. Interpretation of the Multi-Layer Massive Networks

In this section, we review computational strategies to interpret the multiple and massive gene regulatory networks.

4.1. Network Constrained Sparse Common Component Analysis (NetSCCA)

Park et al. [22] considered common structure identification of the multiple matrix datasets to interpret multilayer massive networks. The cell line specific gene regulatory

system can be described by the following regulatory effect of the j^{th} regulator gene on the ℓ^{th} target gene in the α^{th} cell line [19,22],

$$r_{\alpha lj} = \hat{\beta}_{\ell j}(m_{\alpha}) \cdot x_{\alpha j}, \quad \text{for} \quad j = 1, \dots, p,$$
(12)

where $x_{\alpha j}$ is an expression level value of the j^{th} gene in the α^{th} cell line. For the ℓ^{th} target gene, a matrix for the regulatory effect of p regulators is given as $\mathbf{R}_{\ell} = (\mathbf{r}_{1\ell}, \dots, \mathbf{r}_{n\ell})^T \in \mathbb{R}^{n \times p}$, where $\mathbf{r}_{\alpha \ell} = (\mathbf{r}_{\alpha \ell 1}, \dots, \mathbf{r}_{\alpha \ell p})^T$.

To interpret the large-scale gene regulatory networks and identify crucial biomarkers that play a key role in cancer-related mechanism of interest, the network-constrained sparse common component analysis (NetSCCA) was developed. The crucial common component of multiple datasets (R_{ℓ} , $\ell = 1, ..., q$) can be estimated by [23],

$$\arg\min_{A} \{ \sum_{\ell=1}^{q} \| \boldsymbol{R}_{\ell} - \boldsymbol{R}_{\ell} \boldsymbol{A} \boldsymbol{A}^{T} \|_{F}^{2} \},$$
(13)
subject to $\boldsymbol{A}^{T} \boldsymbol{A} = \boldsymbol{I}_{K}.$

As show in (13), the common component analysis of q datasets can be considered as a principal component analysis (PCA) of q datasets. That is, if there is only one dataset R_1 , then the model becomes a standard PCA. Wang et al. [23] showed that the common loading matrix A in (13) can be optimized as the solution to the following problem,

$$\arg \max_{A} \operatorname{Tr}(A^{T}GA),$$
(14)
subject to $A^{T}A = I_{K}.$

where $G = \sum_{\ell=1}^{q} R_{\ell}^{T} R_{\ell}$. It implies that the common loading matrix *A* can be estimated by the standard PCA problem.

$$\underset{A}{\operatorname{arg\,min}} \{ \| Q - QAA^T \|_F^2 \}, \tag{15}$$

subject to $A^T A = I_K,$

where Q is the square root of G, i.e., $Q^T Q = G$. The common component analysis enables us to estimate the common subspace of the multiple massive networks (i.e., $R_{\ell}, \ell = 1, ..., q$), and extract the crucial common component of the datasets.

The common component estimation in (13) provides a fully dense projection matrix *A*. That is, the common component is estimated by a linear combination of all features. It not only leads to difficult to interpret estimated common components but also erroneous estimation results, because the common component analysis is based on crucial and noisy features. To settle the issue, a sparse learning-based strategy was proposed and developed to achieve better biological interpretability, called a NetSCCA [22]. The NetSCCA estimates the projection matrix *A* based only on crucial features without disturbance of noisy features by using sparse learning and incorporates network biology knowledge that the genes with similar molecular interactions may have similar biological function in the common component estimation.

The NetSCCA measures the similarity between genes on networks by using the following jaccard similarity [24]:

$$W_{j,s} = \frac{|N_j \cap N_s|}{|N_j \cup N_s|} \tag{16}$$

where N_j is the set of nodes that are directly connected to the j^{th} gene via an edge in at least one cell line. Then, the similarity between genes $W_{j,s}$ is incorporated into the sparse common loading matrix (*A*) estimation as follows,

$$\arg\min_{\Theta,A} \{ \sum_{\ell=1}^{q} \| Q - Q \Theta A^{T} \|_{F}^{2} \} + \lambda_{1} \sum_{k=1}^{K} \| \theta_{k} \|_{1} + \lambda_{2} \sum_{k=1}^{K} \sum_{j < s} (\theta_{j,k} - \theta_{s,k})^{2} W_{j,s},$$
(17)

where $\theta_k \propto A_k$ is the *p*-dimensional vector and $\Theta = (\theta_1, \dots, \theta_k)$. The last term (penalty term) of (17) enables us to locally smooth the coefficients and encourage the simultaneous selection of related genes. In other words, a large amount of weight is imposed on the coefficients of the two genes with many common interactions, and it encourages similarity in their coefficients of common structure estimation. Thus, the NetSCCA can provide biologically interpretable results of the common component analysis of the multiple networks.

The NetSCCA algorithm is given in Algorithm 1.

Algorithm 1 NetSCCA: Network constrained sparse common component analysis.

- 1: Compute jaccard similarity: *W*.
- 2: For *q* target genes, compute the regulator effect matrices as R_{ℓ} for $\ell = 1, ..., q$ and $G = \sum_{\ell=1}^{q} R_{\ell}^{T} R_{\ell}$.
- 3: For the square root of *G* (i.e., $Q^T Q = G$), compute sparse common loadings of *q* regulator effect matrices R_{ℓ} , $\ell = 1, ..., q$.

3.1: Start *A* at $V = [V_1, V_2, ..., V_K]$, which is the loading matrix from ordinary PCA of *Q*.

3.2: Given a fixed $A = [a_1, a_2, ..., a_K]$, solving the following problem,

$$\hat{\boldsymbol{\theta}}_k \arg\min_{\boldsymbol{\theta}_k} \{ \|\boldsymbol{z}_k - \boldsymbol{Q}\boldsymbol{\theta}_k\|^2 \} + \lambda_1 \|\boldsymbol{\theta}_k\|_1 + \lambda_2 \sum_{j < s} (\theta_{j,k} - \theta_{s,k})^2 W_{j,s} \quad k = 1, 2, \dots, K,$$

where $z_k = Qa_k$. Update $\hat{\Theta} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_K]$.

3.3: For a fixed $\hat{\Theta}$, perform the singular value decomposition of $Q^T Q \hat{\Theta} = U \Gamma V^T$ and update $\hat{A} = U V^T$ (see Zou et al. [25]).

3.4: Repeat Steps 3.2–3.3, until convergence.

- 4: Sparse common loading is given by $\frac{\hat{\theta}_k}{\|\hat{\theta}_k\|}$ for k = 1, ..., K.
- **Limitation:** As pointed out by existing studies on network-based regularization [26,27], the network-constrained regularisation cannot perform well when the connected genes have opposite signs of coefficients. The limitation of the NetSCCA can be overcome by use of the advanced network-constrained regularization methods that incorporate signs of the regression coefficients [27].

4.2. Explainable AI for Gene Network-Based Prediction (Xprediction)

In this section, we review an explainable AI approach for the network-based prediction, called Xprediction [28]. Although the machine learning-based AI approaches provide effective prediction results, most of the existing approaches were developed focusing only on mathematical/statistical accuracy. Thus, the existing AI methodologies cannot explain their decision rules (i.e., the existing AI cannot explain how and why a decision has been made, causing the black-box problem). However, the interpretability and explainability are essential for use of AI strategies in various fields of research, especially medical science.

Xprediction achieves not only prediction accuracy but also interpretability of deep learning-based AI. The method is based on the widely used machine learning and deep learning approaches, e.g., the kernel support vector machine, random forest and deep neural network for prediction models, and describes the cruciality of input on output by comparison with the results of the model without the input. That is, Xprediction constructs a model by removing a feature (i.e., by removing a molecular interaction between ℓ^{th} target and j^{th} regulator genes) individually and performing a prediction, and the prediction is iterated based on the randomly constructed cross-validation datasets. The cruciality of each molecular interaction was measured by comparing with the prediction accuracy based on all molecular interactions Acc(\hat{y}).

The significance of each molecular interaction is computed by the *t*-test between prediction accuracies between models with and without the edge (i.e., interaction). Let *N* be a number of iterations for computing prediction accuracy from the randomly constructed cross-validation dataset, then $Acc(\hat{y})$ and $s_{\hat{y}}$ are mean and standard deviation of the prediction of accuracies in *N* iterations, respectively. In the model without (l, j) interaction, corresponding notations are given $N^{(l,j)}$, $Acc(\hat{y}^{(l,j)})$ and $s_{(\hat{y}^{(l,j)})}$, respectively. We performed the following *t*-test,

$$T_{\ell j} = \frac{\overline{\operatorname{Acc}(\hat{\boldsymbol{y}})} - \overline{\operatorname{Acc}(\hat{\boldsymbol{y}}^{(l,j)})}}{s_p \sqrt{\frac{1}{N} + \frac{1}{N^{(l,j)}}}}$$
(18)

where $s_p = \sqrt{\frac{s_{\hat{y}}(N-1)+s_{\hat{y}}(l,j)(N^{(l,j)}-1)}{N+N^{(l,j)}-2}}$. Then, the cruciality of $(l,j)^{th}$ interaction $I_{\ell j}$) on the prediction result was measured by the *p* value of the *t*-test. The algorithm of Xprediction is given in Algorithm 2.

Algorithm 2 Xprediction: explainable prediction.

- 1: Construct prediction models based on the kernel support vector machine (kSVM), Random Forest (RF), and Neural Network (NN).
- Compute prediction accuracies based on *k*-fold cross-validation (CV). The average of the prediction accuracies of *k* validation sets was given as: Acc(ŷ).
- 3: Step 2 is iterated *N* times for randomly constructed *k*-fold CV datasets.
- 4: If $l \leq q$, then
- 5: If $j \leq p$, then
- 6: Delete (l, j) elements from regulatory effect matrices: $R_{\ell}, \ell = 1, ..., q$
- 7: Compute prediction accuracy of the model without (l, j) elements: Acc $(\hat{y}^{(l,j)})$.
- 8: Step 7 is iterated $N^{(l,j)}$ times for randomly constructed *k*-fold CV datasets.
- 9: Perform *t*-test between Acc(\hat{y}) and Acc($\hat{y}^{(l,j)}$) obtained from *N* and $N^{(l,j)}$ iterations and compute *p* value.
- 10: Cruciality of molecular interplays for AI-based prediction results are measured by on *p* value of the *t*-test.
- **Limitation:** The Xprediction constructs $(q \times p) + 1$ prediction models, because prediction accuracies of the model based on the regulatory effect without the (i, j) element should be compared with the model based on the regulatory effect with all elements. This leads to a great amount of computation. The computational complexity is one of limitations of the Xprediction.

5. Applications

In this section, we introduce an application of the introduced computational strategy to precision cancer network analysis. We consider the application of the explainable AI, Xprediction, to identify anti-cancer drug markers. The drug sensitivity data (i.e., primary-screen-replicate-collapsed-logfold-change) and RNA-expression levels of genes are obtained from the CCLE dataset (https://depmap.org/portal/, accessed on 4 August 2022). For expression levels of genes, we extracted 1922 genes that had the highest 10% variances in cell lines. We focused on anti-cancer drugs, capecitabine, and oxaliplatin, which are used in a chemotherapy combination known as XELOX or CAPEOX. The XELOX is used to colorectal and gastric cancer [29–31].

We first estimated capecitabine's sensitivity specific gene networks by using the NetworkProfiler. We then defined oxaliplatin sensitive and resistant cell lines based on fifth (5P) and ninety-fifth (95P) percentiles of the drug sensitivity (DS) values, i.e., sensitive cells: DS < 5P and resistant cells: DS > 95P. We constructed a prediction model based on the deep learning approach (i.e., deep neural network) to predict the sensitivity of the oxaliplatin. In our analysis, a two hidden layered, fully-connected feed-forward neural network was used. The *ReLU* activation function was used on the hidden layers, and the sigmoid function was used on the output layer. We randomly split the dataset 10-fold and evaluated the prediction accuracy based on the 10-fold cross-validation, i.e., prediction accuracy was given as an average of prediction accuracies of 10 test sets. By using Xprediction, crucial molecular interactions to explain sensitivities of the oxaliplatin were identified based on *p* value<0.05. Table 1 shows the identified crucial interactions and corresponding *p* value.

p Value	Interaction	p Value
0.003	TP63→ITGB4	0.039
0.008	$EHF \rightarrow C6 or f132$	0.040
0.018	PPP1R13L \rightarrow LYPD3	0.040
0.019	IFITM1 \rightarrow TBX2	0.040
0.019	JAM3→FMN2	0.042
0.029	S100A7→KRT14	0.044
0.032	KLK8→IFITM1	0.046
0.033	EYA4→RHOD	0.046
0.033	SYNE1→ZEB2	0.048
0.034	NECTIN4→CLDN4	0.049
0.035		
	<i>p</i> Value 0.003 0.008 0.018 0.019 0.029 0.032 0.033 0.033 0.034 0.035	p Value Interaction 0.003 TP63 \rightarrow ITGB4 0.008 EHF \rightarrow C6orf132 0.018 PPP1R13L \rightarrow LYPD3 0.019 IFITM1 \rightarrow TBX2 0.019 JAM3 \rightarrow FMN2 0.029 S100A7 \rightarrow KRT14 0.032 KLK8 \rightarrow IFITM1 0.033 EYA4 \rightarrow RHOD 0.034 NECTIN4 \rightarrow CLDN4

Table 1. Crucial molecular interplays to explain oxaliplatin sensitivity, where $X \rightarrow Y$ indicates interaction from regulator gene *X* to target gene *Y*.

Figure 3 shows gene regulatory networks consisting of the identified crucial molecular interplays to oxaliplatin sensitivity prediction, where the top and bottom indicates the networks in drug sensitive and resistance cell lines, respectively. The edge sizes represent the median of strengths of interactions between genes in drug-sensitive cell lines and -resistant cell lines, respectively.

As shown in Figure 3, drug-sensitive and-resistant cell lines show different gene regulatory systems of the identified markers. The interplay of SYNE1 \rightarrow IFITM1 can be considered as a oxaliplatin-resistant specific gene regulatory system. The interplays of SPRY2 \rightarrow ETV1 and SLPI \rightarrow PTK1B become weaker from sensitive to resistant cell lines. It was uncovered that high expression levels of the identified drug resistant markers SYNE1 and IFITM1 are associated with poorer chemotherapy efficacy of gastric cancer and resistance to endocrine therapy and chemotherapy [32,33]. The existing studies support our results that the high activities of SYNE1 and IFITM1 are characteristics of capecitabine-resistant cell lines. On the other hand, it was demonstrated that the high expression levels of the SPRY2 are associated with chemotherapy-sensitive cell line MEK inhibitors, BRAF inhibitor-resistant cells, and ovarian cancer cells [34–36]. The results of the literature are consistent with our result that the activity of SPRY2 is a signature of capecitabine-sensitive cell lines. This implies that our gene network analysis results are strongly supported by existing literatures.



Gene network in Capecitabine sensitive cell lines

Gene network in Capecitabine resistant cell lines



Figure 3. Gene regulatory networks of the crucial molecular interplays to oxaliplatin sensitivity prediction. Color of edge indicate negative (red) and positive (blue) effects of regular genes on their target genes.

Table 2 shows that the genes consisted of the crucial interplays, related anti-cancer drugs, and cancer, where the column "Resistant" indicates that the gene was identified as a drug-resistant marker in existing studies. It can be seen from Table 2 that more than half of the identified genes are confirmed as a therapeutic target for not only XELOX (i.e., Oxaliplatin and Capecitabine) but also various anti-cancer drugs (e.g., 5-FU, cisplatin, Paclitaxel, etc.). Furthermore, the cancer-related mechanism of the genes has been verified in the literature. Although the mechanism of some genes has not yet been uncovered, it can be considered through our results and literatures that not just a single gene but the identified molecular interplays may be crucial to understanding the mechanism of anti-cancer drug resistance of cell lines.

We suggest though the application results of precision cancer network analysis and literature that molecular interplays between "SYNE1 and IFITM1" may lead to capecitabine resistance of cancer cell lines, while weakening the molecular regulatory interactions between "SPRY2 and ETV1" and "SLPI and PTK1B" induce capecitabine-resistance in cell lines.

Genes	RG/TG	Drugs	Cancer	Resistant	Evidences
C6orf132	TG	-	-		
CALB1	TG	-	-		
CLDN4	TG	5-FU, cisplatin, Paclitaxel, cDDP	PDC, GS, CRC	*	[37-41]
DDR1	TG	Oxaliplatin	GS, CRC		[42-44]
DMKN	RG	-	CRC		[45]
EHF	RG	-	-		
ETV1	TG	oxaliplatin, 5-FU	HCC, GS, CRC	*	[46-48]
EYA4	RG	-	-		-
FMN2	TG	-	-		-
IFITM1	RG,TG	-	GS, CRC, EAC, GBC		[49–51]
ITGB4	TG	cisplatin, erlotinib, 5-Fu	LC	*	[52–54]
IAM3	RG	-	LIC		[55]
IUP	RG	-	-		-
KLK8	RGTG	oxaliplatin	CRC PC	*	[56 57]
KRT14	TG	Erlotinib	LC.BC		[58,59]
KRT16	RG	Erlotinib	BC		[59]
LY6K	RG	-	GC BC	*	[60-62]
LYPD3	TG	-	AML		[63]
MPZL2	RG	-	-		-
		5-FU, Enfortumab			
NECTIN4	RG	Vedotin	CC, BC, GC, LC	*	[64,65]
PI3	RG	-	-		-
PKP3	TG	5-FU, leucovorin, oxaliplatin	CRC		[66]
PPP1R13L	RG	5-FU	GS	*	[67]
PRSS8	RG	-	-		-
		Midostaurin,			
PTK2B	TG	gilteri- tinib/defactinib, TKI	AML	*	[68]
RHOD	TG	-	-		-
S100A14	TG	-	GS		[69]
S100A7	RG	-	-		-
SH2D3A	TG	-	-		-
SLPI	RG	-	-		-
SPRY2	RG	5-FU	CC		[70]
SYNE1	RG	-	GC		[32]
SYTL1	RG	-	-		-
TD \/2	TC	platinum-ased			
TBX2	IG	chemotherapy	OC, CRC		[71,72]
TMEM265	TG	-	-		-
TP63	RG	apatinib and capecitabine ovaliplatin and	BC		[73]
ZEB2	TG	5-FU, cisplatin, trastuzumab	CRC, GC	*	[74–76]

Table 2. Identified markers and their evidence.

AML: Acute myelogenous leukaemia; BC: breast cancer; CC: colon cancer; CRC: colorectal cancer; EAC: oesophageal adenocarcinoma; GBC: gallbladder cancer; GS: gastric cancer; HCC: hepatocellular carcinoma; LC: lung cancer; OC: ovarian cancer; PC: pancreatic cancer; PDC: pancreatic ductal carcinomas; cDDP: cisdiamminedichloroplatinum. The * indicates that the gene was identified as a drug resistant marker in the existing studies.

6. Discussion

In this article, we reviewed computational tactics for precision cancer network analysis. Although many studies have been conducted to develop computational approaches to gene regulatory network analysis and the gene network-based analysis has been applied to cancer research, the existing studies focused on an averaged gene network for all cell lines. Thus, we cannot extract crucial information for precision cancer research. In this article, we have focused on cancer characteristic-specific gene networks and reviewed the computational strategies for cell line specific modelling to identified cancer characteristic-specific molecular interplays. We also reviewed the studies on analysis and interpretation

of the estimated multiple and massive gene regulatory networks. Finally, we introduced the application results of the introduced computational tactics to anti-cancer drug sensitivity-specific gene network analysis. The application section described cell line's characteristic-(drug sensitivity) specific gene regulatory network analysis. Our analysis can be easily extended to patient's characteristic-specific gene network analysis by using expression levels and drug sensitivities summarized in each patient. We expect that the results of a cancer patient's characteristic-specific gene network analysis provides crucial evidence for precision medicine.

Although we have reviewed some computational tactics for interpretation of multiple and massive gene networks, from cell line characteristic-specific gene network estimation to computational network biology, interpretation and analysis of the large-scale gene networks is still in its infancy. Thus, researchers in various fields of research are faced with a challenge to interpret the estimated large gene networks. Explainable machine learning and, more specifically, interpretable artificial intelligence will be a key tool to overcome this bottleneck in the near future.

Author Contributions: H.P. performed data analysis for the Applications section and drafted the manuscript. S.M. supervised the works. All authors have read and approved the final version of the manuscript.

Funding: This work was supported by MEXT, as a "Program for Promoting Researches on the Supercomputer Fugaku" (Unravelling origin of cancer and diversity by large-scale data analysis and artificial intelligence technology, Project ID: JPMXP1020200102, hp200138, hp210167, hp220163), and by JSPS KAKENHI (JP19K20402, JP22K12259).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in the Application section are from the Dependency Map (DepMap) projects (https://depmap.org/portal/, accessed on 4 August 2022).

Acknowledgments: This research used the computational resources of supercomputer Fugaku, provided by the RIKEN Center for Computational Science and the Super Computer System, Human Genome Center, Institute of Medical Science, University of Tokyo.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Ristevski, B. Overview of Computational Approaches for Inference of MicroRNA-Mediated and Gene Regulatory Networks. *Advan. Comp.* **2015**, *97*, 111–145.
- Aittokallio, T.; Schwikowski, B. Graph-based methods for analysing networks in cell biology. *Brief Bioinform.* 2006, 7, 243–255. [CrossRef] [PubMed]
- Ahmed, K.; Park, S.; Jiang, Q.; Yeu, Y.; Hwang, T.; Zhang, W. Network-based drug sensitivity prediction. BMC Med. Genom. 2020, 13 (Suppl. 11), 193. [CrossRef] [PubMed]
- Epskamp, S.; Waldorp, L.; Mottus, R.; Borsboom, D. The Gaussian Graphical Model in Cross-Sectional and Time-Series Data. *Multivariate Behav. Res.* 2018, 53, 453–480. [CrossRef] [PubMed]
- Liu, E.; Li, J.; Kinnebrew, G.; Zhang, P.; Zhang, Y.; Cheng, L.; Li, L. A Fast and Furious Bayesian Network and Its Application of Identifying Colon Cancer to Liver Metastasis Gene Regulatory Networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2021, 18, 1325–1335. [CrossRef] [PubMed]
- 6. Park, S.; Hwang, K.; Chung, C.; Roy, D.; Yoo, C. Causal Bayesian gene networks associated with bone, brain and lung metastasis of breast cancer. *Clin. Exp. Metastasis* **2020**, *37*, 657–674. [CrossRef]
- 7. Huang, Y.; Zheng, C.; Zhang, X.; Cheng, Z.; Yang, Z.; Hao, Y.; Shen, J. The Usefulness of Bayesian Network in Assessing the Risk of Triple-Negative Breast Cancer. *Acad. Radiol.* **2020**, *37*, 282–291. [CrossRef]
- 8. Xiao, Y. A tutorial on analysis and simulation of boolean gene regulatory network models. *Curr. Genom.* **2009**, *10*, 511–525. [CrossRef]
- 9. Biane, C.; Delaplace, F. Causal Reasoning on Boolean Control Networks Based on Abduction: Theory and Application to Cancer Drug Discovery. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 574–1585. [CrossRef]
- 10. Guo, N.; Wan, Y. Network-based identification of biomarkers coexpressed with multiple pathways. *Cancer Inform.* **2014**, *16*, 37–47. [CrossRef]

- 11. Montagud, A.; Beal, J.; Tobalina, L.; Traynard, P.; Subramanian, V.; Szalai, B.; Alfoldi, R.; Puskas, L.; Valencia, A.; Barillot, E.; et al. Patient-specific Boolean models of signalling networks guide personalised treatments. *eLife* 2022, *11*, e72626. [CrossRef] [PubMed]
- Daoud, M.; Mayo, M.; A survey of neural network-based cancer prediction models from microarray data. Artif. Intell. Med. 2019, 97, 204–214. [CrossRef] [PubMed]
- 13. Cheng, F.; Kovacs, I.; Barabasi, A. Network-based prediction of drug combinations. *Nat. Comm.* **2019**, *10*, 1197. [CrossRef] [PubMed]
- Fout, A.; Byrd, J.; Shariat, B.; Ben-Hur, A. Protein interface prediction using graph convolutional networks. In Proceedings of the NIPS'17: 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6533–6542.
- 15. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Techonometrics* **1970**, *12*, 55–67. [CrossRef]
- 16. Tibshirani, R. Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc. Ser. B 1996, 58, 267–288. [CrossRef]
- 17. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B 2005, 67, 301–320. [CrossRef]
- 18. Hastie, T.; Tibshirani, R. Varying-coefficient models. J. R. Stat. Soc. Ser. B 1993, 55, 757–796. [CrossRef]
- Shimamura, T.; Imoto, S.; Shimada, Y.; Hosono, Y.; Niida, A.; Nagasaki, M.; Yamaguchi, R.; Takahashi, T.; Miyano, S. A novel network profiling analysis reveals system changes in epithelial-mesenchymal transition. *PLoS ONE* 2011, *6*, e20804. [CrossRef]
- Park, H.; Shimamura, T.; Imoto, S.; Miyano S. Adaptive NetworkProfiler for Identifying Cancer Characteristic-Specific Gene Regulatory Networks. J. Comput. Biol. 2018, 25, 130–145. [CrossRef]
- 21. Terrell, G.; Scott, D. Variable kernel density estimation. Ann. Stat. 1992, 20, 1236–1265. [CrossRef]
- 22. Park, H.; Yamaguchi, R.; Imoto, S.; Miyano S. Uncovering Molecular Mechanisms of Drug Resistance via Network-Constrained Common Structure Identification. *J. Comput. Biol.* 2022, *29*, 257–275. [CrossRef] [PubMed]
- Wang, H.; Banerjee, A.; Boley, D. Common component analysis for multiple covariance matrices. In Proceedings of the Publication:KDD 11: 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 956–964.
- 24. Li, Y.; Luo, P.; Wu, C. A new network node similarity measure method and its applications. *arXiv* 2014, arXiv:1403.4303.
- 25. Zou, H.; Hastie, T.; Tibshirani, R. Sparse principal component analysis. J. Comput. Graph. Stat. 2006, 15, 265–286. [CrossRef]
- 26. Li, C.; Li, H. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics* **2008**, *24*, 1175–1182. [CrossRef]
- Sun, H.; Lin, W.; Feng, R.; Li, H. Network-regularized high dimensional cox regression for analysis of genomic data. *Stat. Sin.* 2014, 24, 1433–1459. [CrossRef]
- Park, H.; Yamaguchi, R.; Imoto, S.; Miyano S. Xprediction: Explainable EGFR-TKIs response prediction based on drug sensitivity specific gene networks. *PLoS ONE* 2022, 17, e0261630. [CrossRef] [PubMed]
- 29. Mizushima, T.; Ikeda, M.; Kato, T.; Ikeda, A.; Nishimura, J.; Hata, T.; Matsuda, C.; Satoh, T.; Mori, M.; Doki, M. Postoperative XELOX therapy for patients with curatively resected high-risk stage II and stage III rectal cancer without preoperative chemoradiation: A prospective, multicenter, open-label, single-arm phase II study. *BMC Cancer* **2019**, *19*, 929. [CrossRef]
- 30. Satake, H.; Yasui, H.; Kotake, T.; Okita, Y.; Hatachi, Y.; Kotaka, Y.; Kato, T.; Tsuji, A. First-line chemotherapy with capecitabine/oxaliplatin for advanced gastric cancer: A phase I study. *Mol. Clin. Oncol.* 2017, *7*, 347–350. [CrossRef] [PubMed]
- 31. Quek, R.; Lim, W.; Foo, K.; Koo, W.; A-Manaf, A.; Toh, H. Capecitabine and oxaliplatin (XELOX) is safe and effective in patients with advanced gastric cancer. *Acta. Oncol.* 2007, *46*, 1032–1034. [CrossRef]
- 32. Qu, Y.; Gao, N.; Wu, T. Expression and clinical significance of SYNE1 and MAGI2 gene promoter methylation in gastric cancer. *Medicine* **2021**, *100*, e23788. [CrossRef] [PubMed]
- Ogony, J.; Choi, H.; Lui, A.; Cristofanilli, M.; Lewis-Wambi, J. Interferon-induced transmembrane protein 1 (IFITM1) overexpression enhances the aggressive phenotype of SUM149 inflammatory breast cancer cells in a signal transducer and activator of transcription 2 (STAT2)-dependent manner. *Breast Cancer Res.* 2016, *18*, 25. [CrossRef] [PubMed]
- 34. Ahn, J.; Han, B.; Lee, M. Induction of Resistance to BRAF Inhibitor Is Associated with the Inability of Spry2 to Inhibit BRAF-V600E Activity in BRAF Mutant Cells. *Biomol. Ther.* 2015, 23, 320–326. [CrossRef]
- 35. Li, Y.; Umbach, D.; Krahn, J.; Shats, I.; Li, X.; Li, L. Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines. *BMC Genom.* **2021**, *22*, 272. [CrossRef] [PubMed]
- 36. Sun, L.; Ke, M.; Wang, X.; Yin, M.; Wei, J.; Xu, L.; Tian, X.; Wang, F.; Zhang, H.; Fu, S.; et al. FAP high α-SMA low cancer-associated fibroblast-derived SLPI protein encapsulated in extracellular vesicles promotes ovarian cancer development via activation of PI3K/AKT and downstream signaling pathways. *Mol. Carcinog.* 2022, *online ahead of print*.
- 37. Sasaki, T.; Fujiwara-Tani, R.; Kishi, S.; Mori, S.; Luo, Y.; Ohmori H.; Kawahara, I.; Goto, K.; Nishiguchi, Y.; Mori, T.; et al. Targeting claudin-4 enhances chemosensitivity of pancreatic ductal carcinomas. *Cancer Med.* **2019**, *15*, 6700–6708. [CrossRef] [PubMed]
- 38. Yoshida, H.; Sumi, T.; Zhi, X.; Yasui, T.; Honda, K.; Ishiko, O. Claudin-4: A potential therapeutic target in chemotherapy-resistant ovarian cancer. *Anticancer Res.* 2011, *31*, 1271–1277.
- 39. Breed, C.; Hicks, D.; Webb, P.; Galimanis, C.; Bitler, B.; Behbakht, K.; Baumgartner, H. Ovarian Tumor Cell Expression of Claudin-4 Reduces Apoptotic Response to Paclitaxel. *Mol. Cancer. Res.* **2019**, *17*, 741–750. [CrossRef]
- 40. Hicks, D.; Galimanis, C.; Webb, P.; Spillman, M.; Behbakht, K.; Neville, M.; Baumgartner, H. Claudin-4 activity in ovarian tumor cell apoptosis resistance and migration. *BMC Cancer* **2016**, *16*, 788. [CrossRef]

- 41. Nishiguchi, Y.; Fujiwara-Tani, R.; Sasaki, T.; Luo, Y.; Ohmori, H.; Kishi, S.; Mori, S.; Goto, K.; Yasui, W.; Sho, M.; et al. Targeting claudin-4 enhances CDDP-chemosensitivity in gastric cancer. *Oncotarget* **2019**, *10*, 2189–2202. [CrossRef]
- Chen, Y.; Jiang, K.; Bai, X.; Liu, M.; Lin, S.; Xu, T.; Wei, J.; Li, D.; Xiong, Y.; Xin, W.; et al. ZEB1 Induces Ddr1 Promoter Hypermethylation and Contributes to the Chronic Pain in Spinal Cord in Rats Following Oxaliplatin Treatment. *Neuroch. Res.* 2021, 46, 2181–2191. [CrossRef]
- 43. Tao, Y.; Wang, R.; Lai, Q.; Wu, Q.; Wang, Y.; Jiang, X.; Zeng, L.; Zhou, S.; Li, Z.; Yang, T.; et al. Targeting of DDR1 with antibody-drug conjugates has antitumor effects in a mouse model of colon carcinoma. *Mol. Oncol.* 2019, *13*, 1855–1873. [CrossRef]
- 44. Hur, H.; Ham, I.; Lee, D.; Jin, H.; Aguilera, K.; Oh, H.; Han, S.; Kwon, J.; Kim, Y.; Ding, K.; et al. Discoidin domain receptor 1 activity drives an aggressive phenotype in gastric carcinoma. *BMC Cancer* **2017**, *17*, 87. [CrossRef] [PubMed]
- 45. Menor, M.; Zhu, Y.; Wang, Y; Zhang, J.; Jiang, B.; Deng, Y. Development of somatic mutation signatures for risk stratification and prognosis in lung and colorectal adenocarcinomas. *BMC Med. Genom.* **2019**, *12* (Suppl. 1), 24. [CrossRef] [PubMed]
- 46. Chen, J.; Yuan, D.; Hao, Q.; Zhu, D.; Chen, Z. LncRNA PCGEM1 mediates oxaliplatin resistance in hepatocellular carcinoma via miR-129-5p/ETV1 axis in vitro. *Adv. Clin. Exp. Med.* **2021**, *30*, 831–838. [CrossRef] [PubMed]
- Na, D.; Chae, J.; Cho, S.; Kang, W.; Lee, A.; Min, S.; Kang, J.; Kim, M.; Choi, J.; Lee, W.; et al. Predictive biomarkers for 5-fluorouracil and oxaliplatin-based chemotherapy in gastric cancers via profiling of patient-derived xenografts. *Nat. Commun.* 2021, 12, 4840. [CrossRef] [PubMed]
- 48. Giri, A. ETV5 expression positively correlates with promoter methylation and predicts response for 5-FU-based adjuvant therapy response in proximal colon cancer. *bioRxiv* 2020. [CrossRef]
- 49. Borg, D.; Hedner, C.; Gaber, A.; Nodin, B.; Fristedt, R.; Jirstrom, K.; Eberhard, J.; Johnsson, A. Expression of IFITM1 as a prognostic biomarker in resected gastric and esophageal adenocarcinoma. *Biomark. Res.* **2016**, *58*, 10. [CrossRef]
- Virag, P.; Fischer-Fodor, E.; Perde-Schrepler, M.; Brie, I.; Tatomir, C.; Balacescu, L.; Berindan-Neagoe, I.; Victor, B.; Balacescu, O. Oxaliplatin induces different cellular and molecular chemoresistance patterns in colorectal cancer cell lines of identical origins. BMC Genom. 2013, 14, 480. [CrossRef]
- 51. Lin, H.; Zhang, T.; Chen, M.; Shen, J. Novel biomarkers for the diagnosis and prognosis of gallbladder cancer. J. Dig. Dis. 2021, 22, 62–71. [CrossRef]
- Mohanty, A.; Nam, A.; Pozhitkov, A.; Yang, L.; Srivastava, S.; Nathan, A.; Wu, X.; Mambetsariev, I.; Nelson, M.; Subbalakshmi, A.; et al. A Non-genetic Mechanism Involving the Integrin β4/Paxillin Axis Contributes to Chemoresistance in Lung Cancer. *iScience* 2022, 23, 101496. [CrossRef]
- Penzvalto, Z.; Tegze, B.; Szasz, A.; Sztupinszki, Z.; Liko, I.; Szendroi, A.; Schafer, R.; Gyorffy, B. Identifying resistance mechanisms against five tyrosine kinase inhibitors targeting the ERBB/RAS pathway in 45 cancer cell lines. *PLoS ONE* 2013, *8*, e59503. [CrossRef]
- 54. Riedesser, J.; Ebert, M. Precision medicine for metastatic colorectal cancer in clinical practice. *Ther. Adv. Med. Oncol.* 2022, 14, 17588359211072703. [CrossRef] [PubMed]
- 55. Zhang, Y.; Xia, F.; Liu, X.; Yu, X.; Xie, L.; Liu, L.; Chen, C.; Jiang, H.; Hao, X.; He, X.; et al. JAM3 maintains leukemia-initiating cell self-renewal through LRP5/AKT/β-catenin/CCND1 signaling. J. Clin. Investig. 2018, 128, 1737–1751. [CrossRef] [PubMed]
- Zhang, Y.; Xu, Z.; Sun, Y; Chi, P.; Lu, X. Knockdown of KLK11 reverses oxaliplatin resistance by inhibiting proliferation and activating apoptosis via suppressing the PI3K/AKT signal pathway in colorectal cancer cell. *Onco. Targets Ther.* 2018, *11*, 809–821. [CrossRef] [PubMed]
- 57. Hua, Q.; Li, T.; Liu, Y.; Shen, X; Zhu, X.; Xu, P. Upregulation of KLK8 Predicts Poor Prognosis in Pancreatic Cancer. *Front. Oncol.* **2021**, *11*, 624837. [CrossRef] [PubMed]
- Zhang, Z.; Lee, J.; Lin, L.; Olivas, V.; Au, V.; LaFramboise, T.; Abdel-Rahman, M.; Wang, X.; Levine, A.; Rho, J.; et al. Activation of the AXL kinase causes resistance to *EGFR*-targeted therapy in lung cancer. *Nat. Genet.* 2012, 44, 852–860. [CrossRef]
- 59. Masica, D.; Karchin, R. Collections of simultaneously altered genes as biomarkers of cancer cell drug response. *Cancer Res.* **2013**, 73 1699–1708. [CrossRef]
- 60. Luo, L.; McGarvey, P.; Madhavan, S.; Kumar, R.; Gusev, Y.; Upadhyay, G. Distinct lymphocyte antigens 6 (Ly6) family members Ly6D, Ly6E, Ly6K and Ly6H drive tumorigenesis and clinical outcome. *Oncotarget* **2016**, *7*, 11165–11193. [CrossRef]
- AlHossiny, M.; Luo, L.; Frazier, W.; Steiner, N.; Gusev, Y.; Kallakury, B.; Glasgow, E.; Creswell, K.; Madhavan, S.; Kumar, R.; Upadhyay, G. Ly6E/K Signaling to TGFβ Promotes Breast Cancer Progression, Immune Escape, and Drug Resistance. *Cancer Res.* 2016, 76, 3376–3386. [CrossRef]
- 62. Burg, S. Correlates of immune and clinical activity of novel cancer vaccines. Semin. Immunol. 2018, 39, 119–136. [CrossRef]
- 63. Hu, T.; Zhang, Y.; Yang, T.; He, Q.; Zhao, M. LYPD3, a New Biomarker and Therapeutic Target for Acute Myelogenous Leukemia. *Front. Genet.* **2022**, *13*, 795820. [CrossRef]
- 64. Das, D.; Satapathy, S.; Siddharth, S.; Nayak, A.; Kundu, C. NECTIN-4 increased the 5-FU resistance in colon cancer cells by inducing the PI3K-AKT cascade. *Cancer Chemother. Pharmacol.* **2015**, *76*, 471–479. [CrossRef] [PubMed]
- 65. Jin, S.; Sun, Y; Liang, X.; Gu, X.; Ning, J.; Xu, Y.; Chen, S.; Pan, S. Emerging new therapeutic antibody derivatives for cancer treatment. *Signal Transduct. Target Ther.* **2022**, *7*, 39. [CrossRef] [PubMed]
- Lu, W.; Fu, D.; Kong, X.; Huang, Z.; Hwang, M.; Zhu, Y.; Chen, L.; Jiang, K.; Li, X.; Wu, Y.; et al. FOLFOX treatment response prediction in metastatic or recurrent colorectal cancer patients via machine learning algorithms. *Cancer Med.* 2020, *9*, 419–1429. [CrossRef]

- Zhou, M.; Dong, J.; Huang, J.; Ye, W.; Zheng, Z.; Huang, K.; Pan, Y.; Cen, J.; Liang, Y.; Shu, G.; et al. Chitosan-Gelatin-EGCG Nanoparticle-Meditated LncRNA TMEM44-AS1 Silencing to Activate the P53 Signaling Pathway for the Synergistic Reversal of 5-FU Resistance in Gastric Cancer. *Adv. Sci.* 2022, *9*, e2105077. [CrossRef]
- Allert, C.; Waclawiczek, A.; Zimmermann, S.; Gollner, S.; Heid, D.; Janssen, M.; Renders, S.; Rohde, C.: Bauer, M.; Bruckmann, M.; et al. Maximilian Felix Blank 9 Protein tyrosine kinase 2b inhibition reverts niche-associated resistance to tyrosine kinase inhibitors in AML. *Leukemia* 2022, *online ahead of print*.
- Zhang, Q.; Zhu, M.; Cheng, W.; Xing, R.; Li, W.; Zhao, W.; Xu, L.; Li, E.; Luo, G.; Lu, Y. Downregulation of 425G>a variant of calcium-binding protein S100A14 associated with poor differentiation and prognosis in gastric cancer. *J. Cancer Res. Clin. Oncol.* 2015, 141, 691–703. [CrossRef] [PubMed]
- Feng, Y.; Wu, C.; Shiau, A.; Lee, J.; Chang, J.; Lu, P.; Tung, C.; Feng, L.; Huang, W.; Tsao, C. MicroRNA-21-mediated regulation of Sprouty2 protein expression enhances the cytotoxic effect of 5-fluorouracil and metformin in colon cancer cells. *Int. J. Mol. Med.* 2012, *39*, 920–926.
- 71. Luo, J.; Chen, J.; Zhou, J.; Han, K.; Li, S.; Duan, J.; Cao, C.; Lin, J.; Xie, X.; Wang F. TBX20 inhibits colorectal cancer tumorigenesis by impairing NHEJ-mediated DNA repair. *Cancer Sci.* 2022, 113, 2008–2021. [CrossRef]
- 72. Tasaka, R.; Fukuda, T.; Shimomura, M.; Inoue, Y.; Wada, T.; Kawanishi, M.; Yasui, T.; Sumi, T. TBX2 expression is associated with platinum-sensitivity of ovarian serous carcinoma. *Oncol. Lett.* **2018**, *14*, 3085–3090. [CrossRef]
- Esposito, A.; Bardelli, A.; Criscitiello, C.; Colombo, N.; Gelao, L.; Fumagalli, L.; Minchella, I.; Locatelli, M.; Goldhirsch, A.; Curigliano, G. Monitoring tumor-derived cell-free DNA in patients with solid tumors: Clinical perspectives and research opportunities. *Cancer Treat. Rev.* 2014, 40, 648–655. [CrossRef]
- 74. Sreekumar, R.; Al-Saihati, H.; Emaduddin, M.; Moutasim, K.; Mellone, M.; Patel, A.; Kilic, S.; Cetin, M.; Erdemir, S.; Navio, S.; et al. The ZEB2-dependent EMT transcriptional programme drives therapy resistance by activating nucleotide excision repair genes ERCC1 and ERCC4 in colorectal cancer. *Mol. Oncol.* 2021, 15, 2065–2083. [CrossRef]
- 75. Guo, Q.; Jing, F.; Xu, W.; Li, X.; Li, X.; Sun, J.; Xing, X.; Zhou, C.; Jing, F. Ubenimex induces autophagy inhibition and EMT suppression to overcome cisplatin resistance in GC cells by perturbing the CD13/EMP3/PI3K/AKT/NF-κβ axis. Aging 2019, 12, 80–105. [CrossRef] [PubMed]
- 76. Zhou, X.; Men, X.; Zhao, R.; Han, J.; Fan, Z.; Wang, Y.; Lv, Y.; Zuo, J.; Zhao, L.; Sang, M.; et al. miR-200c inhibits TGF-β-induced-EMT to restore trastuzumab sensitivity by targeting ZEB1 and ZEB2 in gastric cancer. *Cancer Gene Ther.* 2018, 25, 68–76. [CrossRef] [PubMed]



Article



Integrative Analysis of Transcriptome-Wide Association Study and Gene-Based Association Analysis Identifies In Silico Candidate Genes Associated with Juvenile Idiopathic Arthritis

Shuai Liu^{1,2,†}, Weiming Gong^{1,2,†}, Lu Liu^{1,2}, Ran Yan^{1,2}, Shukang Wang^{1,2,*} and Zhongshang Yuan^{1,2,*}

- ¹ Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250012, China
- ² Institute for Medical Dataology, Cheeloo College of Medicine, Shandong University, Jinan 250012, China

Correspondence: wsk2001@sdu.edu.cn (S.W.); yuanzhongshang@sdu.edu.cn (Z.Y.)

+ These authors contributed equally to this work.

Abstract: Genome-wide association study (GWAS) of Juvenile idiopathic arthritis (JIA) suffers from low power due to limited sample size and the interpretation challenge due to most signals located in non-coding regions. Gene-level analysis could alleviate these issues. Using GWAS summary statistics, we performed two typical gene-level analysis of JIA, transcriptome-wide association studies (TWAS) using FUnctional Summary-based ImputatiON (FUSION) and gene-based analysis using eQTL Multi-marker Analysis of GenoMic Annotation (eMAGMA), followed by comprehensive enrichment analysis. Among 33 overlapped significant genes from these two methods, 11 were previously reported, including TYK2 ($P_{\text{FUSION}} = 5.12 \times 10^{-6}$, $P_{\text{eMAGMA}} = 1.94 \times 10^{-7}$ for whole blood), IL-6R $(P_{\text{FUSION}} = 8.63 \times 10^{-7}, P_{\text{eMAGMA}} = 2.74 \times 10^{-6}$ for cells EBV-transformed lymphocytes), and Fas $(P_{\text{FUSION}} = 5.21 \times 10^{-5}, P_{\text{eMAGMA}} = 1.08 \times 10^{-6}$ for muscle skeletal). Some newly plausible JIAassociated genes are also reported, including IL-27 ($P_{\text{FUSION}} = 2.10 \times 10^{-7}$, $P_{\text{eMAGMA}} = 3.93 \times 10^{-8}$ for Liver), LAT ($P_{\text{FUSION}} = 1.53 \times 10^{-4}$, $P_{\text{eMAGMA}} = 4.62 \times 10^{-7}$ for Artery Aorta), and MAGI3 $(P_{\text{FUSION}} = 1.30 \times 10^{-5}, P_{\text{eMAGMA}} = 1.73 \times 10^{-7}$ for Muscle Skeletal). Enrichment analysis further highlighted 4 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and 10 Gene Ontology (GO) terms. Our findings can benefit the understanding of genetic determinants and potential therapeutic targets for JIA.

Keywords: juvenile idiopathic arthritis; transcriptome-wide association study; gene-based association analysis; enrichment analysis

1. Introduction

Juvenile idiopathic arthritis (JIA) is one of the most common rheumatic diseases characterized by arthritis in childhood [1]. It can cause damage to multiple organs such as arthrosis, heart, lung, skin, and eyes, and is an important cause of disability in children under the age of 16 [2]. Previous studies have shown that the concordance rate of JIA in monozygotic twins and the relative risk of the disease in the siblings of JIA patients are higher [3,4], highlighting the important role of genetic factors in the development of JIA [5,6]. Therefore, it is of great significance to probe the complex genetic association of JIA to better understand the genetic mechanisms and investigate potential intervention targets.

Genome-wide association studies (GWAS) have successfully identified 22 risk loci associated with JIA [7–9], however, these GWASs suffer from either small sample size or low case proportions, which may be presumably due to the difficulty in accurate JIA diagnosis and the lack of pathognomonic features [10]. So far, the JIA GWAS with relatively larger sample size and the largest case proportions only includes 3305 JIA cases and 9196 controls [9]. In addition, most genetic variants identified from GWAS of JIA are located in non-coding regions [11], leading to the difficulty in explaining the association signals.

Citation: Liu, S.; Gong, W.; Liu, L.; Yan, R.; Wang, S.; Yuan, Z. Integrative Analysis of Transcriptome-Wide Association Study and Gene-Based Association Analysis Identifies In Silico Candidate Genes Associated with Juvenile Idiopathic Arthritis. *Int. J. Mol. Sci.* **2022**, *23*, 13555. https:// doi.org/10.3390/ijms232113555

Academic Editors: Giuseppe Novelli and Wajid Zaman

Received: 21 August 2022 Accepted: 2 November 2022 Published: 4 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). On the other hand, statistically, GWAS often provide JIA-associated single nucleotide polymorphisms (SNPs), the effects of which are too weak to be detected. Gene-level analysis can not only aggregate many SNPs with small effects to improve the power, but provide good biological interpretations, which is more straightforward to be translated into clinical practice.

With the increase of publicly available GWAS summary data and the well-developed efficient tools, it is feasible to conduct the gene-level analysis for JIA [12]. There are two typical gene-level association analysis methods with different model assumptions, one is transcriptome-wide association studies (TWAS), which has shown great promise in interpreting the GWAS signals and is powerful in detecting the association between the gene expression level and the complex disease [13,14]. Recently, one TWAS analysis of JIA has been conducted, however, it only involves two tissues with lower JIA case proportions [15], which may lead to the power loss and the insufficiency in capturing the tissue information related with JIA. The other is multi-marker analysis, which can assign SNPs to genes based on physical proximity and further conduct gene-based association analysis. Both methods, though have different statistical principles, can produce the gene-level *p* values. We would like to emphasize that using different gene-level methods with different model statistical principles to obtain the common genes can avoid the risk of false discoveries from using single method. Actually, the results from different analysis could complement to each other [16,17].

In the present study, using the GWAS summary data of JIA with relatively larger sample size and case proportions (3305 cases and 9196 controls), we performed TWAS analysis and the gene-based association analysis to identify the tissue-specific JIA-associated genes. We used the false discovery rate (FDR) correction on each tissue to declare the significant genes. Finally, we overlapped the genes significantly detected from these two gene-level methods, and performed the enrichment analysis for these overlapped genes on the Metascape website to identify the significant Gene Ontology (GO) terms as well as the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

2. Methods

2.1. Study Design, Data Source, and Quality Control

The analysis flowchart of this study is shown in Figure 1. We used the GWAS of JIA from a large-scale meta-analysis [9]. We obtained the largest GWAS summary statistics of JIA from the NHGRI-EBI GWAS Catalog (Study Accession Code GCST90010715), where the JIA cases were diagnosed according to The International League Against Rheumatism (ILAR) criteria. The data were restricted to European ancestry with stringent quality control as described previously [9]. Briefly, the GWAS of JIA initially recruited 4520 UK JIA patients and 9965 healthy individuals. Individuals with call rate less than 0.98 and discrepancy between genetically predicted sex and database record were removed. In addition, SNPs that were non-autosomal, had a call rate <0.98 or a minor allele frequency (MAF) <0.01 were further excluded. About 12,501 individuals (3305 cases and 9196 healthy controls) were finally remained. For the summary data, we further excluded the major histocompatibility complex (MHC) region (chromosome 6: 25-35 Mb) due to its complex structure, restricted to biallelic SNPs and removed SNPs with duplicated or missing rs ID for subsequent analyses. Totally 7,415,262 SNPs are finally included. Bearing in mind that using different methods with different model assumptions to obtain the overlapped genes can avoid the risk of false discoveries from using single method, we here applied two gene-level approaches with distinct principles, TWAS analysis and gene-based association analysis, as parallel analyses to obtain the common JIA-associated genes for enrichment analysis.



Figure 1. The flowchart of integrative analysis of FUSION and eMAGMA. ABBR: JIA, Juvenile idiopathic arthritis; GWAS, genome-wide association study; MHC, major histocompatibility complex; GTEx, Genotype-Tissue Expression Project; LD, linkage disequilibrium; EUR, European; FUSION, functional summary-based imputation; eQTL, expression quantitative trait loci; eMAGMA, eQTL Multi-marker Analysis of GenoMic Annotation; FDR, false discovery rate; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology.

2.2. TWAS Analysis

TWAS aim to integrate GWAS and expression quantitative trait loci (eQTL) studies to identify tissue-specific gene-trait associations [18], which has shown great promise both in interpreting GWAS findings and in elucidating the underlying disease mechanisms. We here adopted FUnctional Summary-based ImputatiON (FUSION) method (http://gusevlab.org/projects/fusion/, accessed on 8 April 2022) to conducted TWAS analysis of JIA. FUSION is the most commonly used TWAS method and has shown great promise in large-scale integrative omics data analysis [14]. Once inputting the GWAS summary data and expression weight, FUSION will impute the gene expressions in GWAS, and then perform an association analysis between the predicted gene expression and JIA. We selected Genotype-Tissue Expression Project (GTEx) v8 with pre-computed gene expression weights from totally 49 tissues. Using all the tissues may introduce the nuisance information and increase the computation burden; previous studies recommend using an expression panel from only plausible disease-related tissues in TWAS [13]. Here, we determine the analyzed tissue based on not only previous studies [2,19], but also the clinical symptoms and involved organs of JIA, such as hepatomegaly, splenomegaly, anemia, disseminated intravascular coagulation, arthritis, rash, mesenteric lymphadenopathy, pericarditis, and pneumonia. We finally selected 18 tissues for analysis including artery aorta, artery coronary, artery tibial, cells EBV-transformed lymphocytes, colon sigmoid, colon transverse, heart atrial appendage, heart left ventricle, liver, lung, muscle skeletal, nerve tibial, skin not sun exposed suprapubic, skin sun exposed lower leg, small intestine terminal ileum, spleen, stomach, whole blood. We used the 1000 Genomes European panel as a linkage disequilibrium (LD) reference data, and obtained tissue-specific p value for each gene across different tissues. We finally performed FDR correction on each tissue, and genes with FDR less than 0.05 are declared to be significant.

2.3. Gene-Based Association Analysis

We used eQTL Multi-marker Analysis of GenoMic Annotation (eMAGMA) method (https://github.com/eskederks/eMAGMA-tutorial, accessed on 28 March 2022) to conduct gene-based association analysis for JIA. eMAGMA follows the same statistical framework as MAGMA, which is based on a multiple linear principal component regression model and can provide better statistical performance in gene-based association analysis [16]. Besides this, eMAGMA can further utilize tissue-specific cis-eQTL information to assign SNPs to putative genes, providing more biologically meaningful and interpretable results [17]. Gene-based analysis using eMAGMA typically involves two stages, annotation and gene analysis [16]. We, in the annotation stage, used the same tissues as that in above TWAS analysis and directly used the GTEx v8-based annotation files provided on eMAGMA's website. In the gene analysis stage, we used the 1000 Genomes European panel as the reference panel and tested the association between the annotated genes and JIA. We further performed FDR correction on each tissue, and selected significant genes with FDR less than 0.05.

2.4. Gene Set Enrichment Analysis

We conducted gene set enrichment analysis for the overlapped genes that were significantly identified by both TWAS and eMAGMA analysis. Specifically, these overlapped significant genes were subjected to GO term and KEGG pathway enrichment analysis on the Metascape website (https://metascape.org/gp/index.html#/main/step1, accessed on 26 April 2022) to better understand the biological mechanisms. Metascape essentially utilizes the hypergeometric test and Benjamini-Hochberg *p* value correction algorithm to identify all ontology terms. A large number of terms would make the results redundant and complicate the interpretation, Kappa consistency test was thus performed and terms with Kappa > 0.3 were grouped into a cluster, and the most statistically significant term in the cluster was selected to represent the cluster [20]. The parameters of Min Overlap, *p* Value Cutoff, and Min Enrichment are set to be the default values, respectively. In addition, we also made a protein–protein interaction (PPI) network for the overlapped significant genes on the STRING website (https://cn.string-db.org/, accessed on 15 September 2022).

3. Results

3.1. TWAS Analysis

We analyzed all genes involved in the 18 tissues, among which 275 tissue-specific genes were significantly detected by FUSION with FDR less than 0.05. Note that LAT is significantly detected at the border line ($p = 1.53 \times 10^{-4}$ and FDR = 5.40×10^{-2} for Artery Aorta). These TWAS significant genes included some established JIA-associated genes that have been reported previously, such as CCDC101 ($p = 5.82 \times 10^{-8}$ and FDR = 1.64×10^{-4} for Muscle Skeletal), CLN3 ($p = 5.82 \times 10^{-8}$ and FDR = 2.53×10^{-4} for Whole Blood), ERAP2 ($p = 5.49 \times 10^{-6}$ and FDR = 2.16×10^{-3} for Cells EBV-transformed lymphocytes), LNPEP ($p = 3.53 \times 10^{-6}$ and FDR = 2.36×10^{-3} for Artery Tibial). The consistent results with previous GWAS partly indicates the correctness of the FUSION analysis. All significant genes in FUSION analysis were displayed in Supplementary Table S1.

3.2. Gene-Based Association Analysis Results

Similarly, we analyzed all genes involved in the 18 tissues, among which 380 tissuespecific genes were significantly detected by eMAGMA with FDR less than 0.05. These included some well-known JIA-associated genes, such as SGF29 ($p = 2.22 \times 10^{-8}$ and FDR = 3.38×10^{-5} for Whole Blood), ANKRD55 ($p = 3.09 \times 10^{-8}$ and FDR = 1.08×10^{-4} for Spleen), ATP8B2 ($p = 7.53 \times 10^{-7}$ and FDR = 6.59×10^{-4} for Stomach), PTPN2 ($p = 3.96 \times 10^{-6}$ and FDR = 2.75×10^{-3} for Spleen). Again, all these genes are included in the 22 risk loci identified by previous GWAS illustrate the eMAGMA results more reliable. All significant genes from eMAGMA analysis were summarized in Supplementary Table S2.

3.3. Gene Set Enrichment Analysis

We intersected the significant genes from both TWAS analysis and eMAGMA genebased association analysis according to different tissues, and found a total of 132 tissuespecific genes. A total of 33 unique genes were further identified after removing the duplicated ones, among which 11 genes have been reported in previous studies [7–9], such as TYK2 ($P_{FUSION} = 5.12 \times 10^{-6}$, $P_{eMAGMA} = 1.94 \times 10^{-7}$ for Whole Blood), IL(Interleukin)-6R ($P_{FUSION} = 8.63 \times 10^{-7}$, $P_{eMAGMA} = 2.74 \times 10^{-6}$ for Cells EBV-transformed lymphocytes), and Fas ($P_{FUSION} = 5.21 \times 10^{-5}$, $P_{eMAGMA} = 1.08 \times 10^{-6}$ for Muscle Skeletal). The remaining newly discovered 22 genes are APOBR, ATXN2L, GSDMB, IKZF3, IL27, KIAA1109, LAT, LMAN2L, MAGI3, NFATC2IP, NUPR1, ORMDL3, PHTF1, PSMB7, RGS14, SBK1, SH2B1, STEAP1B, TNFSF15, TUFM, UXS1, ZNF197. Among them, IL-27 ($P_{FUSION} = 2.10 \times 10^{-7}$, $P_{eMAGMA} = 3.93 \times 10^{-8}$ for Liver), LAT ($P_{FUSION} = 1.53 \times 10^{-4}$, $P_{eMAGMA} = 4.62 \times 10^{-7}$ for Artery Aorta) and MAGI3 ($P_{FUSION} = 1.30 \times 10^{-5}$, $P_{eMAGMA} = 1.73 \times 10^{-7}$ for Muscle Skeletal) are the novel genes that are more likely to be associated with JIA. Detailed information for 33 genes were summarized in Table 1.

Table 1. Overlapped gene identified by F	FUSION and eMAGMA.
--	--------------------

Gene Symbol	Chromosome	Gene Start (bp)	Gene End (bp)	P _{FUSION}	FDR _{FUSION}	P _{eMAGMA}	FDR _{eMAGMA}
IL6R	1	154,377,669	154,441,926	$8.63 imes 10^{-7}$	$7.80 imes 10^{-4}$	$2.74 imes 10^{-6}$	$1.56 imes 10^{-3}$
MAGI3	1	113,933,371	114,228,545	$1.30 imes10^{-5}$	$7.35 imes10^{-3}$	$1.73 imes10^{-7}$	$2.62 imes10^{-4}$
PHTF1	1	114,239,453	114,302,111	$6.55 imes10^{-7}$	$1.21 imes 10^{-3}$	$2.93 imes10^{-7}$	$5.60 imes10^{-4}$
LMAN2L	2	97,371,666	97,405,801	$4.16 imes10^{-5}$	$1.69 imes 10^{-2}$	$1.56 imes 10^{-5}$	$6.48 imes10^{-3}$
UXS1	2	106,709,759	106,810,795	$1.39 imes10^{-4}$	$4.78 imes10^{-2}$	$1.39 imes10^{-4}$	$3.52 imes 10^{-2}$
ZNF197	3	44,626,380	44,689,963	$7.92 imes 10^{-5}$	$3.28 imes 10^{-2}$	$2.04 imes 10^{-5}$	$1.23 imes 10^{-2}$
KIAA1109	4	123,073,488	123,283,913	$1.51 imes 10^{-5}$	$9.73 imes10^{-3}$	$2.06 imes 10^{-5}$	$1.44 imes 10^{-2}$
ERAP2	5	96,211,643	96,255,420	$5.49 imes10^{-6}$	$2.16 imes10^{-3}$	$1.03 imes10^{-6}$	$8.68 imes10^{-4}$
LNPEP	5	96,271,098	96,373,219	$3.53 imes 10^{-6}$	$2.36 imes 10^{-3}$	$1.50 imes 10^{-6}$	$9.68 imes10^{-4}$
RGS14	5	176,784,838	176,799,602	$1.02 imes 10^{-6}$	$1.97 imes10^{-3}$	$1.07 imes 10^{-5}$	$9.71 imes 10^{-3}$
AHI1	6	135,604,670	135,818,914	$1.98 imes10^{-6}$	$2.61 imes 10^{-3}$	$3.31 imes 10^{-5}$	$1.10 imes10^{-2}$
STEAP1B	7	22,459,063	22,672,544	$1.23 imes 10^{-5}$	$5.29 imes 10^{-3}$	$2.86 imes 10^{-5}$	2.07×10^{-2}
PSMB7	9	127,115,745	127,177,723	$4.84 imes10^{-5}$	$2.28 imes 10^{-2}$	$1.10 imes10^{-4}$	$3.84 imes 10^{-2}$
TNFSF15	9	117,546,915	117,568,406	$4.03 imes10^{-5}$	$1.60 imes10^{-2}$	$4.09 imes10^{-5}$	$1.61 imes 10^{-2}$
ACTA2	10	90,694,831	90,751,147	$5.64 imes 10^{-6}$	$4.05 imes10^{-3}$	$1.03 imes 10^{-6}$	$7.17 imes10^{-4}$
FAS	10	90,750,414	90,775,542	$5.21 imes 10^{-5}$	2.32×10^{-2}	$1.08 imes 10^{-6}$	$8.93 imes10^{-4}$
APOBR	16	28,505,970	28,510,291	$6.52 imes10^{-7}$	$8.61 imes10^{-4}$	$1.25 imes 10^{-6}$	$9.48 imes10^{-4}$
ATP2A1	16	28,889,726	28,915,830	$5.31 imes 10^{-7}$	$1.18 imes10^{-3}$	$3.66 imes 10^{-7}$	$5.87 imes10^{-4}$
ATXN2L	16	28,834,356	28,848,558	$8.15 imes10^{-6}$	$5.37 imes 10^{-3}$	$1.03 imes10^{-6}$	$7.17 imes 10^{-4}$
CLN3	16	28,477,983	28,506,896	$5.82 imes10^{-8}$	$2.53 imes10^{-4}$	$2.43 imes10^{-8}$	$3.38 imes10^{-5}$
IL27	16	28,510,683	28,523,372	$2.10 imes10^{-7}$	$4.10 imes10^{-4}$	$3.93 imes10^{-8}$	$1.30 imes10^{-4}$
LAT	16	28,996,147	29,002,104	$1.53 imes10^{-4}$	5.40×10^{-2}	$4.62 imes10^{-7}$	$6.13 imes10^{-4}$
NFATC2IP	16	28,962,128	28,978,418	$1.93 imes10^{-8}$	$9.20 imes10^{-5}$	$8.07 imes10^{-8}$	$1.84 imes10^{-4}$
NUPR1	16	28,548,606	28,550,495	$6.39 imes10^{-8}$	$3.04 imes10^{-4}$	$1.94 imes 10^{-8}$	$4.69 imes10^{-5}$
SBK1	16	28,303,840	28,335,170	2.55×10^{-7}	$4.39 imes 10^{-4}$	7.70×10^{-7}	$7.45 imes 10^{-4}$

Gene Symbol	Chromosome	Gene Start (bp)	Gene End (bp)	P _{FUSION}	FDR _{FUSION}	P _{eMAGMA}	FDR _{eMAGMA}
SH2B1	16	28,857,921	28,885,526	$5.91 imes 10^{-9}$	$5.13 imes 10^{-5}$	$7.76 imes 10^{-7}$	$1.10 imes 10^{-3}$
SULT1A1	16	28,616,903	28,634,946	$1.54 imes 10^{-9}$	$8.17 imes10^{-6}$	$8.27 imes10^{-8}$	$1.25 imes 10^{-4}$
SULT1A2	16	28,603,264	28,608,430	$7.19 imes10^{-11}$	$7.63 imes10^{-7}$	$3.63 imes10^{-8}$	$7.79 imes10^{-5}$
TUFM	16	28,853,732	28,857,729	$2.12 imes 10^{-6}$	$2.05 imes 10^{-3}$	$4.10 imes 10^{-7}$	$7.41 imes 10^{-4}$
GSDMB	17	38,060,848	38,076,107	$6.69 imes10^{-5}$	$2.64 imes 10^{-2}$	$2.22 imes 10^{-5}$	$1.16 imes 10^{-2}$
IKZF3	17	37,921,198	38,020,441	$4.01 imes 10^{-5}$	$1.69 imes10^{-2}$	$2.70 imes 10^{-5}$	$9.35 imes10^{-3}$
ORMDL3	17	38,077,294	38,083,854	$4.99 imes10^{-5}$	$1.38 imes 10^{-2}$	$1.64 imes 10^{-5}$	$6.66 imes 10^{-3}$
TYK2	19	10,461,209	10,491,352	$5.12 imes 10^{-6}$	$4.05 imes 10^{-3}$	$1.94 imes10^{-7}$	$2.01 imes 10^{-4}$

Table 1. Cont.

22 novel gene were shown in bold. Gene position (start and end) are based on GRCh37/hg19 by Ensembl.

We further performed KEGG and GO enrichment analysis on the overlapped 33 significant genes on the Metascape website, respectively. For KEGG enrichment analysis, we totally found four significant KEGG pathways (Figure 2), including Th17 cell differentiation $(p = 5.83 \times 10^{-6})$, cytokine–cytokine receptor interaction $(p = 2.92 \times 10^{-4})$, spinocerebellar ataxia ($p = 5.11 \times 10^{-4}$), and Rap1 signaling pathway ($p = 1.55 \times 10^{-3}$). For GO enrichment analysis, we totally found ten significant GO terms (Figure 2), including signaling receptor complex adaptor activity ($p = 2.25 \times 10^{-5}$), apoptotic signaling pathway ($p = 2.09 \times 10^{-4}$), cytokine receptor binding ($p = 2.14 \times 10^{-4}$), regulation of blood pressure ($p = 1.16 \times 10^{-3}$), positive regulation of protein phosphorylation ($p = 1.18 \times 10^{-3}$), inflammatory response $(p = 2.44 \times 10^{-3})$, steroid metabolic process $(p = 2.66 \times 10^{-3})$, regulation of autophagy $(p = 5.96 \times 10^{-3})$, protein homodimerization activity $(p = 6.40 \times 10^{-3})$, and regulation of MAPK cascade ($p = 6.47 \times 10^{-3}$). In addition, chord graphs (Figure 3) were also depicted to show the most significant enrichment pathways of KEGG and GO and visualize the targeting relationship between significant genes and significant pathways, so as to visualize which pathways every gene is enriched to, and which genes are enriched in each pathway. PPI network (Figure 4) shows the interaction between proteins.







Figure 3. Chord graphs for four significant KEGG pathways and ten significant GO terms. (a) Chord graph of four significant KEGG pathways. (b) Chord graph of ten significant GO terms. For each panel, the right semicircle represented significant pathways or terms, and the left semicircle represented the genes enriched in these pathways or terms.



Figure 4. Protein–protein interaction (PPI) network for 33 overlapped genes by STRING. Each circle represents a protein, a line between proteins indicates PPI, line thickness indicates the strength of data support.

4. Discussion

In the present study, we performed a comprehensive large-scale gene-level analysis using co-complementary methods and successfully detected 33 common genes, including 11 previously reported genes such as TYK2, IL-6R, and Fas [7–9], and 22 novel potential genes such as IL-27, LAT, and MAGI3. Enrichment analysis suggested important role of pathways involving Th17 cell differentiation and Rap1 signaling pathway, followed by PPI network illustrating the protein–protein interactions. All these findings provide novel insights into the potential molecular mechanisms underlying the development of JIA.

TYK2, IL-6R, and Fas appear more frequently in the significant KEGG pathways and Go terms, and the expression products of these three genes have been shown to play important roles in the inflammatory and immune responses of JIA [21–23]. Tyrosine kinase 2 encoded by gene TYK2 is a part of janus kinase (JAK), which mediates the activation of signal transducers and activators of transcription (STAT) proteins. That is, TYK2 may play a role in autoimmunity and inflammation through abnormal expression in JAK-STAT pathway, thus leading to JIA [21,24,25]. IL-6R encodes part of the interleukin 6 receptor, as a pro-inflammatory cytokine, IL-6 is significantly elevated in the serum of JIA patients. Inhibition of IL-6R expression reduces IL-6 and IL-6 receptor binding, thereby reducing inflammation and immune responses in JIA patients [22,26]. Fas can induce T-cell apoptosis by binding to the Fas Ligand (FasL), so decreased gene Fas expression may lead to the accumulation of activated T-cells and cause autoimmune diseases [27].

IL-27 is involved in encoding the synthesis of IL-27, a cytokine that plays a role in innate immunity and whose primary function is to promote pro-inflammatory Th1 differentiation and inhibit anti-inflammatory Th2 responses [28–32]. IL-27 promotes Th1 cell differentiation, which in turn produces a large amount of the proinflammatory cytokine interferon- γ (IFN- γ) to play a pathogenic role in JIA [32,33]. The promotion of Th1 differentiation and the inhibition of Th2 differentiation by IL-27 is also dependent on the action of STAT1 [28,31–33], that is, the role of IL-27 is also involved in the JAK-STAT signaling pathway. Therefore, if the expression of gene IL-27 is inhibited, the inflammatory response of JIA patients can be correspondingly reduced.

LAT encodes a protein called T-cell activation adaptor [34]. T-cell receptor (TCR) signaling is an important process in T-cell development and its activation in the periphery [35]. LAT is a key signaling hub connecting TCRs to trigger downstream T-cell responses. If LAT gene expression is reduced, peripheral T-cell development and numbers are inhibited. Decreased numbers of T cells are prone to lead to immunodeficiency and autoimmune diseases [36], and patients with JIA are likely to have decreased autoimmune function due to lack of LAT.

MAGI3 encodes Membrane-associated guanylate kinase, WW and PDZ domaincontaining protein 3. Abnormal expression of MAGI3 may affect Notch signaling and thus affect bone and joint development in children [37]. In addition, MAGI3 is also a risk gene for rheumatoid arthritis (RA), Graves' disease, and other autoimmune diseases, indicating it can also cause JIA by affecting the human immune system [38,39].

Enrichment analysis suggested important role of pathways involving Th17 cell differentiation and Rap1 signaling pathway. The Th17 cell differentiation pathway is involved in inflammation and bone destruction, IL-27, IL-6R, LAT, TYK2 are all on this pathway. Th17 cells are actively differentiated and mainly secrete IL-17, which not only promotes the production of inflammatory cytokines in the JAK-STAT signaling pathway, but also catalyzes the maturation of osteoclasts, leading to osteopenia and joint damage [40,41]. The activation of inflammatory cytokines and osteopenia together lead to arthritis, and so inhibiting the differentiation of Th17 cells may inhibit and treat JIA to a certain extent [42–44].

The Rap1 signaling pathway plays a central role in the functional outcome of T-cell stimulation [45]. Rap1 is a T-cell receptor proximal signaling protein, and its abnormal expression may lead to abnormal T cells [46]. T cells activate macrophages and synovial stromal cells pleiotropically through cell-to-cell contact and interleukin production, leading to synovitis and joint destruction in RA [47,48]. The pathogenic behavior of the above T

cells is caused by Rap1 inactivation, and sustained Rap1 signaling in T cells can effectively reduce the incidence and severity of arthritis [49,50]. Therefore, activation and enhancement of Rap1 signaling also contribute to the prevention and treatment of JIA.

Our study is not without limitations. First, we only focused on European ancestry due to the current large-scale GWASs of JIA was only available for European population. The findings cannot be directly generalized to other ethnic population. Second, the results from data analysis are often less reliable than that from serious and high-cost experiments, which is likely to be the gold standard in biomedicine studies. However, the data analysis is still valuable. For example, it is often hard to pre-specify the experimental target under a hypothesis-free approach, data analysis can help to narrow down the candidate experimental target list and provide the evidence of target priority. In addition, the current analysis pipeline can be easily extended, an alternative way is to search for restriction endonuclease (RE) sites in the non-coding regions and gain insights through RE digestion patterns [51,52]. Third, our findings are obtained from a joint analysis of all subtypes of JIA, and there is no guarantee that the conclusions would be valid for any subtype. Finally, the results must necessarily be confirmed by experiments in the laboratory, given that all the analysis are essentially in silico.

5. Conclusions

In summary, we performed gene-level analysis as well as enrichment analysis on the largest GWAS summary data of JIA. We identified novel JIA-associated genes including IL-27, LAT, and MAGI3, and highlight the important role of Th17 cell differentiation, Rap1 signaling pathway in the development of JIA. Our results can provide new insights into the pathogenic mechanisms as well as potential therapeutic targets of JIA; however, further studies are still required to validate these findings.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms232113555/s1.

Author Contributions: Conceptualization, Z.Y. and S.W.; methodology, W.G. and S.L.; software, W.G. and S.L.; formal analysis, W.G. and S.L.; writing—original draft preparation, S.L.; writing—review and editing, Z.Y., S.W., W.G., S.L., L.L., and R.Y.; supervision, Z.Y. and S.W.; funding acquisition, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 81872712 and No. 82173624), the Natural Science Foundation of Shandong Province (No. ZR2019ZD02), the Cheeloo Young Talent Program of Shandong University, and the Taishan Scholar Project of Shandong Province, all awarded to Z.Y.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The GWAS summary data analyzed during the current study are available in the NHGRI-EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads/summary-statistics, accessed on 9 November 2021) (Study Accession Code GCST90010715). The FUSION software and GTEx v8 gene expression datasets used during the current study are available in the FUSION website (http://gusevlab.org/projects/fusion/, accessed on 8 April 2022). The eMAGMA software and GTEx v8 annotion datasets used during the current study are available in the eMAGMA website (https://github.com/eskederks/eMAGMA-tutorial, accessed on 28 March 2022).

Acknowledgments: We would like to thank GWAS Catalog for providing us with the JIA GWAS summary data.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Cimaz, R. Systemic-onset juvenile idiopathic arthritis. *Autoimmun. Rev.* **2016**, *15*, 931–934. [CrossRef] [PubMed]
- 2. Ravelli, A.; Martini, A. Juvenile idiopathic arthritis. Lancet 2007, 369, 767–778. [CrossRef]

- 3. Giancane, G.; Alongi, A.; Ravelli, A. Update on the pathogenesis and treatment of juvenile idiopathic arthritis. *Curr. Opin. Rheumatol.* **2017**, *29*, 523–529. [CrossRef] [PubMed]
- 4. Cobb, J.E.; Hinks, A.; Thomson, W. The genetics of juvenile idiopathic arthritis: Current understanding and future prospects. *Rheumatology* **2014**, *53*, 592–599. [CrossRef] [PubMed]
- 5. Hersh, A.O.; Prahalad, S. Immunogenetics of juvenile idiopathic arthritis: A comprehensive review. *J. Autoimmun.* 2015, 64, 113–124. [CrossRef]
- 6. Moncrieffe, H.; Prahalad, S.; Thompson, S.D. Genetics of juvenile idiopathic arthritis: New tools bring new approaches. *Curr. Opin. Rheumatol.* **2014**, *26*, 579–584. [CrossRef]
- Thompson, S.D.; Sudman, M.; Ramos, P.S.; Marion, M.C.; Ryan, M.; Tsoras, M.; Weiler, T.; Wagner, M.; Keddache, M.; Haas, J.P.; et al. The susceptibility loci juvenile idiopathic arthritis shares with other autoimmune diseases extend to PTPN2, COG6, and ANGPT1. Arthritis Rheum. 2010, 62, 3265–3276. [CrossRef]
- Hinks, A.; Cobb, J.; Marion, M.C.; Prahalad, S.; Sudman, M.; Bowes, J.; Martin, P.; Comeau, M.E.; Sajuthi, S.; Andrews, R.; et al. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat. Genet.* 2013, 45, 664–669. [CrossRef]
- 9. Lopez-Isac, E.; Smith, S.L.; Marion, M.C.; Wood, A.; Sudman, M.; Yarwood, A.; Shi, C.; Gaddi, V.P.; Martin, P.; Prahalad, S.; et al. Combined genetic analysis of juvenile idiopathic arthritis clinical subtypes identifies novel risk loci, target genes and key regulatory mechanisms. *Ann. Rheum. Dis.* **2021**, *80*, 321–328. [CrossRef]
- 10. Giancane, G.; Consolaro, A.; Lanni, S.; Davi, S.; Schiappapietra, B.; Ravelli, A. Juvenile Idiopathic Arthritis: Diagnosis and Treatment. *Rheumatol. Ther.* **2016**, *3*, 187–207. [CrossRef]
- 11. Tak, Y.G.; Farnham, P.J. Making sense of GWAS: Using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* **2015**, *8*, 57. [CrossRef]
- 12. Li, B.; Ritchie, M.D. From GWAS to Gene: Transcriptome-Wide Association Studies and Other Methods to Functionally Understand GWAS Discoveries. *Front. Genet.* **2021**, *12*, 713230. [CrossRef]
- Wainberg, M.; Sinnott-Armstrong, N.; Mancuso, N.; Barbeira, A.N.; Knowles, D.A.; Golan, D.; Ermel, R.; Ruusalepp, A.; Quertermous, T.; Hao, K.; et al. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* 2019, *51*, 592–599. [CrossRef]
- 14. Gusev, A.; Ko, A.; Shi, H.; Bhatia, G.; Chung, W.; Penninx, B.W.; Jansen, R.; de Geus, E.J.; Boomsma, D.I.; Wright, F.A.; et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **2016**, *48*, 245–252. [CrossRef]
- 15. Xu, J.; Ma, J.; Zeng, Y.; Si, H.; Wu, Y.; Zhang, S.; Shen, B. A Cross-Tissue Transcriptome-Wide Association Study Identifies Novel Susceptibility Genes for Juvenile Idiopathic Arthritis in Asia and Europe. *Front. Immunol.* **2022**, *13*, 941398. [CrossRef]
- De Leeuw, C.A.; Mooij, J.M.; Heskes, T.; Posthuma, D. MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* 2015, 11, e1004219. [CrossRef]
- 17. Gerring, Z.F.; Mina-Vargas, A.; Gamazon, E.R.; Derks, E.M. E-MAGMA: An eQTL-informed method to identify risk genes using genome-wide association study summary statistics. *Bioinformatics* **2021**, *37*, 2245–2249. [CrossRef]
- 18. Zhu, H.; Zhou, X. Transcriptome-wide association studies: A view from Mendelian randomization. Quant. Biol. 2020. [CrossRef]
- 19. Hahn, Y.S.; Kim, J.G. Pathogenesis and clinical manifestations of juvenile rheumatoid arthritis. *Korean J. Pediatr.* **2010**, *53*, 921–930. [CrossRef]
- 20. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A.H.; Tanaseichuk, O.; Benner, C.; Chanda, S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **2019**, *10*, 1523. [CrossRef]
- Tao, J.H.; Zou, Y.F.; Feng, X.L.; Li, J.; Wang, F.; Pan, F.M.; Ye, D.Q. Meta-analysis of TYK2 gene polymorphisms association with susceptibility to autoimmune and inflammatory diseases. *Mol. Biol. Rep.* 2011, *38*, 4663–4672. [CrossRef] [PubMed]
- 22. Yokota, S. Interleukin 6 as a therapeutic target in systemic-onset juvenile idiopathic arthritis. *Curr. Opin. Rheumatol.* **2003**, *15*, 581–586. [CrossRef]
- 23. Van Parijs, L.; Abbas, A.K. Role of Fas-mediated cell death in the regulation of immune responses. *Curr. Opin. Immunol.* **1996**, *8*, 355–361. [CrossRef]
- 24. Kyogoku, C.; Tsuchiya, N. A compass that points to lupus: Genetic studies on type I interferon pathway. *Genes Immun.* 2007, *8*, 445–455. [CrossRef]
- 25. Richter, M.F.; Dumenil, G.; Uze, G.; Fellous, M.; Pellegrini, S. Specific contribution of Tyk2 JH regions to the binding and the expression of the interferon alpha/beta receptor component IFNAR1. *J. Biol. Chem.* **1998**, *273*, 24723–24729. [CrossRef]
- 26. Akioka, S. Interleukin-6 in juvenile idiopathic arthritis. Mod. Rheumatol. 2019, 29, 275–286. [CrossRef]
- Tadaki, H.; Saitsu, H.; Kanegane, H.; Miyake, N.; Imagawa, T.; Kikuchi, M.; Hara, R.; Kaneko, U.; Kishi, T.; Miyamae, T.; et al. Exonic deletion of CASP10 in a patient presenting with systemic juvenile idiopathic arthritis, but not with autoimmune lymphoproliferative syndrome type IIa. *Int. J. Immunogenet.* 2011, *38*, 287–293. [CrossRef]
- 28. Yoshida, H.; Hunter, C.A. The immunobiology of interleukin-27. Annu. Rev. Immunol. 2015, 33, 417–443. [CrossRef]
- 29. Meka, R.R.; Venkatesha, S.H.; Dudics, S.; Acharya, B.; Moudgil, K.D. IL-27-induced modulation of autoimmunity and its therapeutic potential. *Autoimmun. Rev.* 2015, *14*, 1131–1141. [CrossRef]
- 30. Iwasaki, Y.; Fujio, K.; Okamura, T.; Yamamoto, K. Interleukin-27 in T cell immunity. Int. J. Mol. Sci. 2015, 16, 2851–2863. [CrossRef]
- 31. Aparicio-Siegmund, S.; Garbers, C. The biology of interleukin-27 reveals unique pro- and anti-inflammatory functions in immunity. *Cytokine Growth Factor Rev.* 2015, 26, 579–586. [CrossRef]

- 32. Raphael, I.; Nalawade, S.; Eagar, T.N.; Forsthuber, T.G. T cell subsets and their signature cytokines in autoimmune and inflammatory diseases. *Cytokine* **2015**, *74*, 5–17. [CrossRef]
- Takeda, A.; Hamano, S.; Yamanaka, A.; Hanada, T.; Ishibashi, T.; Mak, T.W.; Yoshimura, A.; Yoshida, H. Cutting edge: Role of IL-27/WSX-1 signaling for induction of T-bet through activation of STAT1 during initial Th1 commitment. *J. Immunol.* 2003, 170, 4886–4890. [CrossRef]
- 34. Zhang, W.; Sloan-Lancaster, J.; Kitchen, J.; Trible, R.P.; Samelson, L.E. LAT: The ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. *Cell* **1998**, *92*, 83–92. [CrossRef]
- Bacchelli, C.; Moretti, F.A.; Carmo, M.; Adams, S.; Stanescu, H.C.; Pearce, K.; Madkaikar, M.; Gilmour, K.C.; Nicholas, A.K.; Woods, C.G.; et al. Mutations in linker for activation of T cells (LAT) lead to a novel form of severe combined immunodeficiency. J. Allergy Clin. Immunol. 2017, 139, 634–642.e635. [CrossRef]
- Keller, B.; Zaidman, I.; Yousefi, O.S.; Hershkovitz, D.; Stein, J.; Unger, S.; Schachtrup, K.; Sigvardsson, M.; Kuperman, A.A.; Shaag, A.; et al. Early onset combined immunodeficiency and autoimmunity in patients with loss-of-function mutation in LAT. J. Exp. Med. 2016, 213, 1185–1199. [CrossRef]
- 37. Zuo, C.; Huang, Y.; Bajis, R.; Sahih, M.; Li, Y.P.; Dai, K.; Zhang, X. Osteoblastogenesis regulation signals in bone remodeling. Osteoporos. Int. 2012, 23, 1653–1663. [CrossRef]
- Liu, Y.Q.; Liu, Y.; Zhang, Q.; Xiao, T.; Deng, H.W. Identification of Novel Pleiotropic SNPs Associated with Osteoporosis and Rheumatoid Arthritis. *Calcif. Tissue Int.* 2021, 109, 17–31. [CrossRef]
- Kus, A.; Szymanski, K.; Peeters, R.P.; Miskiewicz, P.; Porcu, E.; Pistis, G.; Sanna, S.; Naitza, S.; Płoski, R.; Medici, M.; et al. The association of thyroid peroxidase antibody risk loci with susceptibility to and phenotype of Graves' disease. *Clin. Endocrinol.* 2015, *83*, 556–562. [CrossRef]
- Kotake, S.; Udagawa, N.; Takahashi, N.; Matsuzaki, K.; Itoh, K.; Ishiyama, S.; Saito, S.; Inoue, K.; Kamatani, N.; Gillespie, M.T.; et al. IL-17 in synovial fluids from patients with rheumatoid arthritis is a potent stimulator of osteoclastogenesis. *J. Clin. Investig.* 1999, 103, 1345–1352. [CrossRef]
- 41. Hueber, A.J.; Asquith, D.L.; Miller, A.M.; Reilly, J.; Kerr, S.; Leipe, J.; Melendez, A.J.; McInnes, I.B. Mast cells express IL-17A in rheumatoid arthritis synovium. *J. Immunol.* 2010, *184*, 3336–3340. [CrossRef]
- 42. Okamoto, K.; Takayanagi, H. Osteoclasts in arthritis and Th17 cell development. *Int. Immunopharmacol.* **2011**, *11*, 543–548. [CrossRef]
- 43. Schett, G.; Teitelbaum, S.L. Osteoclasts and Arthritis. J. Bone Miner. Res. 2009, 24, 1142–1146. [CrossRef]
- 44. Sato, K.; Takayanagi, H. Osteoclasts, rheumatoid arthritis, and osteoimmunology. *Curr. Opin. Rheumatol.* **2006**, *18*, 419–426. [CrossRef]
- 45. Remans, P.H.; Wijbrandts, C.A.; Sanders, M.E.; Toes, R.E.; Breedveld, F.C.; Tak, P.P.; van Laar, J.M.; Reedquist, K.A. CTLA-4IG suppresses reactive oxygen species by preventing synovial adherent cell-induced inactivation of Rap1, a Ras family GTPASE mediator of oxidative stress in rheumatoid arthritis T cells. *Arthritis Rheum.* 2006, 54, 3135–3143. [CrossRef]
- 46. Thomas, R.; Turner, M.; Cope, A.P. High avidity autoreactive T cells with a low signalling capacity through the T-cell receptor: Central to rheumatoid arthritis pathogenesis? *Arthritis Res. Ther.* **2008**, *10*, 210. [CrossRef]
- 47. Sakaguchi, S.; Benham, H.; Cope, A.P.; Thomas, R. T-cell receptor signaling and the pathogenesis of autoimmune arthritis: Insights from mouse and man. *Immunol. Cell Biol.* **2012**, *90*, 277–287. [CrossRef]
- Zhernakova, A.; Stahl, E.A.; Trynka, G.; Raychaudhuri, S.; Festen, E.A.; Franke, L.; Westra, H.J.; Fehrmann, R.S.; Kurreeman, F.A.; Thomson, B.; et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet.* 2011, 7, e1002004. [CrossRef]
- Abreu, J.R.; Krausz, S.; Dontje, W.; Grabiec, A.M.; de Launay, D.; Nolte, M.A.; Tak, P.P.; Reedquist, K.A. Sustained T cell Rap1 signaling is protective in the collagen-induced arthritis model of rheumatoid arthritis. *Arthritis Rheum.* 2010, 62, 3289–3299. [CrossRef]
- 50. Liu, G.; Jiang, Y.; Chen, X.; Zhang, R.; Ma, G.; Feng, R.; Zhang, L.; Liao, M.; Miao, Y.; Chen, Z.; et al. Measles contributes to rheumatoid arthritis: Evidence from pathway and network analyses of genome-wide association studies. *PLoS ONE* **2013**, *8*, e75951. [CrossRef]
- 51. Kalia, V.C.; Kumar, R.; Kumar, P.; Koul, S. A Genome-Wide Profiling Strategy as an Aid for Searching Unique Identification Biomarkers for Streptococcus. *Indian J. Microbiol.* **2016**, *56*, 46–58. [CrossRef]
- 52. Kalia, V.C.; Raju, S.C.; Purohit, H.J. Genomic analysis reveals versatile organisms for quorum quenching enzymes: Acylhomoserine lactone-acylase and -lactonase. *Open Microbiol. J.* **2011**, *5*, 1. [CrossRef]





Developing Genetic Engineering Techniques for Control of Seed Size and Yield

Intikhab Alam^{1,2,3,4}, Khadija Batool⁴, Yuanyuan Huang^{1,3}, Junjie Liu^{1,3} and Liangfa Ge^{1,3,*}

- ¹ Department of Grassland Science, College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China
- ² College of Life Sciences, South China Agricultural University, Guangzhou 510642, China
- ³ Guangdong Subcenter of the National Center for Soybean Improvement, College of Agriculture, South China Agricultural University, Guangzhou 510642, China
- ⁴ State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China
- * Correspondence: lge@scau.edu.cn

Abstract: Many signaling pathways regulate seed size through the development of endosperm and maternal tissues, which ultimately results in a range of variations in seed size or weight. Seed size can be determined through the development of zygotic tissues (endosperm and embryo) and maternal ovules. In addition, in some species such as rice, seed size is largely determined by husk growth. Transcription regulator factors are responsible for enhancing cell growth in the maternal ovule, resulting in seed growth. Phytohormones induce significant effects on entire features of growth and development of plants and also regulate seed size. Moreover, the vegetative parts are the major source of nutrients, including the majority of carbon and nitrogen-containing molecules for the reproductive part to control seed size. There is a need to increase the size of seeds without affecting the number of seeds in plants through conventional breeding programs to improve grain yield. In the past decades, many important genetic factors affecting seed size and yield have been identified and studied. These important factors constitute dynamic regulatory networks governing the seed size in response to environmental stimuli. In this review, we summarized recent advances regarding the molecular factors regulating seed size in Arabidopsis and other crops, followed by discussions on strategies to comprehend crops' genetic and molecular aspects in balancing seed size and yield.

Keywords: seed-specific transcription factors; signaling pathways; seed size; seed development

1. Introduction

The seed-producing crops are essential sources of foodstuffs, fodder, and fuel worldwide. The number and size of seeds are very imperative with respect to the evolution of various plant species for their preservation [1]. In addition, the seeds could be used to produce biofuels, which are becoming more popular as an alternative to fossil fuels [2,3]. Seeds fulfill almost 70% of the world's food demand. Environmental factors significantly affect plant development, consequently reducing seed production. Higher yields are possible with the entire primary and secondary branches, as well as the size and pod number on every plant, flowering time, seeds number per pod, seed weight, and sometimes plant height [4,5]. The seed development in angiosperm plants begins by merging only a single sperm with a single egg cell or beside two polar nuclei, resulting in a diploid embryo or a triploid endosperm, respectively [6]. An endosperm's single layer envelops the embryo in the mature seed of the Arabidopsis, and the seed coat encloses the endosperm, which is designed by many layers of particular maternal tissues that provide safety, enhance latency, and sprouting [7]. The development of zygotic tissues including embryo and endosperm as well as maternal components (e.g., eggs) determines seed size. Thus, the development of the embryo, endosperm, and seed envelope are the major constituents in crops, determining

Citation: Alam, I.; Batool, K.; Huang, Y.; Liu, J.; Ge, L. Developing Genetic Engineering Techniques for Control of Seed Size and Yield. *Int. J. Mol. Sci.* 2022, 23, 13256. https://doi.org/ 10.3390/ijms232113256

Academic Editor: Wajid Zaman

Received: 29 August 2022 Accepted: 15 October 2022 Published: 31 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

the seed size and weight [8]. The seed development process is divided into two phases: the morphogenesis phase, which includes cell division, endosperm and embryo development, and cotyledon differentiation; and the maturing phase, which includes embryo development at the expense of endosperm, seed dehydration, and source collection [9]. Arabidopsis seed development programs are similar to those of dicotyledon types of seed crops, i.e., canola and soybean. Moreover, early phases of seed development are identical in monocots and dicots but vary in later stages [10]. To inhabit various types of habitations, seed plants produce an enormous range of growth farms and many deviations in seed size, accompanied by seed travel ways. Moreover, after a long evolution time, plants exhibit most of the seed sizes, starting from the dust-like seeds of orchids to the 20 kg of the double coconut [11]. Seed size in flowering plants is very important from an evolutionary point of view, as well as essential for crops [12], but the molecular mechanism of seed size is not clear. Seed size is the main component in determining plant fitness and seed yield [11–13]. In plants, variation in seed size is an interesting phenomenon in developmental biology. Seed size is controlled by the inherent inner material regarding maternal tissues and zygotic materials [11], while the progress of the seed is affected by environmental and climatic aspects. Given the significance of seeds as the primary food source, many efforts have been made to understand seed development, size regulation, and total yield in crops. Many crop genomes have been sequenced, and genetic factors have been identified to dissect the network controlling seed growth and development [14,15].

Genetic screening is a powerful strategy widely used to identify the regulators of seed size. The advances in genome-wide association studies (GWAS) and RNA sequencing technology have been extensively applied to investigate the fundamentals of the seed size variation [16]. Plants have the good factor of huge genetic variations occurring naturally, as well as being extensively studied in many species [17]. Seed size is among the most complex characteristics, elaborately controlled by multiple positive and negative factors, constituting a well-balanced network to regulate seed size by modulating biological processes. Genes with a higher population and more density markers and several quantitative trait loci (QTLs) with a novel QTL for grain length, qGL11, were recognized in seven panicle and grain-related traits [18]. In the present review, we summarized recent advances in dissecting the regulatory network that controls seed size and yields to an extent, including the genetic factors and signaling pathways that regulate seed size in Arabidopsis and other various crops.

2. Genetic Factors Controlling Seed Size

Controlling seed size and weight is determined via the genetic composition of zygotic and maternal tissues. In recent times, many scientific studies have been reported to identify genetic factors with plant responses to control seed size and weight (Table 1).

Species	Genes or Mutants	Function	References
Arabidopsis	IKU1, IKU2, and MINI3	Early cellularization of endosperm produces smaller seeds	[19,20]
Arabidopsis	TFL1	Timely endosperm cellularization enhances seed size	[21]
Arabidopsis	SHB1	Up-regulates the functions of <i>IKU2</i> and <i>MINI 3</i> and postpones endosperm cellularization, producing larger seeds	[22,23]
Arabidopsis	TTG2	Identifies integument cell elongation and endosperm growth as the primary regulators of seed size	[24]
Arabidopsis	$TOP1\alpha$ and $UPF1$	The molecular interplay among these three genes explains their biparental gametophytic effect on seed size regulation	[25]
Arabidopsis	LuLEA1	Negatively regulates seed size and yield	[26]
Arabidopsis	AN3-YDA	Disrupts normal sucrose and glucose contents and alters seed size regulation in <i>an3</i> or <i>yda</i> mutants	[27]

Table 1. List of reported genes that control seed size and yield in different plants.

Species	Genes or Mutants	Function	References
Arabidopsis	EIN ₃ -YDA	Both are integral to a sugar-mediated metabolism cascade that regulates seed mass by maternally controlling embryo size	[28]
Arabidopsis	DME	Enhances DNA <i>de</i> -methylation activity and endosperm chromatin composition	[29]
Arabidopsis	CYCB1;4	Overexpressed in transgenic plants to show increased seed size in response to faster cell cycle progression to increase grain size and yield	[30]
Arabidopsis	FIS2, FIS3, FIE, MEA, and MSI1	Control endosperm development	[31]
Arabidopsis	AGL62	Regulates endosperm development by the FIX	[32]
Arabidopsis	PHE1	Controls seed abortion by mea mutant	[33]
Arabidopsis	CDKs and CYC	Control cell cycle development	[34]
Arabidopsis	RBR	Decreases the cell proliferation in the integuments and stops further development of seed coat	[35]
Arabidopsis	AtENO2	Decreases the size of seed and mass with the declined concentration of cytokinin	[36]
Arabidopsis	MET1 and DDM1	DNA methylation of maternal and paternal genome regulates seed size	[37]
Arabidopsis	FIS2, FWA, and MPC	Regulate parental imprinting by DNA methylation	[38–40]
Arabidopsis	MEA	It plays the role as a maternal buffer of paternal effects on seed development	[41]
Arabidopsis	PRC2	Controls turgor pressure during seed development	[42]
Arabidopsis	AtSOB3	Produces heavy and larger seeds with long hypocotyl	[43]
Arabidopsis	p4-siRNAs	Expressions are observed in maternal gametophytes and continue until endosperm development	[44]
Arabidopsis	CEPR1	Promotes reproductive parts including seed size and yield	[45]
Arabidopsis	CYCD3:1 and CYCD7:1	Promote early endosperm and embryo development	[46,47]
Arabidopsis	AtbZIP75	Regulates seed development	[36]
Rice	OsEMF2a	Delayed cellularization and autonomic endosperm development	[48]
Rice	GLW7	Controls grain size and yield	[49]
Rice	OsGW2	Regulates grain size and reduces rice yield	[50]
Rice	GS3	Regulates grain length	[51]
Rice	OML4	Negatively regulates seed size by terminating cell expansion	[52]
Rice	qGL7	Could display a single mendelian characteristic, suggested as producing diverse seed shapes	[53]
Rice	qGL11	Regulates rice grain weight and length	[18]
Rice	MIS2	Controls grain size by coordinately regulating epidermal cell size and cell number	[54]
Rice	MADS78 and MADS79	Regulate endosperm cellularization and early seed formation in rice	[55]
Maize	miRNAs	Control embryogenesis via transcriptional regulation	[56]
Maize	GSI, qZmGW2-CHR4, zMgs3, and qZmGW2-CHR5	Control kernel size and grain production	[57–59]
Maize	$ZmBZR_1$	It attaches to the promoter region of <i>GRACE</i> and <i>KRP6</i> to control their expression and regulate seed size	[60]
Wheat	TaGW2	Controls grain weight	[61]
Cotton	MicroRNA160	Regulates smaller seed size	[62]
Bean	SL9.1 GA	Controls seed length	[63]
Rapeseed	BnaEOD3	Negatively regulates the seed growth and development	[64]

Table 1. Cont.

Species	Genes or Mutants	Genes or Mutants Function	
Soybean	<i>GmSWEET10a</i> and <i>GmSWEET10b</i>	Change seed size, oil quantity, and protein contents	[15]
Soybean	qSL-13-3 _{zy} , -13-4 _{zy} , and qSW-13-4 _{zy}	Control seed size and yield	[65]
Soybean	GmKIX8-1	Increases cell proliferation and size of aerial plant organs such as leaves and seeds with high expression of <i>CYCLIN D3;1-10</i>	[66]
Peanut	SNE and SAP	Regulate seed and fruit size	[67]

 Table 1. Cont.

In Arabidopsis, early endosperm cellularization induces diploids to pollinate tetraploids, ultimately producing smaller seeds. By contrast, late or failed endosperm cellularization causes highly abortive seeds [68,69]. The endosperm cellularization time also regulates seed size, though embryo development covers the seed cavity afterward by replacing the endosperm space. The mutant genes of the IKU-pathway, including haiku (iku1), iku2, and miniseed3 (mini3), developed smaller seeds due to precocious cellularization of the endosperm [19,20,70,71], while the SHORT HYPOCOTYL UNDER BLUEI (SHB1) gain-of-function mutants postponed endosperm cellularization, ultimately producing larger seeds [19,22,23]. Transcriptional co-activators, such as SHB1 linked with IKU2 and MINI3 promoters and stimulate their expression in Arabidopsis [22,23]. Furthermore, adjacent tissues influence endosperm development; for instance, a mutation of maternal sporophytic in TRANSPARENT TESTA GLABRA2 (TTG2) limits integument enlargement and the origins of advanced endosperm cellularization [24]. The IKU pathway interacts with TTG2 genetically by way of double mutants ttg2 and iku2, showing stronger seed phenotypes than single mutants [24]. Furthermore, a DNA topoisomerase, TOPOISO-MERASE I α (TOP1 α), and an ATP-dependent RNA helicase, UP-FRAMESHIFT SUPPRES-SOR 1 (UPF1), biparentally regulate the seed size (regulating TTG2). Loss of function of UPF1 or TOP1 α or induces the ectopic appearance of TTG2 in antipodal cells. Genetic analysis has consistently shown TOP1 α and UPF1 function upstream of TTG2. TTG2 is directly suppressed by TOP1 and UPF1 by causing a chromatin disruption [25]. Further, it has been suggested that comparative TTG2 quantity in antipodal cells to gametes regulates seed size, indicating the role of the maternal and paternal dose-dependent molecular framework. The late embryogenesis abundant (LEA) protein family named LuLEA is expressed at later stages of seed development [26]. LuLEA1 negatively regulates seed size and yield as its higher expression reduces seed size and fatty acid content in Arabidopsis. Recently, it was reported that Arabidopsis TERMINAL FLOWER1 (TFL1) acts as a mobile controller produced in the chalazal endosperm and promotes endosperm cellularization on time when moving to the syncytial peripheral endosperm, and enhances seed size by stabilizing ABA insensitive 5 (ABI5) genes [21]. Most recently, in rice, mutations in the EMBRYONIC FLOWER2a (OsEMF2a) gene resulted in delayed cellularization and autonomic endosperm development [48].

Another research group hypothesized that the seed size and length of the hypocotyl might be regulated by AtSOB3, which belongs to the SUPPRESSOR OF PHYTOCHROME B. They tested this hypothesis and demonstrated that AtSOB3-D produced lighter seeds with a shorter hypocotyl compared to WT, while a dominant-negative mutation, *AtSOB3-6-OX*, produced heavy and larger seeds with a longer hypocotyl in Arabidopsis [43]. Researchers have reported that the impact of accumulated soluble sugar on the seed size is controlled by the ANGUSTIFOLIA3-YODA (AN3-YDA) gene cascade as well as the addition of environmental and/or metabolic factors by sugar and ethylene metabolism to regulate seed mass in Arabidopsis by interacting with ETHYLENE INSENSITIVE 3 (EIN3) [28] and AN3 [27]. The ENHANCER3 OF DA1 (EOD3) exists in the sporophytic tissues of the mother plant, encodes many cytochrome P450/CYP78A6, and stimulates seed growth in plants. Overexpression of EOD3 has been shown to drastically increase the size of the seed, whereas loss-of eod3-ko function mutants produce smaller seeds. Mutants of the *BnaEOD3*

gene were competently produced with stable changes by the stable transformation of the CRISPR/Cas9. These mutations were steadily transferred into the T1 and T2 generations, resulting in an accumulation of homozygous mutants with a combined loss of function alleles. The T1 line contains smaller siliques and seeds, but the higher number of seeds per silique suggests that BnaEOD3 negatively regulates seed growth and development in rapeseed [64].

Rice is one of the most essential crops, feeding a significant amount of the world's population. Many years of studies have focused on identifying QTLs that control the grain size of rice [72]. For example, GLW7 is a major QTL that has been identified as controlling grain size and yield [49]. However, the investigation of the genetic variability of Arabidopsis has significant potential to provide insight into commercially essential traits in crops. Ren et al. (2019) identified new factors responsible for controlling seed size; CYCB1;4 is a cell cycle-related gene that encodes a cyclin protein. Growing regions are enriched in CYCB1;4 and may control the cell cycle that is positive in maternal and zygotic parts, thereby increasing seed and other organ sizes, implying that CYCB1;4 is an essential component in Arabidopsis for regulating seed size. They reported comprehensive information on GWAS in Arabidopsis related to seed size, with 38 important linked loci, and one locus interlinked with CYCB1;4. A higher level of CYCB1;4 was noted in transgenic plants, which increased seed size and grain yield in response to faster cell cycle progression, while the CYCB1;4 mutant produced smaller seeds. In short, CYCB1;4 may potentially target yield improvement in plants as the temporal and spatial expression pattern may inform functioning in both maternal and paternal tissues that are involved in absolutely organizing seed size [30]. Another study reported that overexpression of OsGW2, a QTL on chromosome 2, changes grain size and reduces Indica rice yield. The downregulation of OsGW2 using RNAi technology resulted in larger and plump kernels by regulating cell expansion and cell proliferation in the spikelet hull [50].

Genomic imprinting regulation is very complicated and may include non-coding RNAs, DNA methylation, and histone adaptations. Many scientific studies have evaluated that the endosperm plays an essential role in divergent plant tissues, with imprinted gene expression and a unique DNA methylation [73]. The increased time for the development of endosperm between different seed parts was considered to be under the epigenetic regulator. Nevertheless, it was shown that the maternally expressed in embryo 1 (mee1) gene is imprinted in the embryo and endosperm of the maize plant [74]. Genome-wide methods relating to RNA-seq of mutual crosses in Arabidopsis [75], rice [76], and maize [77], have indicated the occurrence of numerous tentative imprints [74], and the genetic makeup of embryos. Parental imprinting of genes is caused by distinctive DNA methylations at specific loci produced by DEMETER demethylase (DME) in the maternal gametes [78], DNA Methyltransferase 1 (MET1) in the paternal gametophyte, and POLYCOMB REPRESSIVE COMPLEX 2 (PRC2)-facilitated suppression of transcription [29]. Though DNA methylation is not primarily responsible for embryo-specific imprinting, allele-definite siRNAs may induce parent-of-origin and precise genetic expression occupied by the embryo. Hypomethylated regions in the endosperm generate maternal-specific siRNAs, which monitor ROS1-mediated DNA de-methylation of the particular area after moving into the embryo. Sequence-specific siRNAs distinguishing maternal alleles individually could result in allele specificity [79].

CURLY LEAF (*CLF*), which encodes a histone methyltransferase of PRC2 coordinated in tri-methylation of histone H3 Lys 27 (H3K27me3), controls a group of genes that interact with each other during embryo development [80]. Several PcG proteins such as FERTILIZATION INDEPENDENT SEED 2 (FIS2), FIS3, FERTILIZATION INDEPENDENT ENDOSPERM (FIE), FIS1/MEDEA (MEA), and MULTI COPY SUPPRESSOR OF IRA1 (MS11) control the development of the endosperm [31]. By way of methylation of chromatin histones, repressive complexes are formed by these PcG proteins that inhibit the expression of involved genes controlling developmental processes. Numerous PcG constituents can be imprinted, and the functions of FIS complexes in plants are mani-

fested after pollination. Moreover, *fis* mutants decrease endosperm cellularization and, ultimately, the size of seeds. The function of FIS is mediated by AGL62, which belongs to the agamous-like MADS-box proteins, whereas in *fis* mutants, AGL62 expression is suppressed, leading to endosperm cellularization failure [32]. Regarding the FIS gene family, FLOWERING WAGENINGEN (FWA), MEA, and MEA genes are maternally expressed; however, MADS-box I like PHERES1 (PHE1) and a FIS PcG complex are presumed to be paternally expressed. Seed growth is terminated due to peak propagation of the endosperm without embryo development in *medea* (*mea*) mutants [81]. The FIS proteins, including MEA, FIE, and FIS2, strongly regulate the MADS-box gene PHE1 expression. PHE1 expression was normal after fertilization in wild-type plants (WT), whereas *fis* mutants showed higher expression until seed abortion, indicating that FIS-class proteins deregulate PHE1 expression after fertilization to prevent seed abortion [33]. Recently, using CRISPR/Cas9-based genome editing, it was found that the MADS-box TF genes MADS78 and MADS79 are very important for regulating endosperm cellularization and early seed formation in rice [55]. A single MADS78 or MADS79 knockout mutant displayed early endosperm cellularization, while double mutants hindered seed development and produced no viable seeds [55].

DNA methylation plays a significant role in seed development as it may help increase seed size and weight [82]. It also plays an important part in plant genome stability and developmental processes [83]. Chinese super hybrid rice LYP9 (Liangyou Peijiu) is a successful hybrid rice variety with high heterosis [84]. This indicates that DNA methylation in hybrid rice seeds plays a key function in the initiation and maintenance of heterosis [85]. Xiao et al. (2006) showed that methylation of FIS2 and MEA is facilitated by MET1 containing DNA methylase, methylating CpG islands, and a genomic cytosine base. A met1 mutation in the maternal genome increases seed size due to maternal hypomethylation and delays endosperm cellularization [37]. By contrast, *met1* mutations in the paternal genome lead to early cellularization of the endosperm and decreased seed size, thus confirming exclusive maternal control of *MEA* expression, which provides insights into the evolution of imprinting in plants. Similar findings were also obtained in reciprocal crosses with parents with mutations in *ddm1*. FIS2 [38], FWA [39], and MPC [40], which are imprinted genes that regulate parental imprinting by DNA methylation [86]. However, the effects of paternal imprinting are minor or non-existent. Another study in Arabidopsis revealed that a deviation in the paternal alleles regulates the growth of seeds, and these deviations occur in seeds when mutants with the maternal *mea* allele showed loss of function of this gene, indicating that paternal MEA may compensate maternal MEA deficiency during seed development [41]. During the developmental phase of rice seeds, genome-wide DNA methylome analyses showed higher frequencies of DNA methylation in embryos than in the endosperm, and non-transposable element regions were more variable than transposable ones during endosperm and embryo development [87]. Many epigenetic regulatory mechanisms participate in seed development; however, the regulatory mechanisms underlying major genome imprinting processes are not yet clear. Thus, integrated exploration of genome-related DNA methylome differences, histone amendment, transcriptomics, and cleavage positions of miRNA in developing seeds through current genomic approaches, including degradome and bisulfite sequencing, is required for a better understanding of epigenetic mechanisms underlying seed development. The turgor pressure of the endosperm drives seed expansion during the early seed development [88], whereas later, turgor pressure decreases during endosperm cellularization, which is controlled by POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) [42].

MicroRNAs (miRNAs) have become an important part of the basic molecular network of important agronomic seed traits in various plant species. The loss of the specific gene functions that control miRNA biogenesis results in defects in seed development [89]. MicroRNA160 targeting auxin response factors in cotton and suppression of *miR160* resulted in a smaller seed size in cotton [62]. Another study examined the potential role of miR156a in linseed flax (*Linum usitatissimum* L.) by decreasing seed oil content in transgenic lines compared to WT plants [90]. In addition, miRNAs also control embryogenesis by regulating transcription in maize [56]. Recently, 72 existing miRNAs and 39 recently identified miRNAs were discovered to be expressed in the growing seeds of legumes (*Phaseolus vulgaris* L.), and their involvement was more prominent throughout late embryogenesis and desiccation. Numerous TF families have been identified as known miRNA targets, including ARF, HD-ZIP, NF-Y, and SPL, and a majority of new miRNA targets were anticipated to be expressed as functional proteins [91]. In addition, small interfering RNA (siRNAs) can also induce transcriptional gene silencing, which may be associated with epigenetic control of seed expansion. Plant-specific RNA polymerase named RNA polymerase IV (Pol IV) is involved in the production of siRNAs [92]. Genome-related monotonous elements are used to produce p4-siRNAs, and some are interlinked with exclusive regions, comprising numerous characterized imprinted loci that may have a significant function in the initiation or continuation of imprinted gene expression [93]. Pol IV-base p4-siRNAs are produced from transposon elements present in the genome and are expressed exclusively in the endosperm tissues [44].

Quantitative trait locus analysis is useful for providing information on seed size control and other traits which may be relevant in plant breeding. Recently, 39 QTLs were identified on different chromosomes in common beans, and one major QTL, SL9.1 ^{GA}, was found to control seed length [63]. Three epistatic QTLs (qSL-13-3_{zy}, qSL-13-4_{zy}, and qSW-13-4_{zy}) were found to regulate soybean seed size and yield [65], and many well-categorized QTLs have been recognized in different plants, including *GS3*, *qSW5*, and *qGL7* in rice [51,53], *GSI*, *qZmGW2-CHR4*, *zMgs3*, and *qZmGW2-CHR5* in maize [57–59] and *TaGW2* in wheat [61].

Moreover, many QTLs involved in seed size control have been detected in other crops, although their functions have not been categorized [94-96]. Genes with higher abundance and density markers and 43 QTLs were recognized to be linked with seven panicle- and grain-related characteristics. For example, the new locus qGL11 enhances rice grain weight and length, and it occurs in a chromosome segment substitution line (CSSL) obtained through segregation of the population; moreover, it mapped exceptionally well onto a 25-kb sequence that comprises the IAA-amido-synthetase gene OsGH3.13 [18]. A recent study investigated QTL regions at the 168.37 kb chromosome fragment with regard to the SNPs Aradu_A07_1148327 and Aradu_07_1316694. Twenty-two genes were predicted in this region, with two genes of interest, Aradu.RLZ61 and Aradu.DN3DB, encoding a transcriptional regulator, an F-box SNEEZY (SNE), and STERILE APETALAlike (SAP), respectively. These genes play a critical role in regulating seed and fruit size [67]. Moreover, MINI SEED 2 (MIS2) regulates grain size by coordinately controlling epidermal cell number and cell size, and microscopic analysis showed that mis2 mutant rice revealed reduced cell size of spikelet epidermis but higher cell numbers [54]. The MIS2 encodes CRINKLY4 (CR4), the receptor-like kinase through Map-based cloning, which exhibited the strongest expression during panicle development [97].

In rice, the LARGE1 gene encodes the MEI2-LIKE PROTEIN4 (OML4) and negatively regulates seed size through terminating cell enlargement in spikelet hulls. Over-expression of OML4 results in smaller and lighter seeds, whereas loss of function of OML4 leads to larger and heavier grains [52]. In eukaryotes, cyclin-dependent kinases (CDKs) and their governing subunit cyclin (CYC) control the cell cycle progression [34]. CDK inhibitors (CDKIs) are inhibitors of CDK/CYC complex action and were thus the principal cell cycle controllers. In plants, two types of CKIs have been discovered and categorized: SIAMESE/SIAMESE-RELATED (SIM)-related suppressors and INHIBITOR OF CDC2 KINASE/KIP-RELATED (ICK/KPR) [98]. Central-cell rbr mutants have superfluous nuclei, indicating failed cell cycle arrest [99]. Moreover, irregularities in *rbr1* maternal gametes decrease cell propagation in the integument and stop further development of the seed envelope [47]. Furthermore, a complex of MUTICOPY SUPPRESSOR OF IRAI (MSII) RBR1, which stops the expression of *MET1*, is responsible for activating imprinted genes present in germ cells throughout megagametophyte development [35]. Seed termination was observed in the tri-mutant cycd3-1:2:3 (CYCD3 belongs to the CYCD gene family), because of late cell propagation during embryo development. However, cell division
stimulated by the trans-activation of CYCD3:1 or CYCD7:1 in the endosperm and embryo results in fatal embryo deficiencies [100]. Directed upregulation of CYCD7:1 in the endosperm and in the central cell, which resulted in escape from the cell cycle in the central cell, and enhanced endosperm propagation at the syncytial stage, resulting in a larger seed. Nevertheless, death occurred in a few larger seeds, indicating an incomplete seed termination [46]. The ENO gene is strongly expressed in plants encoding glycolytic metalloenzyme enolase (ENO2) that boosts the dehydration of 2-phospho-D-glycerate (2-PGA) to phosphoenolpyruvate (PEP) [101]. Three ENO genes have been identified in Arabidopsis, including At1g74030 (ENO1), At2g36530 (ENO2), and At2g29560 (ENO3) [102]. Sub-cellular occurrence of isoforms of enolase showed that ENO1 and ENO3 are situated in the chloroplast and cytoplasm, while ENO2 has been detected in the nucleus and cytosol [103]. In addition, mutations in AtENO2 decreased the size and mass of seeds with a reduced concentration of cytokinin. Carbohydrate data analyses and RNA sequencing showed that metabolism pathways, particularly regarding the secondary metabolism, occurred in AtENO2 T-DNA mutants [36]. Furthermore, AtENO2 cooperated with Arabidopsis basic-leucine zipper 75 (AtbZIP75) as a substitute for AtMBP-1 [36]. Therefore, AtbZIP75 may contribute to seed development. This study clarified the novel function of AtENO2 in seed growth and improvement and proposed a good target gene for gene manipulation for the purpose of plant breeding. Recently, two SWEET homologs, i.e., GmSWEET10a and GmSWEET10b, were found to simultaneously affect seed size, oil quantity, and protein content in soybean, suggesting that seed size and oil constituents can be controlled by adjusting the combination of both alleles [15]. A different study demonstrated that the reproductive functions of this plant, including fertility, seed size, and yield, can be controlled by C-TERMINALLY ENCODED PEPTIDE RECEPTOR1 (CEPR1) [45]. The authors showed that two cepr1 knockdown mutants produced smaller seeds and reduced yield (by 88–98%). In Arabidopsis, BZR_1 controls organ size via the BR signaling pathway. Recently, studies have shown that overexpression of $ZmBZR_1$ in Arabidopsis causes increased organ and seed size. Moreover, ZmBZR₁ is attached to the promoter of GRACE and KRP6 to control their expression and regulate seed size [60].

3. The Role of Signaling Pathways in Seed Size Regulation

3.1. The Ubiquitin-26s Proteasome Pathway

The ubiquitin–26S pathway regulates the size and development of seeds; the genes regulating the ubiquitin–proteasome pathway are shown in Figure 1A. Ubiquitination and/or protein degradation processes by the proteasome 26S pathway are an important posttranslational protein turnover mechanism in plants [104–106]. The ubiquitin-proteasome system, which is highly controlled, is involved in the regulation of all aspects of plant life and governs numerous developmental and stress-related activities mediated by hormone signals [107]. Ubiquitin (Ub) is a protein of 76 amino acids that affect the target cellular proteins via a multistep reaction including three enzymes, E1 enzyme (Ub-activating), E2 enzyme (Ub-conjugating), and E3 enzyme (Ub ligases), in the ubiquitin–proteasome system (UPS) [108]. Ubiquitination specificity is largely regulated by E3 ligases, which recognize specific substrates and catalyze the bond between the Ub and substrates [108]. E3 ligases are important regulatory components in the ubiquitin-dependent pathway because they mediate substrate specificity [109]. The ubiquitination pathway participates in seed size regulation; ubiquitin receptor, E3 Ubiquitin ligase, ubiquitin-specific protease, 26S proteasome, plant-specific APC/C regulatory factors, and ubiquitination pathway interaction proteins, etc., all play a central function in regulating seed size and development [13,110,111]. In Arabidopsis, ubiquitin receptors such as DA1 and DA1-related protein (DAR1) regulate the size of seeds by restraining cell propagation in the maternal integument [112]. The DA2, E3 Ub-ligases, and BIG BROTHER (BB)/ENHANCER OF DA1 (EOD1) function together with DA1 to limit seed size and the development of other plant parts [110,113], indicating that E3 ubiquitin ligases and DA1 may function in the same complex or have mutual downstream targets. Supporting this concept, DA1 strongly

co-operates with DA2 in vitro as well as in vivo [110]. E3 ubiquitin ligase shows specific substrate relationships between DA1 and DA2, and it may help to identify the precise substrates of DA1 with regard to degradation. Genetic studies have revealed that BB/EOD1 and DA2 function in diverse ways [110], indicating that variant substrates may be targeted by them for degradation. In rice, grain width and weight 2 (GW2) shows sequence resemblance with DA2, encoded by a QTL regulating seed size [13]. GW2 negatively affects the size of grains by controlling cell propagation in spikelet casings. Homologs of GW2 are also associated with seed size regulation in maize and wheat crops [58,61], suggesting that GW2 retains preserved functions of seed size regulation. In Arabidopsis, SUPPRESSOR OF DA1 (SOD2) encoding UBIQUITIN-SPECIFIC PROTEASE 15 (UBP15) is maternally responsible for enhancing seed development. Plants with sod2/ubp15 mutations show reduced seed size and other organs because of the reduction in cell propagation and epistatic effects on DA-1 [114,115]. In Arabidopsis, DA1 interacts physically with UBP15 and modifies its durability, suggesting that UBP15 is the downstream target of DA1. Further study showed that UBP15 individually affects EOD1 and DA2 to regulate grain size, specifying that other unidentified E3 ubiquitin ligases may ubiquitinate UBP15. Moreover, in Arabidopsis, the anaphase-promoting complex/cyclosome (APC/C) ubiquitin ligase and the 26S proteasome, along with many other factors such as RPT2 and SAMBA, affect seed development [116–118]. SAMBA mutations, a negative regulator of APC/C, result in a larger seed size [117]. The mutation in SAMBA markedly increases seed size and structure in eod1-2 da1-1 phenotypes [119], implying that SAMBA may exert substantial effects or act as a complex identical to EOD1 and DA1 regarding grain size regulation. The important QTL qSW5/GW5 is responsible for seed thickness; it is located on chromosome 5 and controls the size of grains by suppressing cell propagation in spikelet casings in rice [120]. GW5 reacts to ubiquitin in vitro, signifying that it may act as a ubiquitin-proteasome pathway [121]. In rice, GW5 may function separately from GW2 to regulate grain width [122]. Histone ubiquitylation controls the transcription of DA1/DA2, which affect seed size. OTU1 de-ubiquitinase de-ubiquitylates DA1/DA2 chromatin and acts as an epigenetic transcriptional suppressor of the DA1/DA2 genes. OTU1 is nucleocytoplasmic and plays a role in nuclear and cytoplasmic functions [123]. SMALL LEAF AND BUSHY1 (SLB1) encodes the F-box protein, a fragment of the SKP1/Cillin/F-box E3 Ub-ligase complex. SLB1 regulates the growth of plant organs and secondary branches by regulating the stability of BIG SEEDS1 (BS1), leading to increased leaf and seed size in soybean and Medicago truncatula [124]. A different study reported that the regulatory complex GW2-WG1–OsbZIP47 regulates grain size in rice [27]. GW2, WG1, and OsbZIP47 work together to control grain width and length through the GW2–WG1–OsbZIP47 regulatory module. Specifically, WG1 encodes a glutaredoxin protein and promotes cell proliferation, leading to enhanced grain growth. WG1 acts as a co-repressor with ASP1 and interacts with the transcription factor (TF) OsbZIP47 to terminate its transcription. OsbZIP47 suppresses seed growth by reducing cell proliferation [28]. Moreover, the E3 Ub-ligase TaPUB1 reduces the seedlings' sensitivity to the abscisic acid (ABA) by interacting with TaPYL4 and TaPY15 (involved in ABA signal transduction and inducing their degradation), resulting in smaller grain size and yield; this suggests that TaPUB1 is a negative regulator of seed development [125].



Figure 1. Signaling pathways and their correlated molecules involved in seed size regulation. (**A**) Ubiquitin–26s proteasome pathway (**B**) MAPK signaling pathway (**C**) IKU pathway (**D**) G-protein signaling pathway (**E**) Transcription regulatory factors.

3.2. The Mitogen-Activated Protein Kinase (MAPK) Signaling Pathway in Plants

MAPK is responsible for many transduction pathways, including hormone signaling and stress responses [126–128]. The genes regulating the MAPK signaling pathway are shown in Figure 1B. A MAPK cascade comprises three kinases: MAPK, MAPK kinase (MAPKK), and MAPKK kinase (MAPKKK). MAPKKKs activate and phosphorylate MAPKKs in response to an external stimulus signal, and then the activated MAPKs phosphorylate a variable number of downstream target substrates, including transcription factors, chromatin remodeling factors, kinases, or different enzymes, resulting in transcriptome and proteome reprogramming in the entire cell. The successive phosphorylation of MAPK proteins and their substrates is critical for MAPK cascade-mediated interactions and signal transduction [129]. Several components of the MITOGEN-ACTIVATED PRO-TEIN KINASE (MAPK) cascade have previously been identified as key regulators of seed development [130]. The small grain 1 (smg1) encoding MAPKK4 mutants generate short grains in rice because of reduced cell proliferation in spikelet exteriors [131]. In addition, a previous study suggested that alterations in OsMAPK6 result in short grains, which resemble those produced by smg1 mutants [132]. Furthermore, loss of function mutations in OsMKP1 produce larger grains, while OsMKP1 overexpression generates smaller grains. OsMKP1 regulates rice grain size by limiting cell proliferation in grain hulls. OsMKP1 directly interacts with and deactivates OsMAPK6 [133]. Moreover, OsMKK4 is physically co-related with OsMAPK6 [132]. Therefore, the OsMKK4–OsMAPK6-controlling module plays an essential role in controlling seed size in rice. It is difficult to investigate the upstream OsMAPKKKs of OsMKK4 and downstream substrates of OsMAPK6 regarding the

control of grain size. Furthermore, OsMKK4 and OsMAPK6 affect brassinosteroid (BR) reactions and many BR-related genes' expression [131,132], suggesting a link between the MAPK pathways and BRs with respect to seed development. Guo et al., 2018 demonstrated that GSN1 functions as a negative regulator of the OsMKK10–OsMKK4–OsMPK60 cascade by inducing precise dephosphorylation of OsMPK6 to coordinate the trade-off between grain size and grain number per panicle [134].

The leucine-rich repeat (LRP)-receptor kinases ER, ERL1, and ERL2 stimulate fruit growth through signaling pathways regulated by EPFL9 in the carpel wall. EPFL2 expression controls the spacing of ovules in the wall of the carpel, and in inter-ovule spaces, this is controlled by ERL1 and ERL2, which may facilitate equal distribution of resources [135]. This may provide an understanding of trade-offs in rice panicle expansion and constitute a basis for increasing crop yield. The novel AGC protein kinase AGC1-4, belonging to the subfamily of AGC VIIIa, encodes a serine-threonine kinase and regulates seed size. Seeds with higher expression of AGC1-4 are smaller, whereas *agc1-4* mutant seeds were significantly larger [136]. A recent study reported that OsINV3 is a positive seed size regulator in rice [137]. They showed that overexpression of OsINV3 increased grain size, while loss of its function reduced grain size, in comparison with WT plants. OsINV2 is a homolog of OsINV3 and has no function in seed size regulation by itself; however, both OsINV2 and OsINV3 play roles in sucrose metabolism in sink organs and increase seed size [137]. In Arabidopsis, the function of the novel gene ZmRLK7, which belongs to the LRP receptorlike protein kinases (LRR-RLKs) isolated from maize, was investigated through ectopic expression to examine its effects on plant development [138]. Most recently, in Arabidopsis, the lectin receptor-like kinase LecRK-VIII.2 has been found to regulate seed production by organizing seed size, seed number, and silique. Lecrk-VIII.2 mutants produce fewer seeds but more seeds and siliques, resulting in a higher yield. On the other hand, overexpressing of LecRK-VIII.2 produces larger seeds but fewer seeds and siliques, resulting in yields comparable to wild-type plants [139]. MPK3 and MPK6 double mutants are toxic to the embryo, but single MPK6 mutants exhibit a variety of abnormal phenotypes related to seeds, similar to those seen in mutants of the kinases MAPK KINASE 4 (MKK4) and MKK5, which function upstream of MPK3 and MPK6, respectively. Given the importance of MPKs in seed development and their function in RLK signaling. (Xiao et al., 2021) investigated whether MPK3 and MPK6 could act downstream of LecRK-VIII.2 [139]. All these findings suggested that the phosphorylation of MPK6 and not MPK3 corresponds with the expression levels of LecRK-VIII.2 in the seed. According to (Xiao et al., 2021), LecRK-VIII.2 promotes growth throughout development. Furthermore, this important receptor positively promotes seed growth from maternal sporophytic tissues during seed development by activating expansions, and possibly through the MAPK cascade and MPK6. This discovery adds to the body of evidence supporting the importance of LecRKs in developmental processes and opens up new paths for advanced research, such as investigating the signal that LecRK-VIII.2 perceives. Understanding the mechanism behind the greater yield in LecRK-VIII.2 KO plants, despite an overall loss in plant growth, will be fascinating. Is it feasible that a plant with lower vegetative biomass has more available resources for seed production before it reaches senescence? Is it possible to manipulate this signaling pathway to generate crops with larger seeds, stronger plants, and higher yields?

3.3. The IKU Pathway

Seed size controlled through an IKU signaling pathway loss-of-function mutation in the VQ motif protein HAI-KU1 (IKU1), receptor kinase IKU2, WRKY TF MINI-SEED 3 (MINI3), and the leucine-rich repeat (LRR) decreased seed sizes due to advanced endosperm cellularization. The genes regulating the IKU pathway are shown in Figure 1C. The embryo and endosperm genome, instead of maternal genetic factors, determine the size of the seed phenotype in these mutants [20,22]. Transcriptional co-activators such as SHORT HYPOCOTYL UNDER BLUE 1 (SHB1) have been shown to link with IKU2 and MINI3 promoters and stimulate their expression in Arabidopsis [22,23]. Seed size is regulated by

endosperm growth through the signaling pathways of IKU1, IKU2, SHB1, and MINI3 [140]. MINI3, which is associated with the cytokinin oxidase (CKX2) promoter stimulates the initiation of CKX2 expression and is responsible for controlling endosperm growth [110]. Higher expression of CKX2 enhances seed size in iku2 seeds, confirming the role of cytokines in down-regulating the IKU pathway with regard to seed size regulation [110]. ABSCISIC ACID-INSENSITIVE 5 (AB15) is a TF that directly attaches to the upstream region of SHB1 and inhibits its expression [141]. ABA controls the development of seeds by ABI5-facilitated transcription and is regulated by SHB1 in the endosperm; thus, the mutants abi5 and ABA-deficient 2 (aba2) abi5 produce the larger size of seeds [141]. In combination with cytokinin and ABA signaling, the IKU pathway affects endosperm growth to control seed size in Arabidopsis and other plants. In Arabidopsis, the function of the new gene ZmRLK7, which belongs to the LRP receptor-like protein kinases (LRR-RLKs) isolated from maize, was investigated using ectopic expression [138]. The roles of WRKY10/MINI3 and *MAPK10* in endosperm development were examined, and they appeared to function in opposite patterns. Furthermore, *mapk10* mutants consistently produce larger seeds, and seed size is positively regulated by WRKY10/MINI3 [142].

3.4. The G-Protein Signaling Pathway

In plants and animals, G-protein signaling regulates many functions related to growth and development. Molecules regulating G-protein signaling are shown in Figure 1D. Many cell surface G-protein-related receptors are connected to intracellular effectors by the G-protein complex. Ligands activate receptors, and effectors control different cellular responses [143]. G-protein complexes comprise Ga, Gb, and Gg subunits, and G-proteincoupled pathways transfer signals through membrane-bearing receptors and heterotrimeric compounds to downstream effectors. Thus, mutations in Ga (GPA1) or Gb (AGB1) result in small leaves and flowers in Arabidopsis [144,145]. Over-expression of the Arabidopsis atypical Gg (AGG3) stimulates the growth of seeds and organs by enhancing cell propagation, and loss-of-function mutations in AGG3 lead to short seeds and organs [146,147]. AGG3 overexpression in *Camelina sativa* resulted in increased seed size, confirming the function of AGG3 in the regulation of seed size [148]. Likewise, loss of function or functional suppression of rice Ga (RGA1) or Gb (RGB1) reduces rice grain size [149–151]. Similarly, the knockdown of the RGB1 gene in rice delayed seed development, and lowered seed weight and starch accumulation [152]. Two QTLs affecting the size of grains and panicles are Rice GRAIN SIZE 3 (GS3) and DENSE AND ERECT PANICLE 1 (DEP1), respectively [146,147]. However, long seeds are produced after the loss of function due to increased cell propagation, whereas short grains are produced in response to GS3 or DEP1 gain-of-function alleles [153–155]. It may be possible that AGG3 in Arabidopsis and GS3 and DEP1 in rice have various cofactors or effectors, suggesting differential effects on seed size. A minor QTL, Small grain 3 (SG3) encoding an R2R3 MYB protein is associated with a major QTL grain size 3 (GS3) and negatively controls rice grain length. The γ subunit of a G protein encoded by GS3 competitively interacts with G β , leading to reduced grain length [156,157]. DEP1 also encodes a γ subunit of a G protein, and GS3 appears to act as a cofactor of OsMADS1 in the control of grain size through the regulation of common target genes [158]. Recently, a study introduced a serine-threonine protein kinase AGC1-4 belonging to the subfamily of AGC VIIIa [136]. AGC1-4 overexpression causes a reduction in seeds, whereas agc1-4 mutants produce significantly larger seeds compared with WT plants. AGC1-4 regulates seed size by regulating the number of embryonic cells [136]. Recently, heterotrimeric G-protein mutants of rice were produced using CRISPR/Cas9 gene editing. The gs3 and dep1 mutants produced more favorable agronomic traits than WT plants, whereas the *rga1* mutation resulted in a dwarf phenotype, leading to an extreme reduction in grain yield. The heterotrimeric G-protein β subunit, RGB1, plays a significant role in plant development, and *rgb1* mutants show suppressed growth and development of the embryo [159]. Most recently, the E3 ligase gene Chang Li Geng 1 (CLG1) has been shown to negatively regulate grain length by targeting the $G\gamma$ protein GS3, and thus control grain

size. CLG1 overexpression increased grain length, and CLG1 mutations with variations in three essential amino acids decreased grain length. CLG1 directly interacts with and ubiquitinates GS3, which is then destroyed via the endosome degradation pathway, resulting in improved grain size [160]. More studies are required to comprehensively elucidate the functions of G-protein signaling in seed size regulation.

3.5. Transcriptional Regulation

Particular transcription factors, i.e., TRANSPARENT TESTA GLABRAA2 (TTG2) and APETALA 2 (AP2), affect seed development in Arabidopsis [24,161–164]. Genes regulating transcription factors are shown in Figure 1E. The transcriptional repressor NGATHA-like (NGAL-2)/SUPPRESSOR-OF-DA1 (SOD7) and its homolog NGLA3/DEVELOPMENTAL-RELATED-TARGET-IN-THE-APEX 4 (DPA4) have been predicted to strongly limit Arabidopsis seed development [165], and sod7-2 dpa4-3 mutants showed enhanced cell proliferation in the maternal integument, resulting in big seeds [165]. The Arabidopsis AINTEGU-MENTA (ANT) gene is associated with integument and organ growth regulation [166,167], and it encodes APETALA2-like TF ant mutants that exhibit smaller leaves, reduced flowering, imperfections in integument origination, and development of ovules. However, Arabidopsis and tobacco plants with higher expression of ANT showed higher production of leaves and flowers because of stimulated cell proliferation, and consequently, they produced larger seeds. An-1 encodes a basic helix-loop-helix protein, which regulates cell division. Transgenic studies have confirmed that An-1 positively regulates awn elongation, but negatively regulates grain number per panicle [168]. APETALA2 is a member of the AP2/EREBP (ethylene-responsive element binding protein) group belonging to the family of TFs that plays a significant role in the specification of floral organs in Arabidopsis [169], and it is responsible for seed size regulation [161,163]. The production of seeds is higher in Arabidopsis *ap2* mutants than in WT plants, regardless of the genotype of the pollen donor, signifying that AP2 is a maternal factor controlling seed size [161]. Tissue-specific epigenetic mechanisms control gene expression and transcriptional regulators at specific stages, and epigenetic processes allow genomic imprinting; thus, the allele's expression after pollination depends on the parent of origin. The "parental conflict theory can describe the importance of biological imprinting". A substantial barrier prevents resource distribution from mothers to offspring, and the main function of the paternal genome is to quickly deliver maternal resources to embryos harboring the paternal genome. However, the maternal genome attempts to distribute resources equally between offspring by downregulating the effects of the paternal genome [170]. Further, ap2-10(+/-) seeds generated by an *ap2* mutant are bigger than WT seeds, even in flowers pollinated with WT pollen, but seeds are smaller compared to ap2-10 (-/) seeds of mutant flowers pollinated with ap2 pollen [163], indicating that AP2 functions maternally and zygotically to regulate seed size. In support of this concept, seeds yielded by cross-pollination of 35S:AP2 antisense transgene mutants with WT flowers were bigger than WT seeds and smaller than ap2 mutant seeds [163]. For instance, SMOS1 (SMALL ORGAN SIZE 1) encodes an auxin-regulated APETAL2-type transcription factor and is positively regulated by OsARF1 in rice [171]. The loss of function of SMOS1 causes pleiotropic developmental phenotypes, including small seed size [171,172], suggesting a crucial role for SMOS1 in seed size regulation.

In *transparent testa glabra2* (*ttg2*) mutants, yellow seeds are produced due to insufficient proanthocyanidin production and deposition of mucilage in the seed coat by the mutant [24]. The *ttg2* mutants' seeds are round and small, causing reduced cell length in the integument, in contrast to WT seeds. Reciprocal crossing trials showed the effects of *ttg2* on seed size in a strongly maternal sporophyte. The *ttg2* mutation also induced advanced endosperm cellularization, resulting in endosperm size reduction. *TTG2* encodes a WRKY family TF which regulates many steps of tannin synthesis [162], and, as several products of the tannin synthesis pathway may affect the cell wall and its capability to increase, mutations in *TTG2* may reduce the elongation ability of the cell wall [24]. The *ALM1* gene encodes Golden

2-like (GLK) a member of the GARP subfamily of Myb TFs, and *alm1.g* and *alm1.a* mutants revealed a decrease in 100-grain weight by 15.8% and 23.1%, respectively [173,174].

Several genes, including GROWTH-REGULATING FACTOR (GRF) genes, encode DNA binding transcription factors that interact with and form a functional transcriptional complex with the transcription cofactor GRF-INTERACTING FACTOR (GIF) [175]. In this functional unit, GIF operates to recruit SWI/SNF chromatin remodeling complexes to their target genes so that they can be transcriptionally activated or inhibited by GRF. GRF expression is post-transcriptionally inhibited by microRNA (miR396) [175].

It is precisely the miR396–GRF/GIF module that influences a wide range of essential plant growth and development traits that could have agricultural implications. However, the most convincing results demonstrating the agronomical values of the miR396– GRF/GIF system come from a set of many studies that independently identified a rare naturally occurring allele of OsGRF4 from *Oryza sativa* landraces with larger grain size in various genetic backgrounds [176–181]. Overexpression of OsGIF1 also improved grain size [179,180,182–184]. Likewise, the expression of a mimic transgene that binds and inactivates miR396 in rice, a strategy also used in Arabidopsis and other species, can increase the yield. Together with the fact that effects similar to those were artificially introduced in Arabidopsis, these findings demonstrate the transferability of knowledge from models like *A. thaliana* to crops, and the general importance of the miR396–GRF/GIF module in other plant species.

The NACs are transcription factors (TFs) that are specific to plants and are involved in many developmental processes [185]. For example, NAC-like TF EjNACL47 might be linked with the larger organ size in triploid loquat. Furthermore, ectopic expression of EjNACL47 results in larger organs i.e., leaves, flowers, and siliques in Arabidopsis, implying a positive role in organ enlargement [186]. In rice, the ONAC020, ONAC026, and ONAC023 genes encoding NAC TFs express specifically during seed development. They have a strong relationship with seed size or weight as sequence changes in the upstream regulatory region [187]. The KIX–PPD complex regulates maternal integument growth and, consequently, regulates seed size through cell propagation and development. In Arabidopsis, the transcription factor MYC3/4 reacts with PPD1/2 and KIX8/9 to form the KIX–PPD–MYC product. The KIX–PPD–MYC product suppresses the expression of the GIFI promoter when interacting with the G-box region located in the GIFI promoter, and controls seed size [188]. Promoters of the grain filling-specific TF gene Opaque2 (O2), which regulates the factors known as B3 domain TF and ZmABI19, directly fuse with the O2 promoter for trans-activation and affect the developmental process of the endosperm and embryo, resulting in smaller seeds [189].

The KIX domain-containing protein family acts as a negative regulator of cell propagation in plants [190]. Loss of function in GmKIX8-1 mutants significantly increases the size of above-ground organs, including leaves and seeds, with increased cell propagation and CYCLIN D3;1-10 expression. In addition, molecular analysis of soybean germplasms showed that increased expression of the qSw17-1 QTL with a large seed phenotype is caused by decreased expression of GmKIX8-1 [66]. A different study reported that a regulatory complex, GW2–WG1–OsbZIP47, regulates rice grain size. GW2, WG1, and OsbZIP47 together regulate grain width and length through the GW2–WG1–OsbZIP47 regulatory module. Specifically, WG1 encodes a glutaredoxin protein and promotes cell proliferation, leading to enhanced grain growth. WG1 acts as a co-repressor with ASP1 and interacts with the transcriptional factor OsbZIP47 to terminate its transcription. OsbZIP47 suppresses seed development by reducing cell proliferation [28]. During development, the shape of the grain is controlled by many genes that interact with one another. GW8 is a member of the SBP transcription factor family. When it is overexpressed, it promotes cell growth and grain filling, which leads to wider grains and higher yields. In addition, GW8 also interacts with the GL7/GW7 promoter, inhibiting transcription and regulating cell proliferation in spikelet glumes [191,192]. Furthermore, GLW7 and GW8 both positively control grain hull cell size, increasing grain length and yield. Small and round seed 5 (SRS5) with a single

amino acid mutation (p.Arg308Leu) decreases rice grain length by inhibiting cell elongation. Although GLW7 can interact with SRS5 to regulate grain size, the exact mechanism remains unclear. Furthermore, the novel gene *OrMKK3* affects morphology and grain size, suggesting a relationship between MAPK and BR-responsive pathways with regard to grain development [193]. Moreover, LARGE2, encoding the HECT E3 UB-ligase OsUPL2, controls the rice panicle size and grain numbers, and *large2* mutants showed increased panicle size and number of seeds per panicle [194].

4. Seed Size Regulation by Phytohormones

Phytohormones control a vast range of developmental and physiological functions in plants. Furthermore, phytohormone levels fluctuate at various stages in different tissues during seed development, playing a significant role in the processes involved [195]. Multi-hormonal regulation, mediated via auxins, cytokinins, brassinolides, and jasmonic acid, plays an important role in endosperm propagation and embryo growth. Phytohormones and related molecules are schematized in Figure 2.



Figure 2. Phytohormones and their correlated molecules involved in seed size regulation: (**A**) auxins, (**B**) cytokinins, (**C**) brassinosteroids, and (**D**) jasmonic acid.

4.1. Auxins

Auxin, a plant hormone, has been linked to a wide range of characteristics of plant growth, including seed production [196]. The key auxin in all plants is indole-3-acetic acid (IAA). It is biosynthesized from the tryptophan (Trp) amino acid in a two-step mechanism which is highly conserved throughout plants [197,198]. First, the amino group is removed from Trp in the first phase, which is catalyzed by transaminases from the TRYPTOPHAN AMINOTRANSFERASE OF the ARABIDOPSIS (TAA/TAR) family, resulting in indole-3pyruvate (IPA). YUCCA is known to upregulate the auxin biosynthesis pathway and the indole-3-pyruvic acid (IPA) pathway. Previously, 11 members of YUC genes have been reported in Arabidopsis and are involved in the development of the basal body during embryogenesis [199]. Throughout embryogenesis, YUCI and YUC4 are expressed in a variety of cell types, and their expression coincides with that of YUC10 and YUC11 in developing seeds [199]. LEAFY COTYLEDON (LEC), BABY BOOM (BBM), and PLETHORA (PLT) transcription factors are known to bind or control auxin-associated genes under normal conditions, which stimulate somatic embryo development. In addition, ectopic expression of LEC2 promotes YUC2 and YUC4 expression early in the seedling somatic embryo development [200], whereas LEC1 ectopic expression promotes YUC gene expression during 2,4-D-induced somatic embryogenesis from immature zygotic embryos (YUC1, YUC4, and YUC10) and seedlings (YUC10) [201,202]. Auxin response factors (ARFs) control auxin-responsive genes' expressions in plants [203]. ARFs are B3-like transcription factors that directly bind to the auxin-responsive element (AuxRE) and activate or repress auxinresponsive genes in plants. In Arabidopsis, twenty-three ARF proteins, including ARF2 and ARF8, have been suggested as funneling auxin signals, to control seed size by reducing cell division [204,205]. In Arabidopsis, twenty-three ARF proteins, including ARF2 and ARF8, have been suggested as funneling auxin signals, to control seed size via reducing cell division [204,205]. ARF3/ETTIN mutations cause significant polarity abnormalities in the gynoecium, including apical tissue over-proliferation and ovary development [206]. ARF6, ARF7, and ARF8 promote cell growth, and hypocotyl elongation [207]. NtARF8 is involved in NtTTG2-regulated seed development in tobacco (Nicotiana tabacum L.) and supports seed quantity in collaboration with NtARF17 and NtARF19 [55]. Another study identified a B3 transcriptional factor, Arabidopsis maternal effect embryo arrest 45 (MEE45), a downstream regulator of auxin biosynthesis in the ovule, which controls seed size maternally and regulates cell proliferation via the transcriptional activation of aintegumenta (ANT). In this case, the largest seed size was observed for MEE45 expression, whereas the mee45 mutant expression resulted in reduced seed size [208]. Auxin also regulates the growth of siliques, which is noteworthy. Overexpression of BnaA9.CYP78A9, a member of the P450 monooxygenase gene family, stimulates the elongation of *Brassica napus* siliques by causing a significant increase in auxin [209]. However, most recently the modified expression of TaCYP78A5 has been shown to increase grain weight and yield per wheat plant by accumulating auxin [210]. Hence, the aforementioned studies based on transgenic and mutant studies imply that the auxin biosynthesis pathway, signaling, and transporter genes are essential in regulating the auxin levels during seed development, controlling embryogenesis, endosperm development, and seed coat development.

4.2. Cytokinins

Cytokinins (CKs) are phytohormones involved in the modulation of cytokinesisrelated enzyme activity and CK-mediated signaling. Additionally, CKs are important for initial endosperm cellularization [211]. The discovery of CK-catabolizing genes, in addition to their distinctive genetic and biochemical properties, has introduced the role of CKX enzymes in CK regulation [212], as well as multi-gene families synthesizing CKX enzymes. Various studies have aimed to determine the role of CKX family members (GFMs) in improving grain yield in many crops, including barley [213,214], rice [159,215], and wheat [216,217]. A transcription factor basic helix-loop-helix, cytokinin growth regulators (CKG) promote CK-mediated regulation of cell expansion and cell cycle development in Arabidopsis. CKG expression has been noted to enhance the size of cotyledons during the reproductive stage, whereas expression of ckg mutants has been noted to mediate the opposite effect [218]. Another study elucidated that CK activity is primarily regulated by the transcription factor receptors ARR1 and AHK3. Similarly, increased levels of CK in the external layer of reproductive tissues and the placenta lead to larger siliques and increased ovule formation, thereby increasing the number of seeds and resulting in higher seed yield [219]. CKs have also been observed as methylthiolated derivatives (MET); however, details of the function of this group of compounds are scarce [220]. The level of CKs during seed development is controlled by balancing the activities of isopentenyl transferase (IPT) and CK oxidase/dehydrogenase (CKX), which play an important role in the biosynthesis and degradation of CKs, respectively. AtIPT4 and AtIPT8 are expressed at higher levels during seed development in the chalazal endosperm area, during morphogenesis [221,222]. Furthermore, Song et al., 2015 suggested that silique and growing seeds have the ability to synthesize CK [223]. The endosperm-specific expression of IPT increases the levels of CK to grow seeds and seed mass without causing any morphological aberrations in transgenic

tobacco plants [224]. Numerous studies have reported that higher CK levels occur during development, upon external application, and that ectopic exposure of IPT is linked with increased CK degradation resulting from increased CKX gene expression [225,226]. Furthermore, the expression of the CKX and IPT GFMs suggests a positive correlation between the two in wheat [227] and *Brassica napus* [223].

CKs have a role in transcription control, and different genes in Arabidopsis influence the activation of its receptor histidine protein kinases (AHKs). AHKs activate AHPs, which are histidine phosphotransferase proteins, in response to CKs, and AHPs stimulate Arabidopsis response regulators (ARRs), thereby inducing the expression of all genes. Triple mutants, ahk-1,2,3 and ahp-1,2,3, have been observed to enlarge the embryo and seeds [221]; moreover, the same study concluded that cytokinin independent 1 (CKI1) expression plays an important function in CK signaling. CKI1 contains a histidine kinase-deficient CK perception domain of AHK. The cki mutants' expression results in fewer seeds that are larger in size, thereby suggesting CKI1 as playing a function in determining seed number and size during seed development [228]. Moreover, Li et al., 2013, elucidated that the expression of cytokinin oxidase 2 (CKX2), which encodes a protein that induces CK degradation, controls endosperm development [89] and up-regulates the CKX2 expression, is controlled by IKU2 and H3K27m3 [86]. These experimental findings explain that CK expression controls seed development phases through members of the CK signaling pathway and other epigenetic approaches [229]. Another study identified 17 GmCKX GFMs in soybean, and natural alterations were probed among cultivars with different yields. A total of 5 out of 17 CKX genes were noted to be responsible for the regulation of CK content during the grain-filling stage, as indicated by the results of introducing single nucleotide polymorphisms (SNPs) in them. GmCKX7-1 was discovered to contain a non-synonymous mutation (H105Q) on histidine 105, one of the amino acid residues in the active site, during the critical grain filling period, to preserve structural reliability of the enzyme, consequently enhancing seed yield [66].

4.3. Brassinosteroids

Brassinosteroids (BRs) are a recently documented class of phytohormones that play significant roles in plant growth and development, such as cell elongation and division [230–232]. BRs mainly take part in regulating the yield determination of agriculture traits, including seed size regulation [233]. In Arabidopsis and rice, molecular analysis of BR-defective mutants permitted the identification of very important genes involved in the BR-mediated regulation of seed development. The DEP1 promoter has been revealed to increase rice grain size and yield by optimizing the expression of BR-like genes [234,235]. Three steps are involved in the synthesis of BR from campesterol: (1) formation of campestanol from campesterol; (2) two concurrent routes from campestanol to castasterone (CS); and (3) B-ring lactonization of CS to 24-epibrassinolide (BL). OsDWARF (OsBR6ox) is a crucial gene in the biosynthesis of BR. The OsDWARF product catalyzes a late stage in bioactive BR formation via combining the early and late C-6 oxidation pathways and encodes BR-6-oxidase, which can convert 6-deoxotyphasterol (TY) and 6-deoxoCS to TY and CS, respectively [236,237]. The OsDWARF loss of function produces a decrease in TY and CS, resulting in the BR-deficient phenotype. Further, dwarf mutant shrink1-D (shk1-D) exhibited a lower level of BRs, which caused reduced seed length in Arabidopsis. In the case of shk1-D mutants, CYP72C1 overexpression and hydroxylation of BRs resulted in low endogenous BR levels and decreased seed size [235]. Furthermore, putative brassinolide receptor brassinosteroid in-sensitive1, a defective mutant, also caused lower seed production [226]. Presumably, this receptor activated the expression of genes encoding BRs that act as positive regulators of seed size (HAI-KU2. MINI3 and SHB1) and inhibited that of genes encoding negative regulators of seed size (ARF2 and AP2) through the TF brassinazole resistant 1 (BZR1) [238]. Higher expression of BR synthesis-related gene OsD11, or that of BR signaling factor OsBZR1 induced an increased sugar-accumulation rate in developing seeds and increased grain yield in rice. By contrast, the knockdown of these genes cleared the imperfect pollen, seed size, and weight reduction compared with the control [239]. In the BR

signaling pathway, cytochrome P450 mutant key enzyme D2 (D11) [234,240], BR-deficient dwarf 1 (BRD1) [236], DWARF 2 (D2) [241], and other genes related to BR biosynthesis, including BRD2 [242], all have very close phenotypes, such as decreased plant height, shoot length, spikelet length, and grain size. Furthermore, GW5 also interacts with and suppresses the activity of glycogen synthase kinase 2 (GSK2), a crucial kinase involved in BR signaling. The BZR1 and DLT transcription factors, downstream of GSK2, enter the nucleus in an un-phosphorylated state to modulate BR signaling-responsive genes, thereby affecting grain size in rice [120,226]. Moreover, GW5 and GW5-related proteins physically bind to the important components of the BR signaling pathway, GSK2, and BIN2, resulting in the accumulation of un-phosphorylated DLT and OsBZR1 [243]. Additionally, GSK2 binds to and phosphorylates members of the OVATE family protein 8 (OFP8) [244]. Furthermore, OFP8, OFP14, and the transcriptional activator GS9 interact with one another. OFP14 inhibits GS9 transcriptional activity, and GS9 overexpression causes circular grain shapes with BR-defective phenotypes. As a result, GS9 appears to negatively regulate the BR signaling [245]. In yeast two-hybrid screens, OFP3 binds to both GSK2 and DLT and negatively regulates the BR response, in addition to OFP8 and OFP14 [246]. Plants with the dss1 mutation were dwarf and showed erect leaves and smaller seeds. Similar results were observed in BR-deficient mutants [247]. In addition, SERK2 has been reported as a BR signaling element for many agronomic traits, especially yield and stress-related traits. Moreover, SERK2 knockdown mutants showed enhanced rice grain size [248].

4.4. Jasmonic Acid

Jasmonic acid (JA), a phytohormone produced by free α -linolenic acid, is similar to animal prostaglandins [249,250] and serves a variety of functions in plant response to biotic and abiotic stressors as well as development [226,251,252]. JA influences pollen formation or pollen shedding [253], embryo and seed development [254,255], spikelet formation [256], and maize sex determination [257–259], during inflorescence development. Linolenic acid is the substrate of JA biosynthesis, which is then metabolized to produce the bioactive isoleucine conjugate, JA-Ile [250]. The interaction of JA-Ile with the COI1 receptor (CORONATINE INSENSITIVE1) activates JA signaling, resulting in the proteasomal degradation of JAZ (JASMONATE ZIM-DOMAIN) transcriptional repressor proteins [260,261]. JAZ repressor degradation releases a number of transcription factors, especially MYCs, which interact with MED25/PFT1 to negatively regulate seed size [262]. The eg1 (extra glume 1) JA biosynthetic mutant displayed aberrant spikelet formation. The gene EG1 encodes a plastid-targeted class I lipase, which is required for the production of JA precursors. Furthermore, eg2-1D showed disrupted floral identity as well as impairment in floral meristem determination [256]. Map-based cloning showed that EG2 encodes for OsJAZ1, a JAZ repressor. Another comprehensive investigation revealed that the spikelet defects in eg2-1D are caused by the inhibited function of the OsMYC2-controlled E-class gene, OsMADS1. In addition, OsJAZ11 regulates the width and weight of rice seeds. When compared to wild-type, transgenic rice lines overexpressing OsJAZ11 showed up to a 14% increase in seed width and a 30% increase in seed weight. The constitutive expression of OsJAZ11 had a significant impact on spikelet development, resulting in additional glumelike structures, an open hull, and an abnormal number of floral organs. Also, transgenic lines accrued increased JA levels in spikelets and developing seeds. Overexpression lines exhibited altered expression of JA signaling and MADS-box genes when compared to WT [263]. According to yeast two-hybrid and pull-down experiments, OsJAZ11 interacts with Os-MADS29 and OsMADS68. Surprisingly, in overexpression lines, the expression of OsGW7, an important negative regulator of grain size, was dramatically decreased. Furthermore, multi-seeded 3 (msd3) sorghum mutants that can quadruple grain number per panicle by expanding panicle size and modifying floral development such that all spikelets are viable and set grain, in turn, decreased the JA levels [263].

5. Possible Strategies Used to Maintain Seed Size

Increasing grain yield is always an important goal in plant breeding programs, especially in the case of domesticated crops. Additionally, increasing seed size without affecting seed number via conventional plant breeding programs has allowed only limited progress because of the trade-off between these two yield components [264]. Many researchers have indicated that modern technology, including transgenic technology, genome editing, marker-assisted selection (MAS), and genomic selection (GS), has major challenges regarding the disruption of grain yield in different crops. The use of molecular markers for plant breeding selection has resulted in major advancements in efficiency and the successive announcement of many new varieties in the last 30 years [265,266]. MAS is often more effectual than traditional methods, resulting in improved accuracy, charge, or time savings, or the ability to screen for diseases that would otherwise be impossible to detect using standard phenotyping methods [267]. One of the major benefits of markers is that homozygosity can quickly be detected. In the private sector, molecular breeding programs have been widely adopted, with reports indicating increasing rates of genetic gain [268,269]. Numerous reports of marker-assisted variety improvement have emerged as a result of the widespread use of MAS in major crop breeding programs [270]. Backcrossing with DNA markers significantly improves selection efficiency. In essence, marker-assisted backcrossing (MABC) allows for very effective detection of the most important genes of interest or QTLs while preserving the recipient variety's original important features, allowing the original variety to be "upgraded" [265]. Whole-genome-based molecular breeding methodologies have been developed as a result of advances in rice genomics. Genomic selection (GS) is a new approach that has recently evolved [271]. Genomic selection, like MAS, is based on making predictions on a genomic scale from many DNA markers rather than focusing on individual genes or QTL [272,273]. Over the last ten years, pilot research in rice, wheat, and maize has yielded promising results in reducing breeding cycles and speeding up variety creation. Genomic selection provides a lot of potential for accurate selection of complicated variables such as yield and shortening the breeding cycle to improve genetic gain [274]. Recently, it has been reported that the GW6 gene is better for increasing grain length and width than Baodali, which was transferred into the recurrent parent 9311 *indica* cultivar and Zhonghua 11 cultivar *japonica* via MAS. Near isogenic lines (NILs) of these two cultivars displayed improved grain weight and grain production [275]. In addition, most studies have suggested that the pyramiding of grain size is a convenient method for QTLs to boost grain production in different plant species. For example, a study found that the pyramid of grain size and weight was better through QTLs and the involvement of the NILs NIL-qGL3, NIL-qGW2a, and NIL-qGS5. Similarly, the implementation of NILs in the genetic study of Zhenshan 97 using MAS and conventional back-crossing has been reported. The combination of three major QTLs for grain size showed improved effects on grain weight without any type of interaction [276]. A similar study reported that NILGS3/qgl3 has been established via crossing NIL-GS3 with NILggl3 by the MAS approach. The GS3 and qGL3 combined effects on grain length were compared among NILs. Furthermore, primary panicle transcription analysis in NILs showed that the gene expression regulated by GS3 and qGL3 did not overlap [277].

Fine-mapping and cloning for crop production, particularly grain weight, have made significant improvements in the last two decades. To date, 20 QTLs have been cloned for grain size and weight, including GL7/GW7 and GS9, which have opposite allelic directions of additive effects on grain length and weight, controlling grain size but having little effect on grain weight [192,278,279]. GSA1 and GW6a have similar results on grain length and weight with similar directions, implying that they significantly impact grain weight [13,280]. The remaining 16 QTLs have an impact on both grain size and grain weight. GW2 [281], TGW2 [282], GS5 [283], qSW5/GW5 [120,121], GW6 [284], and GW8 [191], are six that largely control grain width and weight, while the other ten QTLs containing GS2/GL2 [177,178], OsLG3 [285], qLGY3/OsLG3b [158,286], GS3 [153], GL3.1/qGL322 [287], TGW3/GL3.3 [288,289], GL4 [290], TGW6 [291], GL6 [292] and

GLW7 [49] largely control grain length and weight. All these QTLs characterizations have greatly aided our understanding of the genetic regulation of rice grain size and weight, but further research is required to further improve our understanding of the regulatory mechanism for these important agronomic traits [111].

Numerous studies have reported that alterations in the regulatory genes of quantitative traits, such as oil accumulation in the seed, showed a significantly increasing trend in intrinsic yields [293–295]. Selective genes that have played a vital role in seed development at different developmental phases require expression data of different developmental stage-specific tissues [296]. Furthermore, in polyploid species, the selection of target genes involved in seed size engineering may require specific tissue- and different stage-specific expression outlines of separate GFMs [297]. New technologies in functional genomics have been introduced for the development of seeds, and genetic engineering approaches provide greater scope for engineering to control seed size. Furthermore, increasing evidence shows that a thorough molecular knowledge of seed formation could yield opportunities for controlling seed size in plants [14,234,298]. For example, in a field study, the downregulation of *BnDA1* in *B. napus*, the *AtDA1* ortholog (a harmful regulator of seed formation), improved seed weight and seed yield by 21% and 13%, respectively [14]. In Arabidopsis, the two genes, AtSHB1 and AtKLUH, involved in the seed development were introduced in Brassica juncea and improved seed weight by 40% [299]. AtSHB1 overexpression in B. napus increased seed size 1.6-fold [300]. The loss of ARF18 function mutant in B. napus showed an approximately 8% increase in seed size compared to WT plants [301].

Grain size and yield increase in rice crops through the response to auxins by *big grain* 1 (BG1), encoding a membrane-connected protein (increased by about 15.2% in length and 17.0% in width, respectively). Furthermore, RNAi suppression of BG1 led to a decrease in rice grain size and yield relative to WT plants [301]. Downregulation of big seed 1 (BS1) and BS2 genes in soybean and Medicago resulted in significant enhancement of seed size [302]. Engineering techniques that are better for achieving larger seed size involve the overexpression of important regulators for the development of seeds in Arabidopsis [14,303] and other respective species [298] or the silencing of orthologs that are harmful regulators of seed growth [302]. Genetic material regulating biosynthesis, metabolism, and signaling is useful for gene manipulation, altering hormonal regulation, and producing larger seeds. For example, *BZR1*, a BR-responsive TF expressing most of the positive regulators, *SHB1* and IKU2, while AP2 and ARF2 are harmful regulators [238], may be modified in a spatiotemporal manner. Epigenomes are used as a genome editing technique in locus-specific ways and via methylation or de-methylation. Furthermore, several target genes that control the organized growth of maternal and zygotic tissues could be manipulated to increase seed size and crop yield. For example, KLUH and SHB1 regulate the development of seed coat and endosperm, and embryo formation, in a specific spatiotemporal manner, respectively. Moreover, control of cell cycle core gene expression is an alternative approach to increase seed size and yield. For example, target-specific up-regulation of the core cell cycle component CYCD7;1 in the endosperm helps overcome incomplete seed termination interlinked with ectopic expression [104]. Furthermore, genetic manipulation of genes that govern seed traits may reduce the negative relationship between the number and seed size. In Arabidopsis, FATA2, BZR1, LecRK-VIII.2, and CKX2 expression controls seed number, and expression of many other genes increases seed size and yield.

Genome editing technology, especially the CRISPR/Cas9 editing system, has widely been used in many plant species to knockout an individual gene, the key objective of its editing and repair mechanisms (Figure 3). The widely employed CRISPR/Cas9 and various CRISPR/Cas systems often generate a double-strand break (DSB) by cutting the double-stranded DNA. The double-strand break will often be repaired using non-homologous end joining (NHEJ). Because one or more nucleotide deletions or insertions may inhibit targeted gene expression due to frame shifting in the coding region, genome editing typically results in either gene knockout or silencing. CRISPR/Cas-induced gene silencing has many advantages over T-DNA mutagenesis. The main advantage of CRISPR/Cas-based

editing is its specificity; it may be used to target a particular gene without having any negative effects. Agrobacterium-mediated gene transformation has largely been used to overexpress a specific gene in a plant cell via T-DNA implantation, which includes the foreign DNA sequences. Even so, the haphazard insertion of a specified gene into a genome always has numerous adverse effects, such as the silencing of other useful genes. The CRISPR/Cas system may cleave the DNA sequence at a specified location, and homolog-directed repair (HDR) will then insert a targeted gene sequence into such a cleavage position. As a result, CRISPR/Cas may be used to accurately insert a gene of interest into a particular site inside a genome, avoiding interfering with neighboring genes or preventing position effects. Base editing (BE) is also a new technology that can specifically and proficiently implant point mutations at target positions without the use of DSB formation and donor DNA templates [304]. The cytosine base editor (CBE) and adenine base editor (ABE) are common examples of base editing technology, which has been applied in various plant species to interrogate gene function for crop trait improvement [305,306].



Figure 3. Application diversity of the CRISPR/Cas system for functional genomic research and crop improvement. CRISPR may directly generate gene knockout (silencing) via deletion or addition with a number of bases and repair through NHEJ depending on the DNA double-strand break mechanism. Alternatively, genome editing may replace an undesirable gene and/or overexpress (knock-in) a single gene when homolog-directed repair occurs with a DNA donor. In addition, CRISPR/Cas may also be employed for base editing, and epigenome editing by deactivating the Cas9 enzyme while using transcription effectors or other enzymes coupled with the dCas9.

In gene replacement for gene mutations, as in the assembly of HDR, any sequence may be inserted and replace the existing genomic sequence surrounding the DSB locations. Thus, if the nucleotide sequences have mutations, including point mutations or other undesired sequences, we may replace them with suitable nucleotide sequences that have homologous arms with the existing sequences. It was recently reported that transcript-template HDR was employed to completely replace the rice aceto-lactate synthase gene (ALS) by transport of such a DNA-free ribonucleoprotein complex [307]. Prime editing (PE) is a new DSB-independent precise genome editing technology that can introduce any base conversion, small indels, or a combination thereof, at target positions [304]. The prime editing approach was initially established in the animal system but rapidly utilized in plants to edit a precise gene. Prime editing is now being used to effectively replace genes in numerous key plant species, such as Arabidopsis, tobacco [308], rice [309–312], wheat [313], maize [314], tomato [315], and potato [316].

A CRISPR/dCas9 interference (CRISPRi) system was developed based on the application of the CRISPR/dCas9 system to disrupt transcriptional activities and to be used as a gene knockdown strategy for controlling gene expression [317,318]. CRISPRi could be used to inhibit the expression of genes by restricting transcription start and/or elongation depending on where the CRISPR/dCas system interacts. When the Cas system connects to the upstream area, it will impede TF and RNA polymerase (RNAP) binding and restrict transcription initiation. However, when it attaches to the coding region, it will prevent RNAP interaction and restrict transcription elongation [318]. Regulating a gene's methylation and demethylation can be an effective strategy to influence the expression and, ultimately, trait regulation. As a result, since the CRISPR/Cas system was successfully permitted as an effective genome editing system, research efforts have focused on tweaking it to alter DNA methylation.

CRISPR/Cas9 has great potential for specifically targeting desired genes [319]; thus, it could be used to develop crops with the desired seed size by knocking out specific genes without altering other traits [320]. However, in some cases, seed size may be linked with other important traits. For example, CRISPR-Cas9 genome editing of the TaIPK1 gene in wheat enhanced nutritional value and increased seed size simultaneously [321]. Recently, using CRISPR/Cas9-based genome editing, it was found that the MADS-box TF genes MADS78 and MADS79 are very important for regulating endosperm cellularization and early seed formation in rice [76]. A single MADS78 or MADS79 knockout mutant displayed early endosperm cellularization, while double mutants hindered seed development and produced no viable seeds [76]. Most recently, genome editing of cis-regulatory elements (CREs) has been found to result in both gain- and loss-of-function alleles, which is proving very useful for broadening the phenotypic range of traits related to yield and architecture [322,323].

6. Conclusions and Future Recommendations

Seed size and seed weight regulated by seed development are critical determinants of crop yields. Seed size can be determined through the development of zygotic embryos and endosperm as well as maternal tissues. Many signaling pathways that may determine seed size by the development of endosperm and maternal tissues, such as the IKU pathway, MAPK signaling, G-protein signaling, ubiquitin–protease signaling, etc., induce significant effects on entire features of the growth and development of plants, as well as regulating seed size. Transcription factors are responsible for enhancing cell growth in the maternal ovule and affect seed size. Therefore, there is a need to increase seed size with no effect on seed number through convention breeding programs to improve crop yield.

Many researchers have indicated that modern types of molecular approaches i.e., MAS, GS, and CRISPR/Cas system-based editing and transgenic technology have limitations to their disruption of seed yield hurdles in different plant species. In this review, we summarized many factors regulating seed sizes, such as genetic factors, signaling pathways, and transcriptional factor regulators in Arabidopsis and other crops, followed by engineering seed size by using recently evolved novel transgenic and breeding techniques, and ending with a brief discussion on recent studies, and those conducted over the last decade, which aimed to comprehend the genetic and molecular aspects controlling seed size. However, more research studies are required to understand the seed development pathways via molecular and genetic factors. Moreover, there is a need to introduce more strategies for genetic modifications to improve grain size and yield.

Author Contributions: Writing—original draft preparation, I.A.; writing—review and editing, I.A., K.B., Y.H., J.L. and L.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key-Areas Research and Development Program of Guangdong Province (2020B020220008), and the start-up fund from South China Agricultural University (to L.G.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Purugganan, M.D. Evolutionary insights into the nature of plant domestication. *Curr. Biol.* **2019**, *29*, R705–R714. [CrossRef] [PubMed]
- Arun, N.; Dalai, A.K. Environmental and socioeconomic impact assessment of biofuels from lignocellulosic biomass. In Lignocellulosic Biomass to Liquid Biofuels; Elsevier: Amsterdam, The Netherlands, 2020; pp. 283–299.
- 3. Pachapur, P.K.; Pachapur, V.L.; Brar, S.K.; Galvez, R.; Le Bihan, Y.; Surampalli, R.Y. Food security and sustainability. *Sustain. Fundam. Appl.* **2020**, 357–374.
- 4. Ding, G.; Zhao, Z.; Liao, Y.; Hu, Y.; Shi, L.; Long, Y.; Xu, F. Quantitative trait loci for seed yield and yield-related traits, and their responses to reduced phosphorus supply in *Brassica napus*. *Ann. Bot.* **2012**, *109*, 747–759. [CrossRef] [PubMed]
- 5. Yadava, S.K.; Ramchiary, N. Advanced Molecular Breeding for Yield. In *The Brassica juncea Genome*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 399–412.
- 6. Chaudhury, A.M.; Koltunow, A.; Payne, T.; Luo, M.; Tucker, M.R.; Dennis, E.; Peacock, W. Control of early seed development. *Annu. Rev. Cell Dev. Biol.* 2001, 17, 677–699. [CrossRef] [PubMed]
- Haughn, G.; Chaudhury, A. Genetic analysis of seed coat development in Arabidopsis. *Trends Plant Sci.* 2005, 10, 472–477. [CrossRef] [PubMed]
- 8. Armenta-Medina, A.; Gillmor, C.S.; Gao, P.; Mora-Macias, J.; Kochian, L.V.; Xiang, D.; Datla, R. Developmental and genomic architecture of plant embryogenesis: From model plant to crops. *Plant Commun.* **2021**, *2*, 100136. [CrossRef]
- 9. Locascio, A.; Roig-Villanova, I.; Bernardi, J.; Varotto, S. Current perspectives on the hormonal control of seed development in Arabidopsis and maize: A focus on auxin. *Front. Plant Sci.* 2014, *5*, 412. [CrossRef]
- 10. Agrawal, G.K.; Pedreschi, R.; Barkla, B.J.; Bindschedler, L.V.; Cramer, R.; Sarkar, A.; Renaut, J.; Job, D.; Rakwal, R. Translational plant proteomics: A perspective. J. Proteom. 2012, 75, 4588–4601. [CrossRef]
- 11. Moles, A.T.; Ackerly, D.D.; Webb, C.O.; Tweddle, J.C.; Dickie, J.B.; Westoby, M. A brief history of seed size. *Science* 2005, 307, 576–580. [CrossRef]
- 12. Alonso-Blanco, C.; Blankestijn-de Vries, H.; Hanhart, C.J.; Koornneef, M. Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4710–4717. [CrossRef]
- 13. Song, X.-J.; Huang, W.; Shi, M.; Zhu, M.-Z.; Lin, H.-X. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **2007**, *39*, 623–630. [CrossRef] [PubMed]
- Wang, J.L.; Tang, M.Q.; Chen, S.; Zheng, X.F.; Mo, H.X.; Li, S.J.; Wang, Z.; Zhu, K.M.; Ding, L.N.; Liu, S.Y. Down-regulation of BnDA1, whose gene locus is associated with the seeds weight, improves the seeds weight and organ size in *Brassica napus*. *Plant Biotechnol. J.* 2017, 15, 1024–1033. [CrossRef] [PubMed]
- Wang, S.; Liu, S.; Wang, J.; Yokosho, K.; Zhou, B.; Yu, Y.-C.; Liu, Z.; Frommer, W.B.; Ma, J.F.; Chen, L.-Q. Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. *Natl. Sci. Rev.* 2020, 7, 1776–1786. [CrossRef]
- 16. Huang, X.; Han, B. Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* **2014**, 65, 531–551. [CrossRef]
- 17. Korte, A.; Farlow, A. The advantages and limitations of trait analysis with GWAS: A review. Plant Methods 2013, 9, 29. [CrossRef]
- 18. Wang, D.; Sun, W.; Yuan, Z.; Sun, Q.; Fan, K.; Zhang, C.; Yu, S. Identification of a novel QTL and candidate gene associated with grain size using chromosome segment substitution lines in rice. *Sci. Rep.* **2021**, *11*, 189. [CrossRef] [PubMed]
- Luo, M.; Dennis, E.S.; Berger, F.; Peacock, W.J.; Chaudhury, A. MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. *Proc. Natl. Acad. Sci. USA* 2005, 102, 17531–17536. [CrossRef]
- 20. Wang, A.; Garcia, D.; Zhang, H.; Feng, K.; Chaudhury, A.; Berger, F.; Peacock, W.J.; Dennis, E.S.; Luo, M. The VQ motif protein IKU1 regulates endosperm growth and seed size in Arabidopsis. *Plant J.* **2010**, *63*, 670–679. [CrossRef]
- Zhang, B.; Li, C.; Li, Y.; Yu, H. Mobile TERMINAL FLOWER1 determines seed size in Arabidopsis. Nat. Plants 2020, 6, 1146–1157. [CrossRef]
- 22. Zhou, Y.; Zhang, X.; Kang, X.; Zhao, X.; Zhang, X.; Ni, M. SHORT HYPOCOTYL UNDER BLUE1 associates with MINISEED3 and HAIKU2 promoters in vivo to regulate Arabidopsis seed development. *Plant Cell* **2009**, *21*, 106–117. [CrossRef]
- 23. Kang, X.; Li, W.; Zhou, Y.; Ni, M. A WRKY transcription factor recruits the SYG1-like protein SHB1 to activate gene expression and seed cavity enlargement. *PLoS Genet.* **2013**, *9*, e1003347. [CrossRef] [PubMed]
- 24. Garcia, D.; Gerald, J.N.F.; Berger, F. Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *Plant Cell* **2005**, *17*, 52–60. [CrossRef] [PubMed]
- 25. Li, C.; Gong, X.; Zhang, B.; Liang, Z.; Wong, C.E.; See, B.Y.H.; Yu, H. TOP1α, UPF1, and TTG2 regulate seed size in a parental dosage–dependent manner. *PLoS Biol.* **2020**, *18*, e3000930. [CrossRef]
- Li, Z.; Chi, H.; Liu, C.; Zhang, T.; Han, L.; Li, L.; Pei, X.; Long, Y. Genome-wide identification and functional characterization of LEA genes during seed development process in linseed flax (*Linum usitatissimum* L.). *BMC Plant Biol.* 2021, 21, 193. [CrossRef] [PubMed]

- 27. Meng, L.-S.; Xu, M.-K.; Li, D.; Zhou, M.-M.; Jiang, J.-H. Soluble sugar accumulation can influence seed size via AN3–YDA gene cascade. *J. Agric. Food Chem.* 2017, 65, 4121–4132. [CrossRef] [PubMed]
- 28. Meng, L.-S.; Xu, M.-K.; Wan, W.; Wang, J.-Y. Integration of environmental and developmental (or metabolic) control of seed mass by sugar and ethylene metabolisms in Arabidopsis. *J. Agric. Food Chem.* **2018**, *66*, 3477–3488. [CrossRef]
- 29. Bauer, M.J.; Fischer, R.L. Genome demethylation and imprinting in the endosperm. *Curr. Opin. Plant Biol.* **2011**, *14*, 162–167. [CrossRef]
- 30. Ren, D.; Wang, X.; Yang, M.; Yang, L.; He, G.; Deng, X.W. A new regulator of seed size control in Arabidopsis identified by a genome-wide association study. *New Phytol.* **2019**, *222*, 895–906. [CrossRef]
- 31. Pien, S.; Grossniklaus, U. Polycomb group and trithorax group proteins in Arabidopsis. *Biochim. Et Biophys. Acta (BBA)-Gene Struct. Expr.* **2007**, 1769, 375–382. [CrossRef]
- 32. Kang, I.-H.; Steffen, J.G.; Portereiko, M.F.; Lloyd, A.; Drews, G.N. The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *Plant Cell* **2008**, *20*, 635–647. [CrossRef]
- Köhler, C.; Hennig, L.; Spillane, C.; Pien, S.; Gruissem, W.; Grossniklaus, U. The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev.* 2003, 17, 1540–1553. [CrossRef] [PubMed]
- Kono, A.; Umeda-Hara, C.; Adachi, S.; Nagata, N.; Konomi, M.; Nakagawa, T.; Uchimiya, H.; Umeda, M. The Arabidopsis D-type cyclin CYCD4 controls cell division in the stomatal lineage of the hypocotyl epidermis. *Plant Cell* 2007, 19, 1265–1277. [CrossRef] [PubMed]
- Johnston, A.J.; Matveeva, E.; Kirioukhova, O.; Grossniklaus, U.; Gruissem, W. A dynamic reciprocal RBR-PRC2 regulatory circuit controls Arabidopsis gametophyte development. *Curr. Biol.* 2008, 18, 1680–1686. [CrossRef] [PubMed]
- Liu, Z.; Zheng, L.; Pu, L.; Ma, X.; Wang, X.; Wu, Y.; Ming, H.; Wang, Q.; Zhang, G. ENO2 affects the seed size and weight by adjusting cytokinin content and forming ENO2-bZIP75 complex in Arabidopsis thaliana. *Front. Plant Sci.* 2020, *11*, 1331. [CrossRef]
- Xiao, W.; Brown, R.C.; Lemmon, B.E.; Harada, J.J.; Goldberg, R.B.; Fischer, R.L. Regulation of seed size by hypomethylation of maternal and paternal genomes. *Plant Physiol.* 2006, 142, 1160–1168. [CrossRef]
- Jullien, P.E.; Katz, A.; Oliva, M.; Ohad, N.; Berger, F. Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in Arabidopsis. *Curr. Biol.* 2006, 16, 486–492. [CrossRef]
- 39. Kinoshita, T.; Miura, A.; Choi, Y.; Kinoshita, Y.; Cao, X.; Jacobsen, S.E.; Fischer, R.L.; Kakutani, T. One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science* **2004**, *303*, 521–523. [CrossRef]
- Tiwari, S.; Schulz, R.; Ikeda, Y.; Dytham, L.; Bravo, J.; Mathers, L.; Spielman, M.; Guzmán, P.; Oakey, R.J.; Kinoshita, T. MATERNALLY EXPRESSED PAB C-TERMINAL, a novel imprinted gene in Arabidopsis, encodes the conserved C-terminal domain of polyadenylate binding proteins. *Plant Cell* 2008, 20, 2387–2398. [CrossRef]
- 41. Pires, N.D.; Bemer, M.; Müller, L.M.; Baroux, C.; Spillane, C.; Grossniklaus, U. Quantitative genetics identifies cryptic genetic variation involved in the paternal regulation of seed development. *PLoS Genet.* **2016**, *12*, e1005806. [CrossRef]
- 42. Beauzamy, L.; Fourquin, C.; Dubrulle, N.; Boursiac, Y.; Boudaoud, A.; Ingram, G. Endosperm turgor pressure decreases during early Arabidopsis seed development. *Development* **2016**, *143*, 3295–3299. [CrossRef]
- 43. Neff, M. Improving seed size, seed weight and seedling emergence in Camelina sativa by overexpressing the Atsob3-6 gene variant. *Transgenic Res.* **2020**, *29*, 409–418.
- 44. Mosher, R.A.; Melnyk, C.W.; Kelly, K.A.; Dunn, R.M.; Studholme, D.J.; Baulcombe, D.C. Uniparental expression of PolIVdependent siRNAs in developing endosperm of Arabidopsis. *Nature* 2009, 460, 283–286. [CrossRef] [PubMed]
- 45. Taleski, M.; Chapman, K.; Imin, N.; Djordjevic, M.A.; Groszmann, M. The peptide hormone receptor CEPR1 functions in the reproductive tissue to control seed size and yield. *Plant Physiol.* **2020**, *183*, 620–636. [CrossRef]
- 46. Sornay, E.; Forzani, C.; Forero-Vargas, M.; Dewitte, W.; Murray, J.A. Activation of CYCD 7; 1 in the central cell and early endosperm overcomes cell-cycle arrest in the Arabidopsis female gametophyte, and promotes early endosperm and embryo development. *Plant J.* **2015**, *84*, 41–55. [CrossRef] [PubMed]
- 47. Ingouff, M.; Jullien, P.E.; Berger, F. The female gametophyte and the endosperm control cell proliferation and differentiation of the seed coat in Arabidopsis. *Plant Cell* **2006**, *18*, 3491–3501. [CrossRef]
- Tonosaki, K.; Ono, A.; Kunisada, M.; Nishino, M.; Nagata, H.; Sakamoto, S.; Kijima, S.T.; Furuumi, H.; Nonomura, K.-I.; Sato, Y.; et al. Mutation of the imprinted gene OsEMF2a induces autonomous endosperm development and delayed cellularization in rice. *Plant Cell* 2021, 33, 85–103. [CrossRef] [PubMed]
- 49. Si, L.; Chen, J.; Huang, X.; Gong, H.; Luo, J.; Hou, Q.; Zhou, T.; Lu, T.; Zhu, J.; Shangguan, Y. OsSPL13 controls grain size in cultivated rice. *Nat. Genet.* **2016**, *48*, 447–456. [CrossRef]
- 50. Verma, A.; Prakash, G.; Ranjan, R.; Tyagi, A.K.; Agarwal, P. Silencing of an ubiquitin ligase increases grain width and weight in Indica rice. *Front. Genet.* **2021**, *11*, 1650. [CrossRef]
- 51. Takano-Kai, N.; Jiang, H.; Kubo, T.; Sweeney, M.; Matsumoto, T.; Kanamori, H.; Padhukasahasram, B.; Bustamante, C.; Yoshimura, A.; Doi, K. Evolutionary history of GS3, a gene conferring grain length in rice. *Genetics* **2009**, *182*, 1323–1334. [CrossRef]
- 52. Lyu, J.; Wang, D.; Duan, P.; Liu, Y.; Huang, K.; Zeng, D.; Zhang, L.; Dong, G.; Li, Y.; Xu, R. Control of grain size and weight by the GSK2-LARGE1/OML4 pathway in rice. *Plant Cell* **2020**, *32*, 1905–1918. [CrossRef]

- 53. Bai, X.; Luo, L.; Yan, W.; Kovi, M.R.; Zhan, W.; Xing, Y. Genetic dissection of rice grain shape using a recombinant inbred line population derived from two contrasting parents and fine mapping a pleiotropic quantitative trait locus qGL7. *BMC Genet.* **2010**, *11*, 16. [CrossRef] [PubMed]
- 54. Chun, Y.; Fang, J.; Zafar, S.A.; Shang, J.; Zhao, J.; Yuan, S.; Li, X. MINI SEED 2 (MIS2) Encodes a Receptor-like Kinase that Controls Grain Size and Shape in Rice. *Rice* 2020, *13*, 7. [CrossRef] [PubMed]
- 55. Paul, P.; Dhatt, B.K.; Miller, M.; Folsom, J.J.; Wang, Z.; Krassovskaya, I.; Liu, K.; Sandhu, J.; Yu, H.; Zhang, C. MADS78 and MADS79 are essential regulators of early seed development in rice. *Plant Physiol.* **2020**, *182*, 933–948. [CrossRef] [PubMed]
- 56. Li, D.; Liu, Z.; Gao, L.; Wang, L.; Gao, M.; Jiao, Z.; Qiao, H.; Yang, J.; Chen, M.; Yao, L. Genome-wide identification and characterization of microRNAs in developing grains of *Zea mays* L. *PLoS ONE* **2016**, *11*, e0153168. [CrossRef] [PubMed]
- 57. Martin, A.; Lee, J.; Kichey, T.; Gerentes, D.; Zivy, M.; Tatout, C.; Dubois, F.; Balliau, T.; Valot, B.; Davanture, M. Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production. *Plant Cell* **2006**, *18*, 3252–3274. [CrossRef]
- 58. Li, Q.; Yang, X.; Bai, G.; Warburton, M.L.; Mahuku, G.; Gore, M.; Dai, J.; Li, J.; Yan, J. Cloning and characterization of a putative GS3 ortholog involved in maize kernel development. *Theor. Appl. Genet.* **2010**, *120*, 753–763. [CrossRef]
- 59. Li, Q.; Li, L.; Yang, X.; Warburton, M.L.; Bai, G.; Dai, J.; Li, J.; Yan, J. Relationship, evolutionary fate and function of two maize co-orthologs of rice GW2 associated with kernel size and weight. *BMC Plant Biol.* **2010**, *10*, 143. [CrossRef]
- 60. Zhang, X.; Guo, W.; Du, D.; Pu, L.; Zhang, C. Overexpression of a maize BR transcription factor ZmBZR1 in Arabidopsis enlarges organ and seed size of the transgenic plants. *Plant Sci.* **2020**, *292*, 110378. [CrossRef]
- 61. Su, Z.; Hao, C.; Wang, L.; Dong, Y.; Zhang, X. Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **2011**, *122*, 211–223. [CrossRef]
- 62. Liu, N.; Wu, S.; Li, Z.; Khan, A.Q.; Hu, H.; Zhang, X.; Tu, L. Repression of microRNA 160 results in retarded seed integument growth and smaller final seed size in cotton. *Crop J.* **2020**, *8*, 602–612. [CrossRef]
- 63. Geravandi, M.; Cheghamirza, K.; Farshadfar, E.; Gepts, P. QTL analysis of seed size and yield-related traits in an inter-genepool population of common bean (*Phaseolus vulgaris* L.). *Sci. Hortic.* **2020**, *274*, 109678. [CrossRef]
- 64. Khan, M.H.; Hu, L.; Zhu, M.; Zhai, Y.; Khan, S.U.; Ahmar, S.; Amoo, O.; Zhang, K.; Fan, C.; Zhou, Y. Targeted mutagenesis of EOD3 gene in *Brassica napus* L. regulates seed production. *J. Cell. Physiol.* **2021**, 236, 1996–2007. [CrossRef] [PubMed]
- 65. Hina, A.; Cao, Y.; Song, S.; Li, S.; Sharmin, R.A.; Elattar, M.A.; Bhat, J.A.; Zhao, T. High-resolution mapping in two RIL populations refines major "QTL Hotspot" regions for seed size and shape in soybean (*Glycine max* L.). *Int. J. Mol. Sci.* 2020, 21, 1040. [CrossRef] [PubMed]
- 66. Nguyen, C.X.; Paddock, K.J.; Zhang, Z.; Stacey, M.G. GmKIX8-1 regulates organ size in soybean and is the causative gene for the major seed weight QTL qSw17-1. *New Phytol.* **2021**, *229*, 920–934. [CrossRef] [PubMed]
- 67. Alyr, M.H.; Pallu, J.; Sambou, A.; Nguepjop, J.R.; Seye, M.; Tossim, H.-A.; Djiboune, Y.R.; Sane, D.; Rami, J.-F.; Fonceka, D. Fine-Mapping of a Wild Genomic Region Involved in Pod and Seed Size Reduction on Chromosome A07 in Peanut (*Arachis hypogaea* L.). *Genes* **2020**, *11*, 1402. [CrossRef]
- Scott, R.J.; Spielman, M.; Bailey, J.; Dickinson, H.G. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* 1998, 125, 3329–3341. [CrossRef]
- 69. Adams, S.; Vinkenoog, R.; Spielman, M.; Dickinson, H.G.; Scott, R.J. Parent-of-origin effects on seed development in *Arabidopsis thaliana* require DNA methylation. *Development* **2000**, *127*, 2493–2502. [CrossRef]
- 70. Garcia, D.; Saingery, V.; Chambrier, P.; Mayer, U.; Jurgens, G.; Berger, F. Arabidopsis haiku mutants reveal new controls of seed size by endosperm. *Plant Physiol.* **2003**, *131*, 1661–1670. [CrossRef]
- 71. Wang, D.; Zhang, C.; Hearn, D.J.; Kang, I.-H.; Punwani, J.A.; Skaggs, M.I.; Drews, G.N.; Schumaker, K.S.; Yadegari, R. Identification of transcription-factor genes expressed in the Arabidopsis female gametophyte. *BMC Plant Biol.* **2010**, *10*, 110. [CrossRef]
- 72. Zuo, J.; Li, J. Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Annu. Rev. Genet.* **2014**, *48*, 99–118. [CrossRef]
- 73. Gehring, M.; Satyaki, P. Endosperm and imprinting, inextricably linked. Plant Physiol. 2017, 173, 143–154. [CrossRef] [PubMed]
- 74. Jahnke, S.; Scholten, S. Epigenetic resetting of a gene imprinted in plant embryos. *Curr. Biol.* **2009**, *19*, 1677–1681. [CrossRef] [PubMed]
- 75. Nodine, M.D.; Bartel, D.P. Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **2012**, 482, 94–97. [CrossRef] [PubMed]
- 76. Luo, M.; Taylor, J.M.; Spriggs, A.; Zhang, H.; Wu, X.; Russell, S.; Singh, M.; Koltunow, A. A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet.* **2011**, *7*, e1002125. [CrossRef]
- 77. Waters, A.J.; Bilinski, P.; Eichten, S.R.; Vaughn, M.W.; Ross-Ibarra, J.; Gehring, M.; Springer, N.M. Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc. Natl. Acad. Sci.* USA 2013, 110, 19639–19644. [CrossRef]
- 78. Jullien, P.E.; Kinoshita, T.; Ohad, N.; Berger, F. Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. *Plant Cell* **2006**, *18*, 1360–1372. [CrossRef] [PubMed]
- 79. Jiang, H.; Köhler, C. Evolution, function, and regulation of genomic imprinting in plant seed development. *J. Exp. Bot.* **2012**, 63, 4713–4722. [CrossRef] [PubMed]
- 80. Liu, W.; Stewart Jr, C.N. Plant synthetic promoters and transcription factors. Curr. Opin. Biotechnol. 2016, 37, 36–44. [CrossRef]

- 81. Grossniklaus, U.; Vielle-Calzada, J.-P.; Hoeppner, M.A.; Gagliano, W.B. Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. *Science* **1998**, *280*, 446–450. [CrossRef]
- 82. Rajkumar, M.S.; Gupta, K.; Khemka, N.K.; Garg, R.; Jain, M. DNA methylation reprogramming during seed development and its functional relevance in seed size/weight determination in chickpea. *Commun. Biol.* **2020**, *3*, 340. [CrossRef]
- Kumar, S.; Mohapatra, T. Dynamics of DNA Methylation and Its Functions in Plant Growth and Development. *Front. Plant Sci.* 2021, 12, 596236. [CrossRef] [PubMed]
- 84. Lu, C. Breeding and utilization of two-line intersubspecific hybrid rice Liangyou Peijiu. Hybrid Rice 2000, 15, 4–5.
- 85. Zhou, S.; Xing, M.; Zhao, Z.; Gu, Y.; Xiao, Y.; Liu, Q.; Xue, H. DNA methylation modification in heterosis initiation through analyzing rice hybrid contemporary seeds. *Crop J.* **2021**, *9*, 1179–1190. [CrossRef]
- Hsieh, T.-F.; Shin, J.; Uzawa, R.; Silva, P.; Cohen, S.; Bauer, M.J.; Hashimoto, M.; Kirkbride, R.C.; Harada, J.J.; Zilberman, D. Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc. Natl. Acad. Sci. USA* 2011, 108, 1755–1762. [CrossRef] [PubMed]
- Xing, M.-Q.; Zhang, Y.-J.; Zhou, S.-R.; Hu, W.-Y.; Wu, X.-T.; Ye, Y.-J.; Wu, X.-X.; Xiao, Y.-P.; Li, X.; Xue, H.-W. Global analysis reveals the crucial roles of DNA methylation during rice seed development. *Plant Physiol.* 2015, 168, 1417–1432. [CrossRef] [PubMed]
- 88. Creff, A.; Ali, O.; Bayle, V.; Ingram, G.; Landrein, B. Endosperm turgor pressure both promotes and restricts seed growth and size. *bioRxiv* **2021**. [CrossRef]
- Dhaka, N.; Sharma, R. MicroRNA-mediated regulation of agronomically important seed traits: A treasure trove with shades of grey! Crit. Rev. Biotechnol. 2021, 41, 594–608. [CrossRef]
- 90. Zhang, T.; Li, Z.; Song, X.; Han, L.; Wang, L.; Zhang, J.; Long, Y.; Pei, X. Identification and Characterization of microRNAs in the Developing Seed of Linseed Flax (*Linum usitatissimum* L.). *Int. J. Mol. Sci.* **2020**, *21*, 2708. [CrossRef]
- 91. Parreira, J.R.; Cappuccio, M.; Balestrazzi, A.; Fevereiro, P.; Araújo, S.d.S. MicroRNAs expression dynamics reveal posttranscriptional mechanisms regulating seed development in *Phaseolus vulgaris* L. *Hortic. Res.* **2021**, *8*, 18. [CrossRef]
- 92. He, X.-J.; Hsu, Y.-F.; Zhu, S.; Liu, H.-L.; Pontes, O.; Zhu, J.; Cui, X.; Wang, C.-S.; Zhu, J.-K. A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development. *Genes Dev.* **2009**, *23*, 2717–2722. [CrossRef]
- 93. Mosher, R.A.; Melnyk, C.W. siRNAs and DNA methylation: Seedy epigenetics. *Trends Plant Sci.* 2010, 15, 204–210. [CrossRef] [PubMed]
- 94. Cai, D.; Xiao, Y.; Yang, W.; Ye, W.; Wang, B.; Younas, M.; Wu, J.; Liu, K. Association mapping of six yield-related traits in rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* 2014, 127, 85–96. [CrossRef] [PubMed]
- 95. Chen, J.; Zhang, L.; Liu, S.; Li, Z.; Huang, R.; Li, Y.; Cheng, H.; Li, X.; Zhou, B.; Wu, S. The genetic basis of natural variation in kernel size and related traits using a four-way cross population in maize. *PLoS ONE* **2016**, *11*, e0153428. [CrossRef] [PubMed]
- 96. Liller, C.B.; Walla, A.; Boer, M.P.; Hedley, P.; Macaulay, M.; Effgen, S.; von Korff, M.; van Esse, G.W.; Koornneef, M. Fine mapping of a major QTL for awn length in barley using a multiparent mapping population. *Theor. Appl. Genet.* 2017, 130, 269–281. [CrossRef]
- Pu, C.X.; Ma, Y.; Wang, J.; Zhang, Y.C.; Jiao, X.W.; Hu, Y.H.; Wang, L.L.; Zhu, Z.G.; Sun, D.; Sun, Y. Crinkly4 receptor-like kinase is required to maintain the interlocking of the palea and lemma, and fertility in rice, by promoting epidermal cell differentiation. *Plant J.* 2012, *70*, 940–953. [CrossRef] [PubMed]
- 98. De Veylder, L.; Beeckman, T.; Inzé, D. The ins and outs of the plant cell cycle. Nat. Rev. Mol. Cell Biol. 2007, 8, 655–665. [CrossRef]
- Johnston, A.J.; Kirioukhova, O.; Barrell, P.J.; Rutten, T.; Moore, J.M.; Baskar, R.; Grossniklaus, U.; Gruissem, W. Dosage-sensitive function of retinoblastoma related and convergent epigenetic control are required during the Arabidopsis life cycle. *PLoS Genet.* 2010, *6*, e1000988. [CrossRef]
- Collins, E.S.; Balchand, S.K.; Faraci, J.L.; Wadsworth, P.; Lee, W.-L. Cell cycle–regulated cortical dynein/dynactin promotes symmetric cell division by differential pole motion in anaphase. *Mol. Biol. Cell* 2012, 23, 3380–3390. [CrossRef]
- 101. Liu, H.; Xu, H.; Qiu, W.; Li, X.; Zhang, H.; Zhu, Y.; Li, C.; Wang, L. Roles of bZIP transcription factors in plant growth and development and abiotic stress response. *Acta Agric. Zhejiangensis* **2019**, *31*, 1205–1214.
- Lee, H.; Guo, Y.; Ohta, M.; Xiong, L.; Stevenson, B.; Zhu, J.K. LOS2, a genetic locus required for cold-responsive gene transcription encodes a bi-functional enolase. *EMBO J.* 2002, 21, 2692–2702. [CrossRef]
- 103. Andriotis, V.M.; Kruger, N.J.; Pike, M.J.; Smith, A.M. Plastidial glycolysis in developing Arabidopsis embryos. *New Phytol.* **2010**, *185*, 649–662. [CrossRef] [PubMed]
- 104. Moon, J.; Parry, G.; Estelle, M. The ubiquitin-proteasome pathway and plant development. *Plant Cell* **2004**, *16*, 3181–3195. [CrossRef] [PubMed]
- 105. Schwechheimer, C.; Schwager, K. Regulated proteolysis and plant development. *Plant Cell Rep.* **2004**, *23*, 353–364. [CrossRef] [PubMed]
- Zhou, Y.; Park, S.-H.; Soh, M.Y.; Chua, N.-H. Ubiquitin-specific proteases UBP12 and UBP13 promote shade avoidance response by enhancing PIF7 stability. *Proc. Natl. Acad. Sci. USA* 2021, 118, e2103633118. [CrossRef] [PubMed]
- 107. Iglesias, M.J.; Casalongué, C.A.; Terrile, M.C. 28—Ubiquitin-proteasome system as part of nitric oxide sensing in plants. In *Nitric Oxide in Plant Biology*; Pratap Singh, V., Singh, S., Tripathi, D.K., Romero-Puertas, M.C., Sandalio, L.M., Eds.; Academic Press: Cambridge, MA, USA, 2022; pp. 653–687. [CrossRef]
- 108. Glickman, M.H.; Ciechanover, A. The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. *Physiol. Rev.* **2002**, *82*, 373–428. [CrossRef]

- 109. Morreale, F.E.; Walden, H. Types of ubiquitin ligases. Cell 2016, 165, 248-248.e1. [CrossRef]
- 110. Xia, T.; Li, N.; Dumenil, J.; Li, J.; Kamenski, A.; Bevan, M.W.; Gao, F.; Li, Y. The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in Arabidopsis. *Plant Cell* **2013**, *25*, 3347–3359. [CrossRef]
- 111. Li, N.; Li, Y. Signaling pathways of seed size control in plants. Curr. Opin. Plant Biol. 2016, 33, 23–32. [CrossRef]
- 112. Li, Y.; Zheng, L.; Corke, F.; Smith, C.; Bevan, M.W. Control of final seed and organ size by the DA1 gene family in *Arabidopsis thaliana*. *Genes Dev.* **2008**, *22*, 1331–1336. [CrossRef]
- 113. Disch, S.; Anastasiou, E.; Sharma, V.K.; Laux, T.; Fletcher, J.C.; Lenhard, M. The E3 ubiquitin ligase BIG BROTHER controls Arabidopsis organ size in a dosage-dependent manner. *Curr. Biol.* **2006**, *16*, 272–279. [CrossRef]
- 114. Liu, Y.; Wang, F.; Zhang, H.; He, H.; Ma, L.; Deng, X.W. Functional characterization of the Arabidopsis ubiquitin-specific protease gene family reveals specific role and redundancy of individual members in development. *Plant J.* 2008, 55, 844–856. [CrossRef] [PubMed]
- 115. Du, L.; Li, N.; Chen, L.; Xu, Y.; Li, Y.; Zhang, Y.; Li, C.; Li, Y. The ubiquitin receptor DA1 regulates seed and organ size by modulating the stability of the ubiquitin-specific protease UBP15/SOD2 in Arabidopsis. *Plant Cell* 2014, 26, 665–677. [CrossRef] [PubMed]
- 116. Kurepa, J.; Wang, S.; Li, Y.; Zaitlin, D.; Pierce, A.J.; Smalle, J.A. Loss of 26S proteasome function leads to increased cell size and decreased cell number in Arabidopsis shoot organs. *Plant Physiol.* **2009**, *150*, 178–189. [CrossRef] [PubMed]
- 117. Eloy, N.B.; Gonzalez, N.; Van Leene, J.; Maleux, K.; Vanhaeren, H.; De Milde, L.; Dhondt, S.; Vercruysse, L.; Witters, E.; Mercier, R. SAMBA, a plant-specific anaphase-promoting complex/cyclosome regulator is involved in early development and A-type cyclin stabilization. *Proc. Natl. Acad. Sci. USA* 2012, *109*, 13853–13858. [CrossRef] [PubMed]
- 118. Willems, A.; De Veylder, L. The Plant Anaphase-Promoting Complex/Cyclosome. *Annu. Rev. Cell Dev. Biol.* 2022, 38, 25–48. [CrossRef] [PubMed]
- 119. Vanhaeren, H.; Inzé, D.; Gonzalez, N. Plant growth beyond limits. Trends Plant Sci. 2016, 21, 102–109. [CrossRef]
- 120. Shomura, A.; Izawa, T.; Ebana, K.; Ebitani, T.; Kanegae, H.; Konishi, S.; Yano, M. Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* **2008**, *40*, 1023–1028. [CrossRef]
- 121. Weng, J.; Gu, S.; Wan, X.; Gao, H.; Guo, T.; Su, N.; Lei, C.; Zhang, X.; Cheng, Z.; Guo, X. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* **2008**, *18*, 1199–1209. [CrossRef]
- 122. Ying, J.-Z.; Gao, J.-P.; Shan, J.-X.; Zhu, M.-Z.; Shi, M.; Lin, H.-X. Dissecting the genetic basis of extremely large grain shape in rice cultivar 'JZ1560'. J. Genet. Genom. 2012, 39, 325–333. [CrossRef]
- 123. Keren, I.; Lacroix, B.; Kohrman, A.; Citovsky, V. Histone deubiquitinase OTU1 epigenetically regulates DA1 and DA2, which control Arabidopsis seed and organ size. *Iscience* 2020, 23, 100948. [CrossRef]
- 124. Yin, P.; Ma, Q.; Wang, H.; Feng, D.; Wang, X.; Pei, Y.; Wen, J.; Tadege, M.; Niu, L.; Lin, H. SMALL LEAF AND BUSHY1 controls organ size and lateral branching by modulating the stability of BIG SEEDS1 in Medicago truncatula. *New Phytol.* **2020**, 226, 1399–1412. [CrossRef] [PubMed]
- 125. He, X.; Yin, Q.; Zhou, L.; Meng, L.; Hu, W.; Li, F.; Li, Y.; Han, K.; Zhang, S.; Fu, S. Metagenomic sequencing reveals viral abundance and diversity in mosquitoes from the Shaanxi-Gansu-Ningxia region, China. *PLoS Negl. Trop. Dis.* 2021, 15, e0009381. [CrossRef] [PubMed]
- Bent, A.F. Plant mitogen-activated protein kinase cascades: Negative regulatory roles turn out positive. *Proc. Natl. Acad. Sci. USA* 2001, 98, 784–786. [CrossRef] [PubMed]
- 127. Tena, G.; Asai, T.; Chiu, W.-L.; Sheen, J. Plant mitogen-activated protein kinase signaling cascades. *Curr. Opin. Plant Biol.* 2001, 4, 392–400. [CrossRef]
- 128. Pedley, K.F.; Martin, G.B. Role of mitogen-activated protein kinases in plant immunity. *Curr. Opin. Plant Biol.* 2005, *8*, 541–547. [CrossRef]
- 129. Chen, J.; Wang, L.; Yuan, M. Update on the Roles of Rice MAPK Cascades. Int. J. Mol. Sci. 2021, 22, 1679. [CrossRef]
- 130. Jagodzik, P.; Tajdel-Zielinska, M.; Ciesla, A.; Marczak, M.; Ludwikow, A. Mitogen-Activated Protein Kinase Cascades in Plant Hormone Signaling. *Front. Plant Sci.* **2018**, *9*, 1387. [CrossRef]
- 131. Duan, P.; Rao, Y.; Zeng, D.; Yang, Y.; Xu, R.; Zhang, B.; Dong, G.; Qian, Q.; Li, Y. SMALL GRAIN 1, which encodes a mitogenactivated protein kinase kinase 4, influences grain size in rice. *Plant J.* 2014, *77*, 547–557. [CrossRef]
- 132. Liu, S.; Hua, L.; Dong, S.; Chen, H.; Zhu, X.; Jiang, J.e.; Zhang, F.; Li, Y.; Fang, X.; Chen, F. Os MAPK 6, a mitogen-activated protein kinase, influences rice grain size and biomass production. *Plant J.* **2015**, *84*, 672–681. [CrossRef]
- 133. Xu, R.; Yu, H.; Wang, J.; Duan, P.; Zhang, B.; Li, J.; Li, Y.; Xu, J.; Lyu, J.; Li, N.; et al. A mitogen-activated protein kinase phosphatase influences grain size and weight in rice. *Plant J. Cell Mol. Biol.* **2018**, *95*, 937–946. [CrossRef]
- 134. Guo, T.; Chen, K.; Dong, N.-Q.; Shi, C.-L.; Ye, W.-W.; Gao, J.-P.; Shan, J.-X.; Lin, H.-X. GRAIN SIZE AND NUMBER1 negatively regulates the OsMKKK10-OsMKK4-OsMPK6 cascade to coordinate the trade-off between grain number per panicle and grain size in rice. *Plant Cell* 2018, 30, 871–888. [CrossRef] [PubMed]
- 135. Kawamoto, N.; Del Carpio, D.P.; Hofmann, A.; Mizuta, Y.; Kurihara, D.; Higashiyama, T.; Uchida, N.; Torii, K.U.; Colombo, L.; Groth, G. A peptide pair coordinates regular ovule initiation patterns with seed number and fruit size. *Curr. Biol.* 2020, 30, 4352–4361.e4. [CrossRef] [PubMed]
- 136. Zhang, Y.; Yao, W.; Wang, F.; Su, Y.; Zhang, D.; Hu, S.; Zhang, X. AGC protein kinase AGC1-4 mediates seed size in Arabidopsis. *Plant Cell Rep.* **2020**, *39*, 825–837. [CrossRef] [PubMed]

- 137. Deng, X.; Han, X.; Yu, S.; Liu, Z.; Guo, D.; He, Y.; Li, W.; Tao, Y.; Sun, C.; Xu, P. OsINV3 and its homolog, OsINV2, control grain size in rice. *Int. J. Mol. Sci.* 2020, 21, 2199. [CrossRef]
- 138. He, C.; Wang, J.; Dong, R.; Guan, H.; Liu, T.; Liu, C.; Liu, Q.; Wang, L. Overexpression of an Antisense RNA of Maize Receptor-Like Kinase Gene ZmRLK7 Enlarges the Organ and Seed Size of Transgenic Arabidopsis Plants. *Front. Plant Sci.* 2020, 11, 579120. [CrossRef]
- 139. Xiao, W.; Hu, S.; Zou, X.; Cai, R.; Liao, R.; Lin, X.; Yao, R.; Guo, X. Lectin receptor-like kinase LecRK-VIII. 2 is a missing link in MAPK signaling-mediated yield control. *Plant Physiol.* **2021**, *187*, 303–320. [CrossRef]
- 140. Sun, X.; Shantharaj, D.; Kang, X.; Ni, M. Transcriptional and hormonal signaling control of Arabidopsis seed development. *Curr. Opin. Plant Biol.* **2010**, *13*, 611–620. [CrossRef]
- Cheng, Z.J.; Zhao, X.Y.; Shao, X.X.; Wang, F.; Zhou, C.; Liu, Y.G.; Zhang, Y.; Zhang, X.S. Abscisic acid regulates early seed development in Arabidopsis by ABI5-mediated transcription of SHORT HYPOCOTYL UNDER BLUE1. *Plant Cell* 2014, 26, 1053–1068. [CrossRef]
- 142. Xi, X.; Hu, Z.; Nie, X.; Meng, M.; Xu, H.; Li, J. Cross Inhibition of MPK10 and WRKY10 Participating in the Growth of Endosperm in *Arabidopsis thaliana*. *Front. Plant Sci.* **2021**, *12*, 521. [CrossRef]
- 143. Xu, R.; Li, N.; Li, Y. Control of grain size by G protein signaling in rice. J. Integr. Plant Biol. 2019, 61, 533–540. [CrossRef]
- 144. Ullah, H.; Chen, J.-G.; Young, J.C.; Im, K.-H.; Sussman, M.R.; Jones, A.M. Modulation of cell proliferation by heterotrimeric G protein in Arabidopsis. *Science* 2001, 292, 2066–2069. [CrossRef] [PubMed]
- 145. Lease, K.A.; Wen, J.; Li, J.; Doke, J.T.; Liscum, E.; Walker, J.C. A mutant Arabidopsis heterotrimeric G-protein β subunit affects leaf, flower, and fruit development. *Plant Cell* **2001**, *13*, 2631–2641. [PubMed]
- 146. Li, S.; Liu, Y.; Zheng, L.; Chen, L.; Li, N.; Corke, F.; Lu, Y.; Fu, X.; Zhu, Z.; Bevan, M.W. The plant-specific G protein γ subunit AGG3 influences organ size and shape in *Arabidopsis thaliana*. New Phytol. 2012, 194, 690–703. [CrossRef] [PubMed]
- 147. Chakravorty, D.; Trusov, Y.; Zhang, W.; Acharya, B.R.; Sheahan, M.B.; McCurdy, D.W.; Assmann, S.M.; Botella, J.R. An atypical heterotrimeric G-protein γ-subunit is involved in guard cell K+-channel regulation and morphological development in *Arabidopsis thaliana*. *Plant J.* **2011**, *67*, 840–851. [CrossRef] [PubMed]
- 148. Roy Choudhury, S.; Riesselman, A.J.; Pandey, S. Constitutive or seed-specific overexpression of A rabidopsis G-protein γ subunit 3 (AGG 3) results in increased seed and oil production and improved stress tolerance in C amelina sativa. *Plant Biotechnol. J.* 2014, 12, 49–59. [CrossRef] [PubMed]
- 149. Utsunomiya, Y.; Samejima, C.; Takayanagi, Y.; Izawa, Y.; Yoshida, T.; Sawada, Y.; Fujisawa, Y.; Kato, H.; Iwasaki, Y. Suppression of the rice heterotrimeric G protein β-subunit gene, RGB1, causes dwarfism and browning of internodes and lamina joint regions. *Plant J.* **2011**, *67*, 907–916. [CrossRef] [PubMed]
- 150. Fujisawa, Y.; Kato, T.; Ohki, S.; Ishikawa, A.; Kitano, H.; Sasaki, T.; Asahi, T.; Iwasaki, Y. Suppression of the heterotrimeric G protein causes abnormal morphology, including dwarfism, in rice. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 7575–7580. [CrossRef]
- 151. Ashikari, M.; Wu, J.; Yano, M.; Sasaki, T.; Yoshimura, A. Rice gibberellin-insensitive dwarf mutant gene Dwarf 1 encodes the α-subunit of GTP-binding protein. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10284–10289. [CrossRef]
- 152. Zhang, D.; Zhang, M.; Liang, J. RGB1 Regulates Grain Development and Starch Accumulation through Its Effect on OsYUC11-Mediated Auxin Biosynthesis in Rice Endosperm Cells. *Front. Plant Sci.* **2021**, *12*, 585174. [CrossRef]
- 153. Fan, C.; Xing, Y.; Mao, H.; Lu, T.; Han, B.; Xu, C.; Li, X.; Zhang, Q. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* 2006, 112, 1164–1171. [CrossRef]
- 154. Mao, H.; Sun, S.; Yao, J.; Wang, C.; Yu, S.; Xu, C.; Li, X.; Zhang, Q. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19579–19584. [CrossRef] [PubMed]
- 155. Huang, X.; Qian, Q.; Liu, Z.; Sun, H.; He, S.; Luo, D.; Xia, G.; Chu, C.; Li, J.; Fu, X. Natural variation at the DEP1 locus enhances grain yield in rice. *Nat. Genet.* 2009, *41*, 494–497. [CrossRef] [PubMed]
- 156. Li, Q.; Lu, L.; Liu, H.; Bai, X.; Zhou, X.; Wu, B.; Yuan, M.; Yang, L.; Xing, Y. A minor QTL, SG3, encoding an R2R3-MYB protein, negatively controls grain length in rice. *Theor. Appl. Genet.* **2020**, *133*, 2387–2399. [CrossRef] [PubMed]
- 157. Shen, L.; Wang, C.; Fu, Y.; Wang, J.; Liu, Q.; Zhang, X.; Yan, C.; Qian, Q.; Wang, K. QTL editing confers opposing yield performance in different rice varieties. *J. Integr. Plant Biol.* **2018**, *60*, 89–93. [CrossRef]
- 158. Liu, Q.; Han, R.; Wu, K.; Zhang, J.; Ye, Y.; Wang, S.; Chen, J.; Pan, Y.; Li, Q.; Xu, X.; et al. G-protein βγ subunits determine grain size through interaction with MADS-domain transcription factors in rice. *Nat. Commun.* **2018**, *9*, 852. [CrossRef]
- 159. Gao, Y.; Gu, H.; Leburu, M.; Li, X.; Wang, Y.; Sheng, J.; Fang, H.; Gu, M.; Liang, G. The heterotrimeric G protein β subunit RGB1 is required for seedling formation in rice. *Rice* **2019**, *12*, 53. [CrossRef]
- 160. Yang, W.; Wu, K.; Wang, B.; Liu, H.; Guo, S.; Guo, X.; Luo, W.; Sun, S.; Ouyang, Y.; Fu, X. The RING E3 ligase CLG1 targets GS3 for degradation via the endosome pathway to determine grain size in rice. *Mol. Plant* **2021**, *14*, 1699–1713. [CrossRef]
- 161. Ohto, M.-a.; Fischer, R.L.; Goldberg, R.B.; Nakamura, K.; Harada, J.J. Control of seed mass by APETALA2. *Proc. Natl. Acad. Sci.* USA 2005, 102, 3123–3128. [CrossRef]
- 162. Johnson, C.S.; Kolevski, B.; Smyth, D.R. TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. *Plant Cell* **2002**, *14*, 1359–1375. [CrossRef]
- 163. Jofuku, K.D.; Omidyar, P.K.; Gee, Z.; Okamuro, J.K. Control of seed mass and seed yield by the floral homeotic gene APETALA2. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3117–3122. [CrossRef]

- 164. Ohto, M.-a.; Floyd, S.K.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in Arabidopsis. *Sex. Plant Reprod.* **2009**, *22*, 277–289. [CrossRef] [PubMed]
- 165. Zhang, Y.; Du, L.; Xu, R.; Cui, R.; Hao, J.; Sun, C.; Li, Y. Transcription factors SOD7/NGAL2 and DPA4/NGAL3 act redundantly to regulate seed size by directly repressing KLU expression in *Arabidopsis thaliana*. *Plant Cell* 2015, 27, 620–632. [CrossRef] [PubMed]
- 166. Elliott, R.C.; Betzner, A.S.; Huttner, E.; Oakes, M.P.; Tucker, W.; Gerentes, D.; Perez, P.; Smyth, D.R. AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. *Plant Cell* 1996, *8*, 155–168.
- 167. Mizukami, Y.; Fischer, R.L. Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 942–947. [CrossRef] [PubMed]
- 168. Luo, J.; Liu, H.; Zhou, T.; Gu, B.; Huang, X.; Shangguan, Y.; Zhu, J.; Li, Y.; Zhao, Y.; Wang, Y. An-1 encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. *Plant Cell* **2013**, *25*, 3360–3376. [CrossRef] [PubMed]
- Jofuku, K.D.; Den Boer, B.; Van Montagu, M.; Okamuro, J.K. Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *Plant Cell* 1994, 6, 1211–1225. [PubMed]
- 170. Berger, F.; Chaudhury, A. Parental memories shape seeds. Trends Plant Sci. 2009, 14, 550–556. [CrossRef]
- 171. Aya, K.; Hobo, T.; Sato-Izawa, K.; Ueguchi-Tanaka, M.; Kitano, H.; Matsuoka, M. A novel AP2-type transcription factor, SMALL ORGAN SIZE1, controls organ size downstream of an auxin signaling pathway. *Plant Cell Physiol.* **2014**, *55*, 897–912. [CrossRef]
- 172. Hirano, K.; Yoshida, H.; Aya, K.; Kawamura, M.; Hayashi, M.; Hobo, T.; Sato-Izawa, K.; Kitano, H.; Ueguchi-Tanaka, M.; Matsuoka, M. SMALL ORGAN SIZE 1 and SMALL ORGAN SIZE 2/DWARF AND LOW-TILLERING form a complex to integrate auxin and brassinosteroid signaling in rice. *Mol. Plant* 2017, *10*, 590–604. [CrossRef]
- 173. Taketa, S.; Hattori, M.; Takami, T.; Himi, E.; Sakamoto, W. Barley albino lemma 1 resulted from mutations in a Golden2-like gene reduces seed weight. *Plant Cell Physiol.* **2021**, *62*, 447–457. [CrossRef]
- 174. Alam, I.; Wu, X.; Yu, Q.; Ge, L. Comprehensive Genomic Analysis of G2-like Transcription Factor Genes and Their Role in Development and Abiotic Stresses in Arabidopsis. *Diversity* **2022**, *14*, 228. [CrossRef]
- 175. Omidbakhshfard, M.A.; Proost, S.; Fujikura, U.; Mueller-Roeber, B. Growth-regulating factors (GRFs): A small transcription factor family with important functions in plant biology. *Mol. Plant* **2015**, *8*, 998–1010. [CrossRef] [PubMed]
- 176. Li, S.; Gao, F.; Xie, K.; Zeng, X.; Cao, Y.; Zeng, J.; He, Z.; Ren, Y.; Li, W.; Deng, Q. The OsmiR396c-OsGRF4-OsGIF1 regulatory module determines grain size and yield in rice. *Plant Biotechnol. J.* **2016**, *14*, 2134–2146. [CrossRef] [PubMed]
- 177. Hu, J.; Wang, Y.; Fang, Y.; Zeng, L.; Xu, J.; Yu, H.; Shi, Z.; Pan, J.; Zhang, D.; Kang, S. A rare allele of GS2 enhances grain size and grain yield in rice. *Mol. Plant* **2015**, *8*, 1455–1465. [CrossRef] [PubMed]
- 178. Che, R.; Tong, H.; Shi, B.; Liu, Y.; Fang, S.; Liu, D.; Xiao, Y.; Hu, B.; Liu, L.; Wang, H. Control of grain size and rice yield by GL2-mediated brassinosteroid responses. *Nat. Plants* **2015**, *2*, 15195. [CrossRef]
- 179. Duan, P.; Ni, S.; Wang, J.; Zhang, B.; Xu, R.; Wang, Y.; Chen, H.; Zhu, X.; Li, Y. Regulation of OsGRF4 by OsmiR396 controls grain size and yield in rice. *Nat. Plants* **2015**, *2*, 15203. [CrossRef]
- 180. Sun, P.; Zhang, W.; Wang, Y.; He, Q.; Shu, F.; Liu, H.; Wang, J.; Wang, J.; Yuan, L.; Deng, H. OsGRF4 controls grain shape, panicle length and seed shattering in rice. *J. Integr. Plant Biol.* **2016**, *58*, 836–847. [CrossRef]
- 181. Chen, X.; Jiang, L.; Zheng, J.; Chen, F.; Wang, T.; Wang, M.; Tao, Y.; Wang, H.; Hong, Z.; Huang, Y. A missense mutation in Large Grain Size 1 increases grain size and enhances cold tolerance in rice. *J. Exp. Bot.* **2019**, *70*, 3851–3866. [CrossRef]
- 182. Shimano, S.; Hibara, K.-I.; Furuya, T.; Arimura, S.-I.; Tsukaya, H.; Itoh, J.-I. Conserved functional control, but distinct regulation, of cell proliferation in rice and Arabidopsis leaves revealed by comparative analysis of GRF-INTERACTING FACTOR 1 orthologs. Development 2018, 145, dev159624. [CrossRef]
- 183. He, Z.; Zeng, J.; Ren, Y.; Chen, D.; Li, W.; Gao, F.; Cao, Y.; Luo, T.; Yuan, G.; Wu, X. OsGIF1 positively regulates the sizes of stems, leaves, and grains in rice. *Front. Plant Sci.* 2017, *8*, 1730. [CrossRef]
- 184. Li, S.; Tian, Y.; Wu, K.; Ye, Y.; Yu, J.; Zhang, J.; Liu, Q.; Hu, M.; Li, H.; Tong, Y. Modulating plant growth–metabolism coordination for sustainable agriculture. *Nature* **2018**, *560*, 595–600. [CrossRef] [PubMed]
- 185. Singh, S.; Koyama, H.; Bhati, K.K.; Alok, A. The biotechnological importance of the plant-specific NAC transcription factor family in crop improvement. *J. Plant Res.* 2021, 134, 475–495. [CrossRef] [PubMed]
- 186. Chen, Q.; Jing, D.; Wang, S.; Xu, F.; Bao, C.; Luo, M.; Guo, Q. The Putative Role of the NAC Transcription Factor EjNACL47 in Cell Enlargement of Loquat (*Eriobotrya japonica* Lindl.). *Horticulturae* **2021**, *7*, 323. [CrossRef]
- 187. Mathew, I.E.; Das, S.; Mahto, A.; Agarwal, P. Three rice NAC transcription factors heteromerize and are associated with seed size. *Front. Plant Sci.* **2016**, *7*, 1638. [CrossRef] [PubMed]
- 188. Liu, Z.; Li, N.; Zhang, Y.; Li, Y. Transcriptional repression of GIF1 by the KIX-PPD-MYC repressor complex controls seed size in Arabidopsis. *Nat. Commun.* **2020**, *11*, 1846. [CrossRef]
- 189. Yang, T.; Guo, L.; Ji, C.; Wang, H.; Wang, J.; Zheng, X.; Xiao, Q.; Wu, Y. The B3 domain-containing transcription factor ZmABI19 coordinates expression of key factors required for maize seed development and grain filling. *Plant Cell* 2021, 33, 104–128. [CrossRef]
- 190. Swinnen, G.; Mauxion, J.-P.; Baekelandt, A.; De Clercq, R.; Van Doorsselaere, J.; Inzé, D.; Gonzalez, N.; Goossens, A.; Pauwels, L. SIKIX8 and SIKIX9 are negative regulators of leaf and fruit growth in tomato. *bioRxiv* 2021, *188*, 382–396. [CrossRef]
- 191. Wang, S.; Wu, K.; Yuan, Q.; Liu, X.; Liu, Z.; Lin, X.; Zeng, R.; Zhu, H.; Dong, G.; Qian, Q. Control of grain size, shape and quality by OsSPL16 in rice. *Nat. Genet.* 2012, 44, 950–954. [CrossRef]

- 192. Wang, S.; Li, S.; Liu, Q.; Wu, K.; Zhang, J.; Wang, S.; Wang, Y.; Chen, X.; Zhang, Y.; Gao, C. The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* **2015**, *47*, 949–954. [CrossRef]
- 193. Pan, Y.H.; Gao, L.J.; Liang, Y.T.; Zhao, Y.; Liang, H.F.; Chen, W.W.; Yang, X.H.; Qing, D.J.; Gao, J.; Wu, H. OrMKK3 Influences Morphology and Grain Size in Rice. *J. Plant Biol.* **2021**, 1–14. [CrossRef]
- 194. Huang, L.; Hua, K.; Xu, R.; Zeng, D.; Wang, R.; Dong, G.; Zhang, G.; Lu, X.; Fang, N.; Wang, D. The LARGE2-APO1/APO2 regulatory module controls panicle size and grain number in rice. *Plant Cell* **2021**, *33*, 1212–1228. [CrossRef] [PubMed]
- 195. Blázquez, M.A.; Nelson, D.C.; Weijers, D. Evolution of plant hormone response pathways. *Annu. Rev. Plant Biol.* 2020, 71, 327–353. [CrossRef]
- 196. Figueiredo, D.D.; Köhler, C. Auxin: A molecular trigger of seed development. Genes Dev. 2018, 32, 479–490. [CrossRef] [PubMed]
- 197. Mano, Y.; Nemoto, K. The pathway of auxin biosynthesis in plants. J. Exp. Bot. 2012, 63, 2853–2872. [CrossRef] [PubMed]
- 198. Won, C.; Shen, X.; Mashiguchi, K.; Zheng, Z.; Dai, X.; Cheng, Y.; Kasahara, H.; Kamiya, Y.; Chory, J.; Zhao, Y. Conversion of tryptophan to indole-3-acetic acid by TRYPTOPHAN AMINOTRANSFERASES OF ARABIDOPSIS and YUCCAs in Arabidopsis. *Proc. Natl. Acad. Sci. USA* 2011, 108, 18518–18523. [CrossRef]
- 199. Cheng, Y.; Dai, X.; Zhao, Y. Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. *Genes Dev.* **2006**, *20*, 1790–1799. [CrossRef]
- 200. Stone, S.L.; Braybrook, S.A.; Paula, S.L.; Kwong, L.W.; Meuser, J.; Pelletier, J.; Hsieh, T.-F.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl. Acad. Sci. USA* 2008, 105, 3151–3156. [CrossRef]
- 201. Wójcikowska, B.; Jaskóła, K.; Gąsiorek, P.; Meus, M.; Nowak, K.; Gaj, M.D. LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of Arabidopsis, via YUCCA-mediated auxin biosynthesis. *Planta* **2013**, *238*, 425–440. [CrossRef]
- 202. Junker, A.; Mönke, G.; Rutten, T.; Keilwagen, J.; Seifert, M.; Thi, T.M.N.; Renou, J.P.; Balzergue, S.; Viehöver, P.; Hähnel, U. Elongation-related functions of LEAFY COTYLEDON1 during the development of *Arabidopsis thaliana*. *Plant J.* 2012, 71, 427–442. [CrossRef]
- 203. Guilfoyle, T.J.; Hagen, G. Auxin response factors. Curr. Opin. Plant Biol. 2007, 10, 453-460. [CrossRef]
- 204. Schruff, M.C.; Spielman, M.; Tiwari, S.; Adams, S.; Fenby, N.; Scott, R.J. The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. *Development* **2006**, *133*, 251–261. [CrossRef] [PubMed]
- Goetz, M.; Vivian-Smith, A.; Johnson, S.D.; Koltunow, A.M. AUXIN RESPONSE FACTOR8 is a negative regulator of fruit initiation in Arabidopsis. *Plant Cell* 2006, 18, 1873–1886. [CrossRef] [PubMed]
- 206. Kelley, D.R.; Arreola, A.; Gallagher, T.L.; Gasser, C.S. ETTIN (ARF3) physically interacts with KANADI proteins to form a functional complex essential for integument development and polarity determination in Arabidopsis. *Development* 2012, 139, 1105–1109. [CrossRef] [PubMed]
- 207. Reed, J.W.; Wu, M.-F.; Reeves, P.H.; Hodgens, C.; Yadav, V.; Hayes, S.; Pierik, R. Three auxin response factors promote hypocotyl elongation. *Plant Physiol.* 2018, 178, 864–875. [CrossRef] [PubMed]
- 208. Li, Y.J.; Yu, Y.; Liu, X.; Zhang, X.S.; Su, Y.H. The Arabidopsis MATERNAL EFFECT EMBRYO ARREST45 protein modulates maternal auxin biosynthesis and controls seed size by inducing AINTEGUMENTA. *Plant Cell* **2021**, *33*, 1907–1926. [CrossRef]
- 209. Shi, L.; Song, J.; Guo, C.; Wang, B.; Guan, Z.; Yang, P.; Chen, X.; Zhang, Q.; King, G.J.; Wang, J. A CACTA-like transposable element in the upstream region of BnaA9. CYP 78A9 acts as an enhancer to increase silique length and seed weight in rapeseed. *Plant J.* **2019**, *98*, 524–539. [CrossRef]
- Guo, L.; Ma, M.; Wu, L.; Zhou, M.; Li, M.; Wu, B.; Li, L.; Liu, X.; Jing, R.; Chen, W. Modified expression of TaCYP78A5 enhances grain weight with yield potential by accumulating auxin in wheat (*Triticum Aestivum* L.). *Plant Biotechnol. J.* 2022, 20, 168–182. [CrossRef]
- Rijavec, T.; Jain, M.; Dermastia, M.; Chourey, P.S. Spatial and temporal profiles of cytokinin biosynthesis and accumulation in developing caryopses of maize. *Ann. Bot.* 2011, 107, 1235–1245. [CrossRef] [PubMed]
- Wang, X.; Ding, J.; Lin, S.; Liu, D.; Gu, T.; Wu, H.; Trigiano, R.N.; McAvoy, R.; Huang, J.; Li, Y. Evolution and roles of cytokinin genes in angiosperms 2: Do ancient CKXs play housekeeping roles while non-ancient CKXs play regulatory roles? *Hortic. Res.* 2020, 7, 29. [CrossRef] [PubMed]
- 213. Holubová, K.; Hensel, G.; Vojta, P.; Tarkowski, P.; Bergougnoux, V.; Galuszka, P. Modification of barley plant productivity through regulation of cytokinin content by reverse-genetics approaches. *Front. Plant Sci.* **2018**, *9*, 1676. [CrossRef]
- 214. Gasparis, S.; Przyborowski, M.; Kała, M.; Nadolska-Orczyk, A. Knockout of the HvCKX1 or HvCKX3 gene in barley (*Hordeum vulgare* L.) by RNA-guided Cas9 nuclease affects the regulation of cytokinin metabolism and root morphology. *Cells* 2019, *8*, 782. [CrossRef] [PubMed]
- Mao, C.; He, J.; Liu, L.; Deng, Q.; Yao, X.; Liu, C.; Qiao, Y.; Li, P.; Ming, F. OsNAC2 integrates auxin and cytokinin pathways to modulate rice root development. *Plant Biotechnol. J.* 2020, *18*, 429–442. [CrossRef] [PubMed]
- 216. Lu, J.; Chang, C.; Zhang, H.-P.; Wang, S.-X.; Sun, G.; Xiao, S.-H.; Ma, C.-X. Identification of a novel allele of TaCKX6a02 associated with grain size, filling rate and weight of common wheat. *PLoS ONE* **2015**, *10*, e0144765. [CrossRef] [PubMed]
- 217. Chen, L.; Zhao, J.; Song, J.; Jameson, P.E. Cytokinin dehydrogenase: A genetic target for yield improvement in wheat. *Plant Biotechnol. J.* **2020**, *18*, 614–630. [CrossRef] [PubMed]

- 218. Park, J.; Lee, S.; Park, G.; Cho, H.; Choi, D.; Umeda, M.; Choi, Y.; Hwang, D.; Hwang, I. C YTO K ININ-RESPONSIVE G ROWTH REGULATOR regulates cell expansion and cytokinin-mediated cell cycle progression. *Plant Physiol.* 2021, 186, 1734–1746. [CrossRef]
- Werner, S.; Bartrina, I.; Novák, O.; Strnad, M.; Werner, T.; Schmülling, T. The cytokinin status of the epidermis regulates aspects of vegetative and reproductive development in *Arabidopsis thaliana*. *Front. Plant Sci.* 2021, 12, 613488. [CrossRef]
- 220. Nedvěd, D.; Hošek, P.; Klíma, P.; Hoyerová, K. Differential subcellular distribution of cytokinins: How does membrane transport fit into the big picture? *Int. J. Mol. Sci.* 2021, 22, 3428. [CrossRef]
- Day, R.C.; Herridge, R.P.; Ambrose, B.A.; Macknight, R.C. Transcriptome analysis of proliferating Arabidopsis endosperm reveals biological implications for the control of syncytial division, cytokinin signaling, and gene expression regulation. *Plant Physiol.* 2008, 148, 1964–1984. [CrossRef]
- 222. Belmonte, M.F.; Kirkbride, R.C.; Stone, S.L.; Pelletier, J.M.; Bui, A.Q.; Yeung, E.C.; Hashimoto, M.; Fei, J.; Harada, C.M.; Munoz, M.D. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl. Acad. Sci. USA* 2013, *110*, E435–E444. [CrossRef]
- 223. Song, J.; Jiang, L.; Jameson, P.E. Expression patterns of *Brassica napus* genes implicate IPT, CKX, sucrose transporter, cell wall invertase, and amino acid permease gene family members in leaf, flower, silique, and seed development. *J. Exp. Bot.* 2015, 66, 5067–5082. [CrossRef]
- 224. Ma, Q.-H.; Wang, X.-M.; Wang, Z.-M. Expression of isopentenyl transferase gene controlled by seed-specific lectin promoter in transgenic tobacco influences seed development. *J. Plant Growth Regul.* 2008, 27, 68–76. [CrossRef]
- Motyka, V.; Vaňková, R.; Čapková, V.; Petrášek, J.; Kamínek, M.; Schmülling, T. Cytokinin-induced upregulation of cytokinin oxidase activity in tobacco includes changes in enzyme glycosylation and secretion. *Physiol. Plant.* 2003, 117, 11–21. [CrossRef]
- 226. Yu, H.Q.; Sun, F.A.; Feng, W.Q.; Lu, F.Z.; Li, W.C.; Fu, F.L. The BES1/BZR1 transcription factors regulate growth, development and stress resistance in plants. *Yi Chuan* **2019**, *41*, 206–214.
- 227. Song, J.; Jiang, L.; Jameson, P.E. Co-ordinate regulation of cytokinin gene family members during flag leaf and reproductive development in wheat. *BMC Plant Biol.* **2012**, *12*, 78. [CrossRef] [PubMed]
- 228. Deng, Y.; Dong, H.; Mu, J.; Ren, B.; Zheng, B.; Ji, Z.; Yang, W.-C.; Liang, Y.; Zuo, J. Arabidopsis histidine kinase CKI1 acts upstream of histidine phosphotransfer proteins to regulate female gametophyte development and vegetative growth. *Plant Cell* 2010, 22, 1232–1248. [CrossRef] [PubMed]
- 229. Li, S.; Zhao, B.; Yuan, D.; Duan, M.; Qian, Q.; Tang, L.; Wang, B.; Liu, X.; Zhang, J.; Wang, J. Rice zinc finger protein DST enhances grain production through controlling Gn1a/OsCKX2 expression. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 3167–3172. [CrossRef]
- 230. Nolan, T.M.; Vukašinović, N.; Liu, D.; Russinova, E.; Yin, Y. Brassinosteroids: Multidimensional regulators of plant growth, development, and stress responses. *Plant Cell* **2020**, *32*, 295–318. [CrossRef]
- 231. Zhao, B.; Lv, M.; Feng, Z.; Campbell, T.; Liscum, E.; Li, J. TWISTED DWARF 1 associates with BRASSINOSTEROID-INSENSITIVE 1 to regulate early events of the brassinosteroid signaling pathway. *Mol. Plant* **2016**, *9*, 582–592. [CrossRef]
- Tong, H.; Chu, C. Functional specificities of brassinosteroid and potential utilization for crop improvement. *Trends Plant Sci.* 2018, 23, 1016–1028. [CrossRef]
- 233. Xiong, M.; Yu, J.; Wang, J.; Gao, Q.; Huang, L.; Chen, C.; Zhang, C.; Fan, X.; Zhao, D.; Liu, Q.-Q. Brassinosteroids regulate rice seed germination through the BZR1-RAmy3D transcriptional module. *Plant Physiol.* **2022**, *189*, 402–418. [CrossRef]
- 234. Wu, Y.; Fu, Y.; Zhao, S.; Gu, P.; Zhu, Z.; Sun, C.; Tan, L. Clustered primary branch 1, a new allele of DWARF 11, controls panicle architecture and seed size in rice. *Plant Biotechnol. J.* **2016**, *14*, 377–386. [CrossRef] [PubMed]
- 235. Wu, C.-y.; Trieu, A.; Radhakrishnan, P.; Kwok, S.F.; Harris, S.; Zhang, K.; Wang, J.; Wan, J.; Zhai, H.; Takatsuto, S. Brassinosteroids regulate grain filling in rice. *Plant Cell* **2008**, *20*, 2130–2145. [CrossRef] [PubMed]
- 236. Hong, Z.; Ueguchi-Tanaka, M.; Shimizu-Sato, S.; Inukai, Y.; Fujioka, S.; Shimada, Y.; Takatsuto, S.; Agetsuma, M.; Yoshida, S.; Watanabe, Y. Loss-of-function of a rice brassinosteroid biosynthetic enzyme, C-6 oxidase, prevents the organized arrangement and polar elongation of cells in the leaves and stem. *Plant J.* 2002, *32*, 495–508. [CrossRef] [PubMed]
- 237. Mori, M.; Nomura, T.; Ooka, H.; Ishizaka, M.; Yokota, T.; Sugimoto, K.; Okabe, K.; Kajiwara, H.; Satoh, K.; Yamamoto, K. Isolation and characterization of a rice dwarf mutant with a defect in brassinosteroid biosynthesis. *Plant Physiol.* 2002, 130, 1152–1161. [CrossRef] [PubMed]
- 238. Jiang, W.-B.; Huang, H.-Y.; Hu, Y.-W.; Zhu, S.-W.; Wang, Z.-Y.; Lin, W.-H. Brassinosteroid regulates seed size and shape in Arabidopsis. *Plant Physiol.* **2013**, *162*, 1965–1977. [CrossRef]
- Zhu, X.; Liang, W.; Cui, X.; Chen, M.; Yin, C.; Luo, Z.; Zhu, J.; Lucas, W.J.; Wang, Z.; Zhang, D. Brassinosteroids promote development of rice pollen grains and seeds by triggering expression of Carbon Starved Anther, a MYB domain protein. *Plant J.* 2015, *82*, 570–581. [CrossRef]
- 240. Tanabe, S.; Ashikari, M.; Fujioka, S.; Takatsuto, S.; Yoshida, S.; Yano, M.; Yoshimura, A.; Kitano, H.; Matsuoka, M.; Fujisawa, Y. A novel cytochrome P450 is implicated in brassinosteroid biosynthesis via the characterization of a rice dwarf mutant, dwarf11, with reduced seed length. *Plant Cell* 2005, 17, 776–790. [CrossRef]
- 241. Feng, Z.; Wu, C.; Wang, C.; Roh, J.; Zhang, L.; Chen, J.; Zhang, S.; Zhang, H.; Yang, C.; Hu, J. SLG controls grain size and leaf angle by modulating brassinosteroid homeostasis in rice. *J. Exp. Bot.* **2016**, *67*, 4241–4253. [CrossRef]

- 242. Hong, Z.; Ueguchi-Tanaka, M.; Fujioka, S.; Takatsuto, S.; Yoshida, S.; Hasegawa, Y.; Ashikari, M.; Kitano, H.; Matsuoka, M. The rice brassinosteroid-deficient dwarf2 mutant, defective in the rice homolog of Arabidopsis DIMINUTO/DWARF1, is rescued by the endogenously accumulated alternative bioactive brassinosteroid, dolichosterone. *Plant Cell* 2005, 17, 2243–2254. [CrossRef]
- 243. Tian, P.; Liu, J.; Mou, C.; Shi, C.; Zhang, H.; Zhao, Z.; Lin, Q.; Wang, J.; Wang, J.; Zhang, X. GW5-Like, a homolog of GW5, negatively regulates grain width, weight and salt resistance in rice. *J. Integr. Plant Biol.* **2019**, *61*, 1171–1185. [CrossRef]
- 244. Yang, C.; Shen, W.; He, Y.; Tian, Z.; Li, J. OVATE family protein 8 positively mediates brassinosteroid signaling through interacting with the GSK3-like kinase in rice. *PLoS Genet.* 2016, *12*, e1006118. [CrossRef] [PubMed]
- 245. Tilman, D.; Balzer, C.; Hill, J.; Befort, B.L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. USA* 2011, *108*, 20260–20264. [CrossRef] [PubMed]
- Xiao, Y.; Zhang, G.; Liu, D.; Niu, M.; Tong, H.; Chu, C. GSK2 stabilizes OFP3 to suppress brassinosteroid responses in rice. *Plant J.* 2020, 102, 1187–1201. [CrossRef] [PubMed]
- 247. Li, Y.; Huang, R.; Li, J.; Huang, X.; Zeng, X.; Zhao, D. DWARF and SMALL SEED1, a Novel Allele of OsDWARF, Controls Rice Plant Architecture, Seed Size, and Chlorophyll Biosynthesis. *Phyton* **2021**, *90*, 111. [CrossRef]
- 248. Dong, N.; Yin, W.; Liu, D.; Zhang, X.; Yu, Z.; Huang, W.; Liu, J.; Yang, Y.; Meng, W.; Niu, M. Regulation of brassinosteroid signaling and salt resistance by SERK2 and potential utilization for crop improvement in rice. *Front. Plant Sci.* 2020, 11, 621859. [CrossRef]
- 249. Wasternack, C.; Hause, B. Jasmonates: Biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Ann. Bot.* **2013**, *111*, 1021–1058. [CrossRef]
- Yuan, Z.; Zhang, D. Roles of jasmonate signalling in plant inflorescence and flower development. *Curr. Opin. Plant Biol.* 2015, 27, 44–51. [CrossRef]
- 251. Li, C.; Liu, G.; Xu, C.; Lee, G.I.; Bauer, P.; Ling, H.-Q.; Ganal, M.W.; Howe, G.A. The tomato suppressor of prosystemin-mediated responses2 gene encodes a fatty acid desaturase required for the biosynthesis of jasmonic acid and the production of a systemic wound signal for defense gene expression. *Plant Cell* **2003**, *15*, 1646–1661. [CrossRef]
- Lyons, R.; Manners, J.M.; Kazan, K. Jasmonate biosynthesis and signaling in monocots: A comparative overview. *Plant Cell Rep.* 2013, 32, 815–827. [CrossRef]
- 253. Browse, J. The power of mutants for investigating jasmonate biosynthesis and signaling. *Phytochemistry* **2009**, *70*, 1539–1546. [CrossRef]
- 254. Goetz, S.; Hellwege, A.; Stenzel, I.; Kutter, C.; Hauptmann, V.; Forner, S.; McCaig, B.; Hause, G.; Miersch, O.; Wasternack, C. Role of cis-12-oxo-phytodienoic acid in tomato embryo development. *Plant Physiol.* **2012**, *158*, 1715–1727. [CrossRef] [PubMed]
- 255. Li, L.; Zhao, Y.; McCaig, B.C.; Wingerd, B.A.; Wang, J.; Whalon, M.E.; Pichersky, E.; Howe, G.A. The tomato homolog of CORONATINE-INSENSITIVE1 is required for the maternal control of seed maturation, jasmonate-signaled defense responses, and glandular trichome development. *Plant Cell* 2004, *16*, 126–143. [CrossRef] [PubMed]
- Cai, Q.; Yuan, Z.; Chen, M.; Yin, C.; Luo, Z.; Zhao, X.; Liang, W.; Hu, J.; Zhang, D. Jasmonic acid regulates spikelet development in rice. *Nat. Commun.* 2014, *5*, 3476. [CrossRef]
- 257. Acosta, I.F.; Laparra, H.; Romero, S.P.; Schmelz, E.; Hamberg, M.; Mottinger, J.P.; Moreno, M.A.; Dellaporta, S.L. tasselseed1 is a lipoxygenase affecting jasmonic acid signaling in sex determination of maize. *Science* **2009**, *323*, 262–265. [CrossRef]
- 258. DeLong, A.; Calderon-Urrea, A.; Dellaporta, S.L. Sex determination gene TASSELSEED2 of maize encodes a short-chain alcohol dehydrogenase required for stage-specific floral organ abortion. *Cell* **1993**, *74*, 757–768. [CrossRef]
- 259. Yan, Y.; Christensen, S.; Isakeit, T.; Engelberth, J.; Meeley, R.; Hayward, A.; Emery, R.N.; Kolomiets, M.V. Disruption of OPR7 and OPR8 reveals the versatile functions of jasmonic acid in maize development and defense. *Plant Cell* 2012, 24, 1420–1436. [CrossRef]
- 260. Chini, A.; Fonseca, S.; Fernandez, G.; Adie, B.; Chico, J.; Lorenzo, O.; García-Casado, G.; López-Vidriero, I.; Lozano, F.; Ponce, M. The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* 2007, 448, 666–671. [CrossRef]
- 261. Yan, J.; Zhang, C.; Gu, M.; Bai, Z.; Zhang, W.; Qi, T.; Cheng, Z.; Peng, W.; Luo, H.; Nan, F. The Arabidopsis CORONATINE INSENSITIVE1 protein is a jasmonate receptor. *Plant Cell* **2009**, *21*, 2220–2236. [CrossRef]
- Ali, M.S.; Baek, K.-H. Jasmonic acid signaling pathway in response to abiotic stresses in plants. *Int. J. Mol. Sci.* 2020, 21, 621. [CrossRef]
- 263. Mehra, P.; Pandey, B.K.; Verma, L.; Prusty, A.; Singh, A.P.; Sharma, S.; Malik, N.; Bennett, M.J.; Parida, S.K.; Giri, J.; et al. OsJAZ11 regulates spikelet and seed development in rice. *Plant Direct* **2022**, *6*, e401. [CrossRef]
- 264. Gnan, S.; Priest, A.; Kover, P.X. The genetic basis of natural variation in seed size and seed number and their trade-off using *Arabidopsis thaliana* MAGIC lines. *Genetics* **2014**, *198*, 1751–1758. [CrossRef] [PubMed]
- Collard, B.C.; Mackill, D.J. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* 2008, 363, 557–572. [CrossRef] [PubMed]
- 266. Xu, Y.; Crouch, J.H. Marker-assisted selection in plant breeding: From publications to practice. Crop Sci. 2008, 48, 391–407. [CrossRef]
- 267. William, H.M.; Morris, M.; Warburton, M.; Hoisington, D.A. Technical, economic and policy considerations on marker-assisted selection in crops: Lessons from the experience at an international agricultural research centre. *Marker-Assist. Sel.* 2007, *19*, 381.

- 268. Crosbie, T.M.; Eathington, S.R.; Johnson Sr, G.R.; Edwards, M.; Reiter, R.; Stark, S.; Mohanty, R.G.; Oyervides, M.; Buehler, R.E.; Walker, A.K. Plant breeding: Past, present, and future. In *Plant Breeding: The Arnel R. Hallauer International Symposium*; Wiley: Hoboken, NJ, USA, 2006; pp. 3–50.
- Eathington, S.R.; Crosbie, T.M.; Edwards, M.D.; Reiter, R.S.; Bull, J.K. Molecular markers in a commercial breeding program. Crop Sci. 2007, 47, S-154–S-163. [CrossRef]
- 270. Gregorio, G.; Islam, M.; Vergara, G.; Thirumeni, S. Recent advances in rice science to design salinity and other abiotic stress tolerant rice varieties. *SABRAO J. Breed. Genet.* **2013**, *45*, 31–40.
- 271. Heffner, E.L.; Sorrells, M.E.; Jannink, J.-L. Genomic selection for crop improvement. Crop Sci. 2009, 49, 1–12. [CrossRef]
- 272. Desta, Z.A.; Ortiz, R. Genomic selection: Genome-wide prediction in plant improvement. *Trends Plant Sci.* 2014, 19, 592–601. [CrossRef]
- Heslot, N.; Jannink, J.L.; Sorrells, M.E. Perspectives for genomic selection applications and research in plants. *Crop Sci.* 2015, 55, 1–12. [CrossRef]
- Sandhu, K.S.; Mihalyov, P.D.; Lewien, M.J.; Pumphrey, M.O.; Carter, A.H. Combining genomic and phenomic information for predicting grain protein content and grain yield in spring wheat. *Front. Plant Sci.* 2021, 12, 170. [CrossRef]
- 275. Li, Y.-Y.; Tao, H.-J.; Zhao, X.-Q.; Jie, X.; Li, G.-M.; Hu, S.-K.; Dong, G.-J.; Shi, Z.-Y.; Wu, L.-W.; Jiang, H. Molecular improvement of grain weight and yield in rice by using GW6 gene. *Rice Sci.* 2014, 21, 127–132. [CrossRef]
- 276. Sun, L.; Ma, D.; Yu, H.; Zhou, F.; Li, Y.; Luo, L.; Gao, G.; Zhang, Q.; Xu, C.; He, Y. Identification of quantitative trait loci for grain size and the contributions of major grain-size QTLs to grain weight in rice. *Mol. Breed.* **2013**, *31*, 451–461. [CrossRef]
- 277. Gao, X.; Zhang, X.; Lan, H.; Huang, J.; Wang, J.; Zhang, H. The additive effects of GS3 and qGL3 on rice grain length regulation revealed by genetic and transcriptome comparisons. *BMC Plant Biol.* **2015**, *15*, 156. [CrossRef]
- 278. Wang, Y.; Xiong, G.; Hu, J.; Jiang, L.; Yu, H.; Xu, J.; Fang, Y.; Zeng, L.; Xu, E.; Xu, J. Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat. Genet.* **2015**, *47*, 944–948. [CrossRef] [PubMed]
- 279. Zhao, D.-S.; Li, Q.-F.; Zhang, C.-Q.; Zhang, C.; Yang, Q.-Q.; Pan, L.-X.; Ren, X.-Y.; Lu, J.; Gu, M.-H.; Liu, Q.-Q. GS9 acts as a transcriptional activator to regulate rice grain shape and appearance quality. *Nat. Commun.* **2018**, *9*, 1240. [CrossRef]
- 280. Dong, N.-Q.; Sun, Y.; Guo, T.; Shi, C.-L.; Zhang, Y.-M.; Kan, Y.; Xiang, Y.-H.; Zhang, H.; Yang, Y.-B.; Li, Y.-C. UDPglucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. *Nat. Commun.* **2020**, *11*, 2629. [CrossRef]
- 281. Song, X.J.; Kuroha, T.; Ayano, M.; Furuta, T.; Nagai, K.; Komeda, N.; Segami, S.; Miura, K.; Ogawa, D.; Kamura, T. Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. *Proc. Natl. Acad. Sci.* USA 2015, 112, 76–81. [CrossRef]
- Ruan, B.; Shang, L.; Zhang, B.; Hu, J.; Wang, Y.; Lin, H.; Zhang, A.; Liu, C.; Peng, Y.; Zhu, L. Natural variation in the promoter of TGW2 determines grain width and weight in rice. *New Phytol.* 2020, 227, 629–640. [CrossRef]
- 283. Li, Y.; Fan, C.; Xing, Y.; Jiang, Y.; Luo, L.; Sun, L.; Shao, D.; Xu, C.; Li, X.; Xiao, J. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nat. Genet.* **2011**, *43*, 1266–1269. [CrossRef]
- 284. Shi, C.L.; Dong, N.Q.; Guo, T.; Ye, W.W.; Shan, J.X.; Lin, H.X. A quantitative trait locus GW6 controls rice grain size and yield through the gibberellin pathway. *Plant J.* **2020**, *103*, 1174–1188. [CrossRef]
- 285. Yu, J.; Xiong, H.; Zhu, X.; Zhang, H.; Li, H.; Miao, J.; Wang, W.; Tang, Z.; Zhang, Z.; Yao, G. OsLG3 contributing to rice grain length and yield was mined by Ho-LAMap. *BMC Biol.* **2017**, *15*, 28. [CrossRef] [PubMed]
- 286. Yu, J.; Miao, J.; Zhang, Z.; Xiong, H.; Zhu, X.; Sun, X.; Pan, Y.; Liang, Y.; Zhang, Q.; Abdul Rehman, R.M.; et al. Alternative splicing of OsLG3b controls grain length and yield in japonica rice. *Plant Biotechnol. J.* 2018, 16, 1667–1678. [CrossRef] [PubMed]
- 287. Qi, P.; Lin, Y.-S.; Song, X.-J.; Shen, J.-B.; Huang, W.; Shan, J.-X.; Zhu, M.-Z.; Jiang, L.; Gao, J.-P.; Lin, H.-X. The novel quantitative trait locus GL3. 1 controls rice grain size and yield by regulating Cyclin-T1; 3. *Cell Res.* **2012**, 22, 1666–1680. [CrossRef]
- 288. Ying, J.-Z.; Ma, M.; Bai, C.; Huang, X.-H.; Liu, J.-L.; Fan, Y.-Y.; Song, X.-J. TGW3, a major QTL that negatively modulates grain length and weight in rice. *Mol. Plant* 2018, *11*, 750–753. [CrossRef]
- 289. Xia, D.; Zhou, H.; Liu, R.; Dan, W.; Li, P.; Wu, B.; Chen, J.; Wang, L.; Gao, G.; Zhang, Q. GL3. 3, a novel QTL encoding a GSK3/SHAGGY-like kinase, epistatically interacts with GS3 to produce extra-long grains in rice. *Mol. Plant* 2018, 11, 754–756. [CrossRef]
- 290. Wu, W.; Liu, X.; Wang, M.; Meyer, R.; Luo, X.; Ndjiondjop, M.; Tan, L.; Zhang, J.; Wu, J.; Cai, H. A single-nucleotide polymorphism causes smaller grain size and loss of seed shattering during African rice domestication. *Nat. Plants* 2017, *3*, 17064. [CrossRef] [PubMed]
- 291. Ishimaru, K.; Hirotsu, N.; Madoka, Y.; Murakami, N.; Hara, N.; Onodera, H.; Kashiwagi, T.; Ujiie, K.; Shimizu, B.-I.; Onishi, A. Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nat. Genet.* 2013, 45, 707–711. [CrossRef]
- Wang, A.; Hou, Q.; Si, L.; Huang, X.; Luo, J.; Lu, D.; Zhu, J.; Shangguan, Y.; Miao, J.; Xie, Y. The PLATZ transcription factor GL6 affects grain length and number in rice. *Plant Physiol.* 2019, 180, 2077–2090. [CrossRef] [PubMed]
- 293. Savadi, S.; Naresh, V.; Kumar, V.; Bhat, S.R. Seed-specific overexpression of Arabidopsi s DGAT1 in Indian mustard (*Brassica Juncea*) increases seed oil content and seed weight. *Botany* **2016**, *94*, 177–184. [CrossRef]

- 294. Kim, H.; Park, J.H.; Kim, A.Y.; Suh, M.C. Functional analysis of diacylglycerol acyltransferase1 genes from *Camelina sativa* and effects of CsDGAT1B overexpression on seed mass and storage oil content in *C. sativa*. *Plant Biotechnol. Rep.* 2016, 10, 141–153. [CrossRef]
- 295. Liu, B.; Hua, C.; Song, G.; Wu, M.; Cui, R.; Zhang, A.; Liu, Y.; Huang, L.; Yan, A.; Ali, I. The SPATULA transcription factor regulates seed oil content by controlling seed specific genes in *Arabidopsis thaliana*. *Plant Growth Regul.* 2017, *82*, 111–121. [CrossRef]
- 296. Le, D.T.; Nishiyama, R.; Watanabe, Y.; Vankova, R.; Tanaka, M.; Seki, M.; Yamaguchi-Shinozaki, K.; Shinozaki, K.; Tran, L.-S.P. Identification and expression analysis of cytokinin metabolic genes in soybean under normal and drought conditions in relation to cytokinin levels. *PLoS ONE* 2012, 7, e42411. [CrossRef] [PubMed]
- 297. Jameson, P.E.; Song, J. Cytokinin: A key driver of seed yield. J. Exp. Bot. 2016, 67, 593-606. [CrossRef] [PubMed]
- 298. Zhao, B.; Dai, A.; Wei, H.; Yang, S.; Wang, B.; Jiang, N.; Feng, X. ArabidopsisKLU homologue GmCYP78A72 regulates seed size in soybean. *Plant Mol. Biol.* 2016, *90*, 33–47. [CrossRef] [PubMed]
- Sasaki, K.; Kim, M.-H.; Kanno, Y.; Seo, M.; Kamiya, Y.; Imai, R. Arabidopsis COLD SHOCK DOMAIN PROTEIN 2 influences ABA accumulation in seed and negatively regulates germination. *Biochem. Biophys. Res. Commun.* 2015, 456, 380–384. [CrossRef] [PubMed]
- 300. Xiao, Y.G.; Sun, Q.B.; Kang, X.J.; Chen, C.B.; Ni, M. SHORT HYPOCOTYL UNDER BLUE 1 or HAIKU 2 mixepression alters canola and Arabidopsis seed development. *New Phytol.* **2016**, 209, 636–649. [CrossRef]
- 301. Liu, L.; Tong, H.; Xiao, Y.; Che, R.; Xu, F.; Hu, B.; Liang, C.; Chu, J.; Li, J.; Chu, C. Activation of Big Grain1 significantly improves grain size by regulating auxin transport in rice. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 11102–11107. [CrossRef] [PubMed]
- Ge, L.; Yu, J.; Wang, H.; Luth, D.; Bai, G.; Wang, K.; Chen, R. Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. *Proc. Natl. Acad. Sci. USA* 2016, 113, 12414–12419. [CrossRef]
- 303. Savadi, S.; Naresh, V.; Kumar, V.; Dargan, S.; Gupta, N.C.; Chamola, R.; Bhat, S.R. Effect of overexpression of Arabidopsis thaliana SHB1 and KLUH genes on seed weight and yield contributing traits in Indian mustard (*Brassica juncea* L.(Czern.)). Indian J. Genet. Plant Breed 2015, 75, 349–356. [CrossRef]
- Rees, H.A.; Liu, D.R. Base editing: Precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 2018, 19, 770–788. [CrossRef]
- Molla, K.A.; Yang, Y. CRISPR/Cas-mediated base editing: Technical considerations and practical applications. *Trends Biotechnol.* 2019, 37, 1121–1142. [CrossRef] [PubMed]
- 306. Mishra, R.; Joshi, R.K.; Zhao, K. Base editing in crops: Current advances, limitations and future implications. *Plant Biotechnol. J.* 2020, 18, 20–31. [CrossRef] [PubMed]
- 307. Li, S.; Li, J.; He, Y.; Xu, M.; Zhang, J.; Du, W.; Zhao, Y.; Xia, L. Precise gene replacement in rice by RNA transcript-templated homologous recombination. *Nat. Biotechnol.* **2019**, *37*, 445–450. [CrossRef] [PubMed]
- 308. Wang, L.; Kaya, H.B.; Zhang, N.; Rai, R.; Willmann, M.R.; Carpenter, S.C.; Read, A.C.; Martin, F.; Fei, Z.; Leach, J.E. Spelling changes and fluorescent tagging with prime editing vectors for plants. *Front. Genome Ed.* 2021, 3, 617553. [CrossRef] [PubMed]
- Anzalone, A.V.; Randolph, P.B.; Davis, J.R.; Sousa, A.A.; Koblan, L.W.; Levy, J.M.; Chen, P.J.; Wilson, C.; Newby, G.A.; Raguram, A.; et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019, 576, 149–157. [CrossRef]
- 310. Hua, K.; Jiang, Y.; Tao, X.; Zhu, J.K. Precision genome engineering in rice using prime editing system. *Plant Biotechnol. J.* **2020**, *18*, 2167. [CrossRef]
- 311. Li, H.; Li, J.; Chen, J.; Yan, L.; Xia, L. Precise modifications of both exogenous and endogenous genes in rice by prime editing. *Mol. Plant* **2020**, *13*, 671–674. [CrossRef]
- 312. Xu, R.; Li, J.; Liu, X.; Shan, T.; Qin, R.; Wei, P. Development of plant prime-editing systems for precise genome editing. *Plant Commun.* 2020, *1*, 100043. [CrossRef]
- 313. Lin, Q.; Zong, Y.; Xue, C.; Wang, S.; Jin, S.; Zhu, Z.; Wang, Y.; Anzalone, A.V.; Raguram, A.; Doman, J.L.; et al. Prime genome editing in rice and wheat. *Nat. Biotechnol.* 2020, *38*, 582–585. [CrossRef]
- 314. Jiang, Y.-Y.; Chai, Y.-P.; Lu, M.-H.; Han, X.-L.; Lin, Q.; Zhang, Y.; Zhang, Q.; Zhou, Y.; Wang, X.-C.; Gao, C. Prime editing efficiently generates W542L and S621I double mutations in two ALS genes in maize. *Genome Biol.* 2020, 21, 257. [CrossRef]
- 315. Lu, Y.; Tian, Y.; Shen, R.; Yao, Q.; Zhong, D.; Zhang, X.; Zhu, J.K. Precise genome modification in tomato using an improved prime editing system. *Plant Biotechnol. J.* **2021**, *19*, 415. [CrossRef] [PubMed]
- 316. Perroud, P.-F.; Guyon-Debast, A.; Veillet, F.; Kermarrec, M.-P.; Chauvin, L.; Chauvin, J.-E.; Gallois, J.-L.; Nogué, F. Prime Editing in the model plant Physcomitrium patens and its potential in the tetraploid potato. *Plant Sci.* 2022, 316, 111162. [CrossRef] [PubMed]
- 317. Zhao, Y.; Li, L.; Zheng, G.; Jiang, W.; Deng, Z.; Wang, Z.; Lu, Y. CRISPR/dCas9-Mediated Multiplex Gene Repression in Streptomyces. *Biotechnol. J.* 2018, 13, e1800121. [CrossRef] [PubMed]
- 318. Qi, L.S.; Larson, M.H.; Gilbert, L.A.; Doudna, J.A.; Weissman, J.S.; Arkin, A.P.; Lim, W.A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **2013**, *152*, 1173–1183. [CrossRef] [PubMed]
- 319. Jurkowski, T.P.; Ravichandran, M.; Stepper, P. Synthetic epigenetics—Towards intelligent control of epigenetic states and cell identity. *Clin. Epigenetics* 2015, 7, 18. [CrossRef]
- 320. Zafar, S.A.; Zaidi, S.S.-E.-A.; Gaba, Y.; Singla-Pareek, S.L.; Dhankher, O.P.; Li, X.; Mansoor, S.; Pareek, A. Engineering abiotic stress tolerance via CRISPR/Cas-mediated genome editing. *J. Exp. Bot.* **2020**, *71*, 470–479. [CrossRef]

- 321. Ibrahim, S.; Saleem, B.; Rehman, N.; Zafar, S.A.; Naeem, M.K.; Khan, M.R. CRISPR/Cas9 mediated disruption of Inositol Pentakisphosphate 2-Kinase 1 (TaIPK1) reduces phytic acid and improves iron and zinc accumulation in wheat grains. *J. Adv. Res.* 2021, 37, 33–41. [CrossRef]
- 322. Liu, L.; Gallagher, J.; Arevalo, E.D.; Chen, R.; Skopelitis, T.; Wu, Q.; Bartlett, M.; Jackson, D. Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes. *Nat. Plants* **2021**, *7*, 287–294. [CrossRef]
- 323. Rodríguez-Leal, D.; Lemmon, Z.H.; Man, J.; Bartlett, M.E.; Lippman, Z.B. Engineering quantitative trait variation for crop improvement by genome editing. *Cell* **2017**, *171*, 470–480.e8. [CrossRef]



Article Ferritin Heavy Chain Binds Peroxiredoxin 6 and Inhibits Cell Proliferation and Migration

Maddalena Di Sanzo ^{1,†}, Flora Cozzolino ^{2,3,†}[®], Anna Martina Battaglia ¹, Ilenia Aversa ¹[®], Vittoria Monaco ^{2,3}[®], Alessandro Sacco ¹, Flavia Biamonte ^{1,4}, Camillo Palmieri ¹[®], Francesca Procopio ⁵, Gianluca Santamaria ¹[®], Francesco Ortuso ⁵[®], Piero Pucci ³, Maria Monti ^{2,3}[®] and Maria Concetta Faniello ^{1,*}

- ¹ Research Center of Biochemistry and Advanced Molecular Biology, Department of Experimental and Clinical Medicine, "Magna Graecia" University of Catanzaro, 88100 Catanzaro, Italy
- ² Department of Chemical Sciences, Università degli Studi di Napoli "Federico II", Via Cinthia 21, 80126 Napoli, Italy
- CEINGE Biotecnologie Avanzate, Via G. Salvatore 486, 80145 Napoli, Italy
- ⁴ Interdepartmental Centre of Services, "Magna Graecia" University of Catanzaro, 88100 Catanzaro, Italy
- ⁵ Department of Health Science, "Magna Graecia" University of Catanzaro, 88100 Catanzaro, Italy
- Correspondence: faniello@unicz.it; Tel.: +39-0961-369-4345
- + These authors contributed equally to this work.

Abstract: The H Ferritin subunit (FTH1), as well as regulating the homeostasis of intracellular iron, is involved in complex pathways that might promote or inhibit carcinogenesis. This function may be mediated by its ability to interact with different molecules. To gain insight into the FTH1 interacting molecules, we analyzed its interactome in HEK293T cells. Fifty-one proteins have been identified, and among them, we focused our attention on a member of the peroxiredoxin family (PRDX6), an antioxidant enzyme that plays an important role in cell proliferation and in malignancy development. The FTH1/PRDX6 interaction was further supported by co-immunoprecipitation, in HEK293T and H460 cell lines and by means of computational methods. Next, we demonstrated that FTH1 could inhibit PRDX6-mediated proliferation and migration. Then, the results so far obtained suggested that the interaction between FTH1/PRDX6 in cancer cells might alter cell proliferation and migration, leading to a less invasive phenotype.

Keywords: H Ferritin subunit; PRDX6; protein-protein interaction

1. Introduction

Ferritin, the major intracellular iron-storage protein, binds iron in a soluble, non-toxic form and makes it available for many cellular processes. It is localized in cytoplasm [1], nucleus [2] and mitochondria [3] and is composed of 24 subunits of two different types, a light chain (L; FTL) 19 kDa and a heavy chain (H; FTH1) 21 kDa. The two subunits, which share extensive homology, are functionally distinct: FTH1 is responsible for the ferroxidase activity of the ferritin molecule, while FTL is mainly involved in iron storage [4,5]. FTH1 and FTL are coded by two different genes [6,7] whose activity is regulated at transcriptional [8-10] and post-transcriptional levels. The post-transcriptional control of the ferritin genes is iron-dependent [11] and acts mainly on the translational efficiency of the FTH1 and FTL mRNAs [12]. FTH1 controls the cellular pool of chelatable and redox-active iron, named the labile iron pool (LIP) [13]. It has been shown that FTH1 repression evokes an increase in the cellular LIP level and activity, while its overexpression decreases the LIP content [14]. A high level of LIP results in oxidative damage by catalyzing ROS generation in mitochondria [15]. Besides its role in iron metabolism, FTH1 is involved in signaling pathways related to physiologic and pathologic processes by interacting with different protein partners. FTH1 regulates angiogenesis during inflammation and malignancy by antagonizing the cleavage of HKa (high molecular weight kininogen), an

Citation: Di Sanzo, M.; Cozzolino, F.; Battaglia, A.M.; Aversa, I.; Monaco, V.; Sacco, A.; Biamonte, F.; Palmieri, C.; Procopio, F.; Santamaria, G.; et al. Ferritin Heavy Chain Binds Peroxiredoxin 6 and Inhibits Cell Proliferation and Migration. *Int. J. Mol. Sci.* 2022, 23, 12987. https://doi.org/10.3390/ ijms232112987

Academic Editor: Wajid Zaman

Received: 16 September 2022 Accepted: 23 October 2022 Published: 26 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). endogenous inhibitor of angiogenesis [16], which is also involved in the onset of HIV-mediated neuropathology in patients with a history of drug abuse [16]. FTH1 interacts with the Alacrima-Acalasia-Adrenal Insufficiency Neurological Disorder (ALADIN) protein [17], whose mutations determine the triple A syndrome characterized by impaired FTH1 nuclear uptake [17]. FTH1 may also physically interact with the CXCR4 chemokine receptor, and in response to CXCL12 stimulation, FTH1 is phosphorylated and translocated into nuclei [18]. FTH1/CXCR4 interaction also has functional feedback because CXCR4-mediated ERK1/2 activation is inhibited by FTH1 overexpression [18]. FTH1, under oxidative stress, interacts with p53, thus increasing its transcriptional activity [19]. Moreover, it also binds Nuclear Receptor Coactivator-4 (NCOA4), which targets its degradation to iron recycling [20]. Furthermore, FTH1 may form a complex with death domain-associated nuclear proteins (Daxx), and together they can participate in the apoptosis pathway [21]. FTH1 might function as a tumor suppressor or as an oncogene; indeed, FTH1 knockdown reduces, both in vivo and in vitro, melanoma cell proliferation [22] and may modulate the MHC class I molecules expression leading to NK cells activation [23] while in the erythroleukemia K562 cell line determines an increased expression of a specific set of onco-miRNAs [24], activation of H19/miR-675 axis [25] and severe protein misfolding [26]. In the MCF7 and in NCI-H460 cells induced epithelial-to-mesenchymal transition (EMT) [27], while in ovarian cancer cells, FTH1 is involved in the inhibition of cancer cell proliferation and in cancer stem cell propagation [28]. The evidence that FTH1 may physically interact with molecules expressed in various human malignancies raises the question of the functional role of FTH1 in critical cellular pathways. This function may be mediated by its ability to interact with several signaling molecules. Therefore, studies focused on FTH1 "interactome" analysis are expected to allow an understanding of both physiological and pathological processes involving it. The investigation of new potential FTH1 interacting proteins was performed by using a functional proteomic approach, relying on the immunoprecipitation of a 3xFlag-FTH1 containing complexes from HEK-293T protein extract and their identification by nanoLC-MS/MS methodologies. In this study, we identified Prdx6 as a novel FTH1 interacting protein and that both were involved in the cellular response to oxidative stress. We demonstrated that the FTH1/PRDX6 interaction also occurred in NCI-H460 cells and investigated the functional role of this binding in cell migration and proliferation.

2. Results

2.1. Identification of FTH1 Interacting Proteins

To unravel novel functional roles of ferritin-heavy chain (FTH1), its interactome has been investigated in order to identify new putative protein partners by immunoprecipitation of a tagged 3xFlag-FTH1 and its interacting proteins from HEK-293T protein extracts. HEK-293T cells, transfected with an empty vector, were used as control. Immunoprecipitated proteins were fractionated by SDS-PAGE, and the 39 slices cut from both sample and control lanes (Figure S1) were in situ digested by trypsin. Peptide mixtures were then analyzed by LC-MS/MS, and proteins were identified by Mascot software.

The final list of FTH1's putative partners, obtained by removing common proteins between the sample and control, is reported in Table S1.

Subsequently, a functional analysis of identified FTH1 putative interactors was conducted according to the STRING database, up to a maximum of five co-interactors.

The graphical representation of the inferred network (Figure 1), weighted by protein interactions, provided a detailed view of the putative functional linkage among the 51 proteins used as input and FTH1, facilitating the comprehension of the modularity analysis in biological processes. The analysis generates a putative protein–protein interaction score. We sorted interactions by applying a score threshold of 0.3 and being FTH1 interactors.



Figure 1. String Interaction Analysis: In silico interaction analysis by String (https://string-db.org/ (accessed on 15 September 2022)) suggests putative interaction of FTH1 and PRDX6.

The majority of FTH1 interacting proteins belongs to the cytoskeleton, cell proliferation and traffic, signal transduction, spindle organization and ferroptosis process. Among the proteins already known to interact with FTH1, NCOA4 has been identified, and the interaction has also been confirmed by immunoprecipitation (data not shown). Among the novel putative FTH1 interacting proteins was identified the HSPB1 protein involved in iron metabolism. HSPB1inhibits transferrin receptor 1 (TFR1) expression by modulating intracellular iron accumulation. CDKAL1 belongs to the methyltransferase family, whose function is impaired by cellular iron deficiency. Another interactor is Vimentin, whose function is also modulated by iron metabolism.

Additionally, a novel partner, Peroxiredoxin 6 (PRDX6), captured our attention, being co-expressed with FTH1 in different tissue (see Table S2) and involved in several biological processes, including focal adhesion (FDR 1.35×10^{-5} , cell junction (FDR 0.006), signaling by ROBO receptors (FDR 0.01) and activation of ATR in response to replication stress (FDR 0.02), as predicted by the STRING functional association analysis. (Table S3). PRDX6 is an enzyme involved in the containment of oxidative stress preventing cell peroxidation causative of membrane lesions [29]. Moreover, it has been demonstrated that its catalytic activity favors cancer cell proliferation, motility promoting invasion and metastasis processes [30,31]. In light of these important roles in promoting pathological states, the interaction between FTH1 and PRDX6 was further investigated.

2.2. FTH1/PRDX6 Interaction in HEK-293T and NCI-H460 Cells

To assess PRDX6-FTH1 physical interaction, HEK-293T cells transiently transfected with 3xFlag-FTH1 or 3xFlag control vector were employed for the co-immunoprecipitation experiment. The cell lysates were incubated with the ANTI-FLAG M2 affinity gel, and the

immunoprecipitated complexes were separated by SDS-PAGE and analyzed by Western blot with anti-PRDX6 and anti-Flag antibodies, respectively. As shown in Figure 2A, the interaction between the two proteins was confirmed. The same result was obtained by investigating the interaction between the endogenous proteins expressed at basal levels in wild-type HEK-293T cells (Figure 2B). Confocal microscopy analysis demonstrated the cyto-plasmatic colocalization of FTH1 and PRDX6 (Figure 2C), as expected by their localization. Furthermore, to verify if the interaction between FTH1 and PRDX6 is independent by cell lines, we replicated the same experiment in non-small cell lung cancer (NSCLC) NCI-H460 cells, transiently transfected with 3xFlag-FTH1 or 3xFlag control vector. In NCI-H460 cells, as well as in HEK-293T, FTH1 interacted with PRDX6 (Figure 3A) and colocalized in the cytoplasm (Figure 3B).



Figure 2. Co-IP and colocalization analysis of the interaction between FTH1 and PRDX6. (**A**) HEK-293T cells were transfected with Flag-FTH1 or a control vector (3x-Flag). Cell lysates were immunoprecipitated with anti-Flag M2 resin and then analyzed by immunoblotting with anti-PRDX6 and anti-Flag antibodies. (**B**) HEK-293T cell lysates were immunoprecipitated with FTH1 antibody or anti-mouse IgG, respectively. Eluates were analyzed by immunoblotting with anti-FTH1 or anti-PRDX6 antibodies. (**C**) HEK-293T cells were grown on a coverslip, fixed with 4% paraformaldehyde and processed for double-label immunofluorescence with anti-FTH1 antibody and anti-PRDX6 antibody. Images were collected using confocal microscopy system (40×). Representative data from one of three experiments.

A



B



Figure 3. Co-IP and colocalization analysis of the interaction between FTH1 and PRDX6. (**A**) NCI-H460 cells were transfected with Flag-FTH1 or a control vector (3x-Flag). Cell lysates were immunoprecipitated with anti-Flag M2 resin and then analyzed by immunoblotting with anti-PRDX6 and anti-Flag antibodies. (**B**) NCI-H460 cells were grown on a coverslip, fixed with 4% paraformaldehyde and processed for double-label immunofluorescence with anti-FTH1 antibody and anti-PRDX6 antibody. Images were collected using confocal microscopy system ($40 \times$). Representative data from one of three experiments.

2.3. Molecular Recognition in FTH1/PRDX6 Complex

The interaction between FTH1 and PRDX6 was investigated by means of computational methods. Theoretical models of FTH1 and PRDX6, the latter analyzed both as reduced (PRDX6_r) and sulphinic acid (PRDX6_s) states, were built and submitted to docking simulation (as reported in Materials and Methods). Results clearly indicated a productive recognition of FTH1 against both models of PRDX6 states. Globally, 50 and 55 possible configurations of FTH1 were predicted against PRDX6_r and PRDX6_s targets, respectively. The average interaction energies were very similar and were estimated at -15.44 kcal/mol for FTH1•PRDX6_r and at -15.59 kcal/mol for FTH1•PRDX6_s. Even considering the most stable configuration only, the difference was lower than 1 kcal/mol. All docking-generated complexes were graphically inspected, revealing a wide exploration of FTH1 around both targets (Figure 4).

On the theoretical complexes, the in-house GBPM method was applied to highlight and classify, by quartiles, the most relevant interacting residues (Table 1).



Figure 4. Superimposition of all theoretical complexes between FTH1 and (**a**) PRDX6_r and (**b**) PRDX6_s. Ferritin protein is depicted as wireframe CPK colored, and PRDX6 chains are shown as green and magenta cartoon.

FTH1 Residues Interacting to		PRDX6_r	PRDX6_s		
PRDX6_r	PRDX6_s	Residues Interacting with FTH1			
Thr	5	Gl	y4		
		As	р9		
Gln	14		Asn13		
Argź	22		Glu15		
Tyr3	2 ^a	Arş	<u>3</u> 24		
Tyr3	39	Asp31			
Asp42		Arg53			
	Lys53	Lys	\$56		
Arg	63		Tyr89		
	Glu64		Asn90		
Glue	67	Glu	109		
Lys7	1 ^a	Lys	122		
	Asn74 ^a		Lys125		
Arg	79	Lys141			
Gln8	83	Lys	142		
Asp	84	Asp	158		
Lys8	6 ^a	Glu	172		
Lys87	7 a,b	Arg174			
Asp	89	Asp	180		
Asp9	1 ^a	Lys	182		
	Lys108	Asp	183		
Glu1	16	Gly184			
Asn125			Asp185		
Asn139		Ser	186		
Lys143		Lys200			
	Lys157		Lys204		
		Lys	209		
		Tyr	220		

	Table 1. FTH1	(left) and PRDX6	(right)	most relevant,	quartile 1	, interacting	g residues
--	---------------	-------	-------------	---------	----------------	------------	---------------	------------

^a FTH1 residues interacting to PRDX6 phospholipase site ^b FTH1 residues interacting to PRDX6 peroxidase site.

Interestingly, in both PRDX6 models, FTH1 was able to directly recognize several residues close to His39, Cys47 and Arg132, corresponding to the PRDX6 peroxidase site, and to His26, Ser32 and Asp140 that are involved in the PRDX6 phospholipase activity (Table 1). These interactions could rationalize the experimentally observed FTH1 modulation of PRDX6. To deeply investigate the above-reported event, all theoretical complexes were searched to identify, among the highest complexes stabilizing ones, the FTH1 residues recognizing PRDX6 phospholipase or peroxidase sites. Of course, the phospholipase site, with more solvent exposed than the peroxidase one, allowed larger interaction with FTH1 (Figure 5).



Figure 5. Most stable theoretical complexes of FTH1 with (**a**) PRDX6_r and (**b**) PRDX6_s. Ferritin protein is depicted as wireframe CPK colored, and PRDX6 chains are shown as green and magenta cartoon. PRDX6 residues at peroxidase and phospholipase sites are reported as space-filling yellow carbon or cyan carbon colored, respectively. Ferritin-bound Fe²⁺ ions are shown as orange space-filling.

2.4. FTH1 Inhibits Proliferation and Migration in PRDX6 Overexpressed NCI-H460 Cells

Accumulated evidence suggests that PRDX6 exerts specific functions in cancer progression, affecting cell growth, survival, migration, differentiation, invasiveness, and metastasis. In NSCLC, PRDX6 over-expression promotes cancer cell proliferation and invasive phenotype [30]. To investigate whether the interaction between FTH1 and PRDX6 might affect cell migration and proliferation, NCI-H460 cells were transiently transfected with 3xFlag-PRDX6, 3xFlag-PRDX6/3xFlag-FTH1, 3xFlag-PRDX6/siFTH1 or 3xFlag control vector. The confluent NCI-H460 cells were scraped to create a wound, and cell migration was assessed 24, 48 and 72 h later. The results of a triplicate set of independent assays (A and B of Figure 6) demonstrate that at 72 h after the scratch, the wound area of 3xFlag-PRDX6 cells was significantly narrower than that of the control. Interestingly, when FTH1 and PRDX6 were contemporaneously overexpressed, the wound area was significantly larger than in the presence of PRDX6 alone. Conversely, silencing of FTH1 induced by siRNA transfection and PRDX6 overexpression significantly increased the migratory ability compared to PRDX6 and FTH1 overexpressed cells. Next, we compared the proliferation ability of NCI-H460 FTH1/PRDX6 vs. NCI-H460 PRDX6 and vs. NCI-H460 siFTH1/PRDX6 cells. The results of the MTT test, reported in Figure 6C, showed that, at 72 h, the simultaneous overexpression of FTH1 and PRDX6 was able to significantly reduce cell proliferation with respect to the PRDX6 overexpression, the FTH1 knockdown and PRDX6 overexpression increased proliferation rate compared to FTH1 and PRDX6 overexpression. Western blot analysis (Figure 6D) demonstrated that FTH1 levels are markedly reduced in 3xFlag-PRDX6 cells compared to both control and 3xFlag-FTH1/3xFlag-PRDX6 overexpression. These data strongly indicate that the presence of FTH1 might counteracts cell proliferation and migration processes induced by PRDX6 overexpression.


Figure 6. Scratch wound healing and MTT proliferation assays. (A) After transfection with control vector, (3xFlag) Flag-PRDX6, Flag-PRDX6/Flag-FTH1 and Flag-PRDX6/siFTH1 confluent NCI-H460 cells monolayers were scratched to induce horizontal migration (magnification of $10 \times$) at 0, 24, 48 and 72 h. (B) The wound areas were measured using Image J 1.42q software. The histogram indicates wound area/% of initial area. Value, represent mean \pm SD; n = 3; p-value Flag-PRDX6_vs._3xFlag, Flag-PRDX6_vs._Flag-PRDX6/Flag-FTH1 and Flag-PRDX6_vs_Flag-PRDX6/siFTH1 is considered not statistically significant at 24 h. Flag-PRDX6_vs_3xFlag, Flag-PRDX6_vs_Flag-PRDX6/Flag-FTH1 is considered statistically significant at 48 and 72 h, * p < 0.005; Flag-PRDX6_vs._Flag-PRDX6/siFTH1 is considered not statistically significant at 48 and 72 h. (C) NCI-H460 cells, transfected with control vector, Flag-PRDX6, Flag-PRDX6/Flag-FTH1, and Flag-PRDX6/siFTH1 seeded in 24 wells and were cultured for 0, 24, 48 and 72 h, and their proliferation was determined by absorbance at 490 nm. Final results represent mean \pm SD of three independent experiments, each performed in triplicate. p-value Flag-PRDX6_vs._3xFlag, Flag-PRDX6_vs._Flag-PRDX6/Flag-FTH1 and Flag-PRDX6_vs._Flag-PRDX6/siFTH1 are considered not statistically significant at 24 and 48 h. Flag-PRDX6_vs._3xFlag p < 0.05 and Flag-PRDX6_vs_Flag-PRDX6/Flag-FTH1 * p < 0.005 at 72 h; Flag-PRDX6_vs_Flag-PRDX6/Flag-PRDX6/Flag-FTH1 * p < 0.005 at 72 h; Flag-PRDX6/FFRDX6/FFRD PRDX6_vs._Flag-PRDX6/siFTH1 is considered not statistically significant at 72 h. (D) Western Blot analysis was performed to confirm PRDX6 and PRDX6/FTH1 overexpression and siFTH1 silencing. γ -tubulin was used as a loading control.

3. Discussion

In recent years, our work has mainly focused on the analysis of the iron-independent roles of FTH1 in human-transformed cell lines [22,27,28,32]. Taken all together, the data from our and other groups strongly suggest that the FTH1 is involved in the complex pathways that might promote or inhibit carcinogenesis [33]. Although a lot of experimental evidence suggests that altered expression of FTH1 is a common event in cancer, most of the available studies do not clarify the specific role of FTH1 in this process. This is mainly due to incomplete knowledge of the overall scenario in which FTH1 and its partners interact to promote tumor progression. Therefore, the knowledge of molecules interacting with FTH1 can clarify this issue. Identification and characterization of each component of protein networks is a critical step to fully understand these processes at the molecular level.

The aim of this paper was the identification of FTH1 protein partners to further shed light on its functions in cancer cells. To this purpose, we have pursued a proteomic-based approach relying on the IP-MS (Immunoprecipitation—Mass Spectrometry) of 3xFLAG-FTH1 protein overexpressed in HEK-293T cells.

In order to further investigate the FTH1 role in tumorigenesis, we focused our attention on the PRDX6, a member of the peroxiredoxin family. PRDX6 is a bifunctional enzyme and possesses both peroxidase and calcium-independent phospholipase A2 (iPLA2) activity, protects against oxidative stress and prevents cell peroxidation causative of membrane lesions [29]. Elevated levels of PRDX6 expression have been shown to be associated with a variety of human cancers, among which lung and breast [30,34]. In lung cancer cells, it has been demonstrated that the upregulation of PRDX6 results in the activation of Akt via phosphoinositide 3-kinase (PI3K) and p38 kinase [35]. It has been demonstrated that PRDX6 promotes cell proliferation and inhibits cancer cell apoptosis; its peroxidase activity promotes the growth of cancer cells, whereas its PLA2 activity promotes the invasion and metastasis of cancer cells [30]. Their physical interaction was also confirmed by coimmunoprecipitation in both FTH1 overexpressing cells and in basal conditions. Confocal images further confirmed the interaction between FTH1 and PRDX6 and showed their cytoplasmic predominantly colocalization. Our data, in agreement with the literature, showed that PRDX6 overexpression induced NCI-H460 cell proliferation and migration; on the contrary, NCI-H460 cells show a strong reduction of their proliferative and migratory capability when simultaneous transfected with FTH1 and PRDX6. These results demonstrate that the interaction between FTH1/PRDX6 alters cell proliferation and migration leading to a less invasive phenotype, suggesting for FTH1 an inhibitory effect on PRDX6 oncogenic role.

It has been demonstrated that PRDX6 is highly expressed in human cancer cells and plays a critical role in cancer development [36,37]. Indeed, high levels of PRDX6 contribute to the proliferation of cancer cells [30]. So, our findings demonstrate the inhibitory effects of FTH1 on PRDX6 and suggest a new potential therapeutic target for the development of novel therapeutic strategies for the treatments of such cancers overexpressing PRDX6.

4. Materials and Methods

4.1. Cell Cultures

HEK-293T cells (Sigma-Aldrich, St. Louis, MO, USA) human embryonic kidneys were cultured in adherent conditions in a DMEM (Sigma-Aldrich) medium with 10% FBS and 1% Penicillin-Streptomycin (Sigma-Aldrich). NCI-H460 human non-small lung cancer cells (ATCC, Manassas, VA, USA) were cultured in an RPMI 1640 (Sigma-Aldrich) medium supplemented with 10% fetal bovine serum (FBS, Belize City, Belize) and 1% Penicillin-Streptomycin (Sigma-Aldrich). The two cell lines were maintained at 37 °C in a humidified 5% CO₂ atmosphere.

4.2. Transient Transfection and Cell Lysis

HEK-293T cells were transiently transfected with an expression vector containing the full length of human FTH1 cDNA (3xFlag-FTH1) (N-Flag tag-Sino Biological, Beijing, China), using Lipofectamine 3000 transfection reagent (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions [38]. Cells transfected with empty vector 3xFlag were used as a negative control. Cells were incubated in the medium containing the transfection mixture for up to 72 h. Experiments were performed at least three times. Cell lysis was conducted using a buffer composed of the following components: 50 mM Tris/HCl pH 7.4, 150 mM NaCl, 10% Glycerol, 1 mM Na₃(VO)₄, 1 mM NaF, 0, 5 mM PMSF, 1% Triton and 1 mini pill of protease inhibitor (EDTA free ROCHE). Cell pellets were resuspended, left for 10 min in ice, and then put on the tube rotator at 4 °C for 45 min. Finally, the lysates were centrifuged for 30 min at 13,000 rpm and 4 °C. A quantitative evaluation of protein extracts was carried out with Bradford assay (BIORAD, Hercules, CA, USA). Bovine serum albumin was used as the standard protein for the calibration curve.

4.3. Isolation of Protein Complexes by Immunoprecipitation

Lysates from 3xFlag-FTH1 transfected cells and 3xFlag negative control (cells transfected with empty vector) were subjected simultaneously pre-cleared onto DynabeadsTM Protein G (Invitrogen, Waltham, MA, USA); thus, to remove the background of unspecific proteins from cellular extracts. The latter were then subjected to immunoprecipitation protocol by using anti-FLAG M2 (Sigma Aldrich) magnetic beads to isolate the bait, and its putative interactors, as previously reported [39]. Surnatants containing the unbound proteins were removed, and the beads were washed with two different NaCl concentrations (150 and 300 mM) in lysis buffer. Elution was performed using the free 3xFLAG peptide with a concentration of 200 μ g/mL for 5 h at 4 °C. Protein samples were precipitated in a mixture of methanol and chloroform and then dried by a Speed-Vac system (Thermo Fisher, Waltham, MA, USA).

4.4. Mass Spectrometry Analysis and Protein Identification

Immunoprecipitated protein complexes were fractionated by SDS-PAGE on a 16×16 cm, 8–15% gradient acrylamide/bis-acrylamide gel. The latter was stained with Colloidal Blue Coomassie (PIERCE, Hercules, CA, USA), and the excess of the dye was removed by washing with deionized water. Thirty-nine bands were excised from both sample and control lanes and subjected to an in-situ hydrolysis protocol, as previously reported [40]. Each peptide mixture was dried and then suspended again in 10 µL of 0.2% HCOOH (J.T. Baker, Waltham, MA, USA) and analyzed by LC-MS/MS on an LTQ Orbitrap XL system (Thermo Fisher) equipped with a nano-LC Proxeon nanoEasy-II system. Peptides were fractionated onto a C18 reverse-phase capillary column (5 µm biosphere, 75 µm ID, 200 mm length) working at 250 nL/min flow rate and adopting a linear gradient from 10% to 60% of eluent B (0.2% formic acid, 95% acetonitrile LC-MS Grade) over 69 min. Mass spectrometric analyses were carried out in Data Dependent Acquisition mode (DDA): from each MS scan, spanning from 300 to 1800 m/z, the five most abundant ions were selected and fragmented.

Output data were processed generating. mgf files employed for protein identification procedure in NCBI database according to Mascot licensed software (Matrix Science, Boston, MA, USA). Protein identification was carried out by using 10 ppm as peptides mass tolerance for MS and 0.6 Da for MS/MS search; *Homo sapiens* as taxonomy, carbamidomethyl (C) as fixed modification and Gln \rightarrow pyro-Glu (N-term Q), Oxidation (M), Pyro-carbamidomethyl (N-term C) as variable modifications. The proteins were identified with at least two significant peptides, overcoming the Mascot score threshold.

4.5. Co-Immunoprecipitation

HEK-293T cells were lysated in a Ripa buffer assay (Radioimmunoprecipitation assay) [41–43] and incubated overnight with 1 μ g of anti-FTH1 antibody or with non-immune serum at 4 °C and then incubated for 1h at 4 °C with Protein A/G plus-agarose. The beads were collected by centrifugation, washed four times with lysis buffer and then loaded onto an SDS 12 % (w/v) polyacrylamide gel [44].

4.6. Western Blotting Analysis

A total of 60 µg of protein extract was resolved on 15% SDS-PAGE and then transferred to a nitrocellulose membrane by electroblotting [32]. Non-specific reactivity was blocked in nonfat dry milk in TTBS 1X [5% (w/v) milk in TBS 1X (pH 7.4) and 0.1% Tween [20] for 2 h at room temperature. The membrane was incubated with specific primary antibodies anti-FTH1 (sc-376594 Santa Cruz, Santa Cruz, CA, USA), Anti-Flag M2 (F-1804 Sigma Aldrich) and anti-PRDX6 (ab59543 Abcam, Cambridge, UK) overnight at 4 °C. After incubation, and the membranes were washed three times with TTBS 1X for 10 min and incubated with an appropriate horseradish peroxidase-conjugated secondary antibody at room temperature for 1 h. According to the manufacturer's instructions. The membranes were washed three times with TTBS 1X, and signals were visualized by ECL-Western blot detection reagents (Santa Cruz Biotechnology, Dallas, TX, USA [45]).

4.7. Molecular Modeling

Protein Data Bank (PDB) [46] entries 3AJO [47], 5B6M and 5B6N [48] were selected for building our theoretical models of FTH1 and PRDX6 in reduced (PRDX6_r) and sulphinic

acid (PRDX6_s) states, respectively. The original PDB 3AJO structure was modified by removing co-crystallized water molecules and Mg²⁺ ions. To identify the most favorable FTH1 sites for iron recognition, FE + 2 probe molecular interaction field (MIF), as implemented in GRID ver. 22d [49], was computed. To obtain a more accurate position of the iron ions, the NPLA GRID keyword was equal to 3, while other parameters were set to the default value. MIF was filtered by means of the GRID MINIM utility using an energy cutoff equal to 10 kcal/mol above the global minimum energy point (-85.27 kcal/mol). Such an approach allowed us to identify two favorable positions for the iron ion with respect to 3AJO PDB entry. Therefore, two iron ions were included in our FTH1 preliminary model.

Both PRDX6_r and PRDX6_s PDB entries contained, reported as chains A, B and C, three conformationally different structures of the protein. These models were divided by the chain, and each of them was considered a conformer of PRDX6.

To prepare FTH1 and PRDX6 structures for further simulation, these were optimized using MacroModel ver. 11.9 (Schrödinger Release 2018-1: MacroModel, Schrödinger, LLC, New York, NY, USA, 2021). In detail, hydrogen atoms were added, and 10,000 steps of the Polack Ribiere Conjugate Gradient energy minimization method, coupled to the AMBER* force field, were applied. Aqueous environment effects were mimicked by means of the implicit solvation model GB/SA water.

The aim was to investigate the recognition between FHT1 and PRDX6. The resulting optimized structures were submitted to AutoDock Vina v. 1.2.1 molecular docking [50]. Using the MGL-tool ver. 1.5.6, Kollman charge distribution was computed on all protein models. PRDX6_r and _s optimized conformations were considered as receptor models and FHT1 as ligand. Therefore, for each receptor model, an AutoDock Vina simulation was performed, considering a regular box equal to 4,000,752 Å³ entirely surrounding targets. For each FHT1·PRDX6 theoretical complex, a maximum of 20 configurations were allowed (num_modes = 30), and the other docking parameters were set to default values. Docking results highlighted 57 and 56 FHT1 poses with respect to PRDX6_r and PRDX6_s, respectively. The affinity of FHT1 against both PRDX6_r and PRDX6_s was estimated by computing the AutoDock Vina average interaction energy values.

To consider the induced fit phenomena, all theoretical complexes were submitted to the same energy minimization protocol previously reported for the protein models preparation. The most relevant interacting residues were highlighted by analyzing DRY, N1 and O GRID MIFs using a modified GBPM method against all optimized structures [51,52].

4.8. Immunofluorescence Assay

For immunofluorescence analysis, HEK-293T and NCI-H460 cells were treated as previously described by Aversa et al. [27]. Briefly, HEK-293T and NCI-H460 cells were incubated with primary antibodies anti-FTH1 (sc-376594 Santa Cruz) and anti-PRDX6 (ab59543 Abcam) overnight at 4 °C. Appropriate secondary antibodies (anti-mouse IgG Alexa Fluor 488 and anti-rabbit IgG Alexa Fluor 555, Thermo Fisher Scientific) diluted in PBS 1X were applied for 1 h at room temperature. After three washes, Nuclear Dapi (1:1000, Invitrogen, Carlsbad, CA, USA) was added for 20 min. The slides were mounted on microscope slides using a mounting solution ProLong Gold antifade reagent (Thermo Fisher Scientific). Images were collected using a Leica TCS SP2 confocal microscopy system (Leica Microsystems, Wetzlar, Germany) [53].

4.9. Proliferation Assays

For the proliferation assay 3×10^3 3xFlag control vector, 3xFlag-PRDX6, 3xFlag FTH1/PRDX6 and 3xFlag-PRDX6/siFTH1 NCI-H460 cells were plated in a 96-well flat bottom tissue culture plate. At 24, 48 and 72 h of culture, 10 µL of 3-(4,5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT) (Sigma-Aldrich) solution (2 mg/mL) were added per well. After 4 h of incubation, the culture medium was discarded and replaced with 200 µL of isopropanol. Optical density (OD) was read on a multi-well scanning

spectrophotometer (ELISA reader) (BIORAD) at 595 nm. The proliferation assay was performed in triplicate.

4.10. Wound Healing Assay

NCI-H460 cells 3xFlag control vector, 3xFlag-PRDX6, 3xFlag FTH1/PRDX6 and 3xFlag-PRDX6/siFTH1 were seeded in 60 mm dishes at a density of 4×10^5 . After 24 h, a yellow pipette tip was used to make a scratch. Scratch closure was monitored, and images were captured at 0, 24, 48 and 72 h using a light microscope (using the Leica DFC420 C and Leica Application Suite X Software 3.7.4.23463) (Leica Microsystems). Wound closure was measured by calculating the density of the pixels in the area where the cut was made and expressed as a percentage of wound closure in the area. The percentage of wound closure was calculated by ImageJ64 software.

4.11. Statistical Analyses

Data were presented as the mean \pm standard error (SD) of three independent experiments. Statistical and data analysis was carried out using GraphPad Prism 9 software. Statistical differences between samples were assessed by Student's *t*-test.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms232112987/s1.

Author Contributions: M.C.F., M.M. and P.P. writing—original draft, writing—review and editing, validation, visualization, project administration, and supervision. M.D.S., F.C. and V.M. methodology, formal analysis, investigation, validation, and visualization. G.S., F.P. and F.O. software. C.P., A.M.B., I.A., A.S. and F.B. formal analysis, and data curation. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article or Supplementary Materials.

Acknowledgments: We thank Caterina Alessi for providing technical support. We thank the Centre of Interdepartmental Services (CIS), 'Magna Graecia' University of Catanzaro, Italy, for supporting part of this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Theil, E.C. Ferritin: Structure, Gene Regulation, and Cellular Function in Animals, Plants, and Microorganisms. *Annu. Rev. Biochem.* **1987**, *56*, 289–315. [CrossRef]
- Surguladze, N.; Patton, S.; Cozzi, A.; Fried, M.G.; Connor, J.R. Characterization of Nuclear Ferritin and Mechanism of Translocation. *Biochem. J.* 2005, 388, 731–740. [CrossRef] [PubMed]
- 3. Levi, S.; Corsi, B.; Bosisio, M.; Invernizzi, R.; Volz, A.; Sanford, D.; Arosio, P.; Drysdale, J. A Human Mitochondrial Ferritin Encoded by an Intronless Gene. *J. Biol. Chem.* **2001**, *276*, 24437–24440. [CrossRef] [PubMed]
- 4. Arosio, P.; Levi, S. Ferritin, Iron Homeostasis, and Oxidative Damage. Free Radic. Biol. Med. 2002, 33, 457–463. [CrossRef]
- 5. Theil, E.C. Ferritin: At the Crossroads of Iron and Oxygen Metabolism. J. Nutr. 2003, 133, 1549S–1553S. [CrossRef] [PubMed]
- 6. Costanzo, F.; Colombo, M.; Staempfli, S.; Santoro, C.; Marone, M.; Frank, R.; Delius, H.; Cortese, R. Structure of Gene and Pseudogenes of Human Apoferritin H. *Nucleic Acids Res.* **1986**, *14*, 721. [CrossRef]
- Santoro, C.; Marone, M.; Ferrone, M.; Costanzo, F.; Colombo, M.; Minganti, C.; Cortese, R.; Silengo, L. Cloning of the Gene Coding for Human L Apoferritin. *Nucleic Acids Res.* 1986, 14, 2863. [CrossRef]
- 8. Bevilacqua, M.A.; Faniello, M.C.; Russo, T.; Cimino, F.; Costanzo, F. Transcriptional Regulation of the Human H Ferritin-Encoding Gene (FERH) in G418-Treated Cells: Role of the B-Box-Binding Factor. *Gene* **1994**, *141*, 287–291. [CrossRef]
- 9. Miller, L.L.; Miller, S.C.; Torti, S.v.; Tsuji, Y.; Torti, F.M. Iron-Independent Induction of Ferritin H Chain by Tumor Necrosis Factor. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4946–4950. [CrossRef]
- Costanzo, F.; Bevilacqua, M.A.; Giordano, M.; Cimino, F. Expression of Genes of Ferritin Subunits in Human Hepatoma Cell Lines. *Biochem. Biophys. Res. Commun.* 1989, 161, 902–909. [CrossRef]

- 11. Munro, H.N. Iron Regulation of Ferritin Gene Expression. J. Cell. Biochem. 1990, 44, 107–115. [CrossRef] [PubMed]
- 12. Klausner, R.D.; Rouault, T.A.; Harford, J.B. Regulating the Fate of MRNA: The Control of Cellular Iron Metabolism. *Cell* **1993**, *72*, 19–28. [CrossRef]
- 13. Kakhlon, O.; Cabantchik, Z.I. The Labile Iron Pool: Characterization, Measurement, and Participation in Cellular Processes(1). *Free Radic. Biol. Med.* **2002**, *33*, 1037–1046. [CrossRef]
- 14. Tian, Y.; Lu, J.; Hao, X.; Li, H.; Zhang, G.; Liu, X.; Li, X.; Zhao, C.; Kuang, W.; Chen, D.; et al. FTH1 Inhibits Ferroptosis Through Ferritinophagy in the 6-OHDA Model of Parkinson's Disease. *Neurotherapeutics* **2020**, *17*, 1796–1812. [CrossRef] [PubMed]
- Zhang, K.H.; Tian, H.Y.; Gao, X.; Lei, W.W.; Hu, Y.; Wang, D.M.; Pan, X.C.; Yu, M.L.; Xu, G.J.; Zhao, F.K.; et al. Ferritin Heavy Chain-Mediated Iron Homeostasis and Subsequent Increased Reactive Oxygen Species Production Are Essential for Epithelial-Mesenchymal Transition. *Cancer Res.* 2009, *69*, 5340–5348. [CrossRef] [PubMed]
- 16. Coffman, L.G.; Parsonage, D.; D'Agostino, R.; Torti, F.M.; Torti, S.v. Regulatory Effects of Ferritin on Angiogenesis. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 570–575. [CrossRef] [PubMed]
- Storr, H.L.; Kind, B.; Parfitt, D.A.; Chapple, J.P.; Lorenz, M.; Koehler, K.; Huebner, A.; Clark, A.J.L. Deficiency of Ferritin Heavy-Chain Nuclear Import in Triple a Syndrome Implies Nuclear Oxidative Damage as the Primary Disease Mechanism. *Mol. Endocrinol.* 2009, 23, 2086–2094. [CrossRef]
- Li, R.; Luo, C.; Mines, M.; Zhang, J.; Fan, G.H. Chemokine CXCL12 Induces Binding of Ferritin Heavy Chain to the Chemokine Receptor CXCR4, Alters CXCR4 Signaling, and Induces Phosphorylation and Nuclear Translocation of Ferritin Heavy Chain. J. Biol. Chem. 2006, 281, 37616–37627. [CrossRef]
- Lee, J.H.; Jang, H.; Cho, E.J.; Youn, H.D. Ferritin Binds and Activates P53 under Oxidative Stress. *Biochem. Biophys. Res. Commun.* 2009, 389, 399–404. [CrossRef]
- 20. Bellelli, R.; Federico, G.; Matte', A.; Colecchia, D.; Iolascon, A.; Chiariello, M.; Santoro, M.; de Franceschi, L.; Carlomagno, F. NCOA4 Deficiency Impairs Systemic Iron Homeostasis. *Cell Rep.* **2016**, *14*, 411–421. [CrossRef]
- Liu, F.; Du, Z.Y.; He, J.L.; Liu, X.Q.; Yu, Q.B.; Wang, Y.X. FTH1 Binds to Daxx and Inhibits Daxx-Mediated Cell Apoptosis. *Mol. Biol. Rep.* 2012, 39, 873–879. [CrossRef] [PubMed]
- di Sanzo, M.; Gaspari, M.; Misaggi, R.; Romeo, F.; Falbo, L.; de Marco, C.; Agosti, V.; Quaresima, B.; Barni, T.; Viglietto, G.; et al. H Ferritin Gene Silencing in a Human Metastatic Melanoma Cell Line: A Proteomic Analysis. J. Proteome Res. 2011, 10, 5444–5453. [CrossRef] [PubMed]
- Sottile, R.; Federico, G.; Garofalo, C.; Tallerico, R.; Faniello, M.C.; Quaresima, B.; Cristiani, C.M.; di Sanzo, M.; Cuda, G.; Ventura, V.; et al. Iron and Ferritin Modulate MHC Class I Expression and NK Cell Recognition. *Front. Immunol.* 2019, 10, 224. [CrossRef] [PubMed]
- Biamonte, F.; Zolea, F.; Bisognin, A.; di Sanzo, M.; Saccoman, C.; Scumaci, D.; Aversa, I.; Panebianco, M.; Faniello, M.C.; Bortoluzzi, S.; et al. H-Ferritin-Regulated MicroRNAs Modulate Gene Expression in K562 Cells. *PLoS ONE* 2015, 10, e0122105. [CrossRef]
- di Sanzo, M.; Chirillo, R.; Aversa, I.; Biamonte, F.; Santamaria, G.; Giovannone, E.D.; Faniello, M.C.; Cuda, G.; Costanzo, F. ShRNA Targeting of Ferritin Heavy Chain Activates H19/MiR-675 Axis in K562 Cells. *Gene* 2018, 657, 92–99. [CrossRef]
- Zolea, F.; Biamonte, F.; Candeloro, P.; di Sanzo, M.; Cozzi, A.; di Vito, A.; Quaresima, B.; Lobello, N.; Trecroci, F.; di Fabrizio, E.; et al. H Ferritin Silencing Induces Protein Misfolding in K562 Cells: A Raman Analysis. *Free Radic. Biol. Med.* 2015, *89*, 614–623. [CrossRef]
- Aversa, I.; Zolea, F.; Ieranò, C.; Bulotta, S.; Trotta, A.M.; Faniello, M.C.; de Marco, C.; Malanga, D.; Biamonte, F.; Viglietto, G.; et al. Epithelial-to-Mesenchymal Transition in FHC-Silenced Cells: The Role of CXCR4/CXCL12 Axis. *J. Exp. Clin. Cancer Res.* 2017, *36*, 104. [CrossRef]
- Lobello, N.; Biamonte, F.; Pisanu, M.E.; Faniello, M.C.; Jakopin, Ž.; Chiarella, E.; Giovannone, E.D.; Mancini, R.; Ciliberto, G.; Cuda, G.; et al. Ferritin Heavy Chain Is a Negative Regulator of Ovarian Cancer Stem Cell Expansion and Epithelial to Mesenchymal Transition. *Oncotarget* 2016, 7, 62019–62033. [CrossRef]
- 29. Fisher, A.B. Peroxiredoxin 6: A Bifunctional Enzyme with Glutathione Peroxidase and Phospholipase A₂ Activities. *Antioxid. Redox Signal.* **2011**, *15*, 831–844. [CrossRef]
- 30. Li, H.; Zhang, D.; Li, B.; Zhen, H.; Chen, W.; Men, Q. PRDX6 Overexpression Promotes Proliferation, Invasion, and Migration of A549 Cells in Vitro and in Vivo. *Cancer Manag. Res.* 2021, *13*, 1245–1255. [CrossRef]
- Huang, W.S.; Huang, C.Y.; Hsieh, M.C.; Kuo, Y.H.; Tung, S.Y.; Shen, C.H.; Hsieh, Y.Y.; Teng, C.C.; Lee, K.C.; Lee, K.F.; et al. Expression of PRDX6 Correlates with Migration and Invasiveness of Colorectal Cancer Cells. *Cell. Physiol. Biochem.* 2018, 51, 2616–2630. [CrossRef] [PubMed]
- Misaggi, R.; di Sanzo, M.; Cosentino, C.; Bond, H.M.; Scumaci, D.; Romeo, F.; Stellato, C.; Giurato, G.; Weisz, A.; Quaresima, B.; et al. Identification of H Ferritin-Dependent and Independent Genes in K562 Differentiating Cells by Targeted Gene Silencing and Expression Profiling. *Gene* 2014, 535, 327–335. [CrossRef] [PubMed]
- Aung, W.; Hasegawa, S.; Furukawa, T.; Saga, T. Potential Role of Ferritin Heavy Chain in Oxidative Stress and Apoptosis in Human Mesothelial and Mesothelioma Cells: Implications for Asbestos-Induced Oncogenesis. *Carcinogenesis* 2007, 28, 2047–2052. [CrossRef]
- 34. Chang, X.Z.; Li, D.Q.; Hou, Y.F.; Wu, J.; Lu, J.S.; Di, G.H.; Jin, W.; Ou, Z.L.; Shen, Z.Z.; Shao, Z.M. Identification of the Functional Role of Peroxiredoxin 6 in the Progression of Breast Cancer. *Breast. Cancer Res.* **2007**, *9*, R76. [CrossRef]

- Lee, S.B.; Ho, J.N.; Yoon, S.H.; Kang, G.Y.; Hwang, S.G.; Um, H.D. Peroxiredoxin 6 Promotes Lung Cancer Cell Invasion by Inducing Urokinase-Type Plasminogen Activator via P38 Kinase, Phosphoinositide 3-Kinase, and Akt. *Mol. Cells* 2009, 28, 583–588. [CrossRef]
- Hu, X.; Lu, E.; Pan, C.; Xu, Y.; Zhu, X. Overexpression and Biological Function of PRDX6 in Human Cervical Cancer. J. Cancer 2020, 11, 2390–2400. [CrossRef]
- López-Grueso, M.J.; Lagal, D.J.; García-Jiménez, Á.F.; Tarradas, R.M.; Carmona-Hidalgo, B.; Peinado, J.; Requejo-Aguilar, R.; Bárcena, J.A.; Padilla, C.A. Knockout of PRDX6 Induces Mitochondrial Dysfunction and Cell Cycle Arrest at G2/M in HepG2 Hepatocarcinoma Cells. *Redox Biol.* 2020, *37*, 101737. [CrossRef]
- 38. Zolea, F.; Biamonte, F.; Battaglia, A.M.; Faniello, M.C.; Cuda, G.; Costanzo, F. Caffeine Positively Modulates Ferritin Heavy Chain Expression in H460 Cells: Effects on Cell Proliferation. *PLoS ONE* **2016**, *11*, e0163078. [CrossRef] [PubMed]
- Venuto, S.; Monteonofrio, L.; Cozzolino, F.; Monti, M.; Appolloni, I.; Mazza, T.; Canetti, D.; Giambra, V.; Panelli, P.; Fusco, C.; et al. TRIM8 Interacts with KIF11 and KIFC1 and Controls Bipolar Spindle Formation and Chromosomal Stability. *Cancer Lett.* 2020, 473, 98–106. [CrossRef]
- Scieuzo, C.; Salvia, R.; Franco, A.; Pezzi, M.; Cozzolino, F.; Chicca, M.; Scapoli, C.; Vogel, H.; Monti, M.; Ferracini, C.; et al. An Integrated Transcriptomic and Proteomic Approach to Identify the Main Torymus Sinensis Venom Components. *Sci. Rep.* 2021, 11, 5032. [CrossRef]
- Battaglia, A.M.; Sacco, A.; Perrotta, I.D.; Faniello, M.C.; Scalise, M.; Torella, D.; Levi, S.; Costanzo, F.; Biamonte, F. Iron Administration Overcomes Resistance to Erastin-Mediated Ferroptosis in Ovarian Cancer Cells. *Front. Oncol.* 2022, *12*, 868351. [CrossRef] [PubMed]
- Vecchio, E.; Nisticò, N.; Golino, G.; Iaccino, E.; Maisano, D.; Mimmi, S.; Aloisio, A.; Renna, M.; Avagliano, A.; Arcucci, A.; et al. IBtkα Activates the β-Catenin-Dependent Transcription of MYC through Ubiquitylation and Proteasomal Degradation of GSK3β in Cancerous B Cells. *Int. J. Mol. Sci.* 2022, 23, 2044. [CrossRef] [PubMed]
- 43. Biamonte, F.; Zolea, F.; Santamaria, G.; Battaglia, A.M.; Cuda, G.; Costanzo, F. Human Haematological and Epithelial Tumor-Derived Cell Lines Express Distinct Patterns of Onco-MicroRNAs. *Cell. Mol. Biol.* **2017**, *63*, 75–85. [CrossRef] [PubMed]
- 44. di Caprio, R.; Ciano, M.; Montano, G.; Costanzo, P.; Cesaro, E. KAP1 Is a Novel Substrate for the Arginine Methyltransferase PRMT5. *Biology* **2015**, *4*, 41. [CrossRef]
- 45. Sodaro, G.; Cesaro, E.; Montano, G.; Blasio, G.; Fiorentino, F.; Romano, S.; Jacquel, A.; Aurberger, P.; Costanzo, P. Role of ZNF224 in C-Myc Repression and Imatinib Responsiveness in Chronic Myeloid Leukemia. *Oncotarget* **2017**, *9*, 3417–3431. [CrossRef]
- 46. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]
- 47. Masuda, T.; Goto, F.; Yoshihara, T.; Mikami, B. The Universal Mechanism for Iron Translocation to the Ferroxidase Site in Ferritin, Which Is Mediated by the Well Conserved Transit Site. *Biochem. Biophys. Res. Commun.* **2010**, 400, 94–99. [CrossRef]
- 48. Kim, K.H.; Lee, W.; Kim, E.E.K. Crystal Structures of Human Peroxiredoxin 6 in Different Oxidation States. *Biochem. Biophys. Res. Commun.* 2016, 477, 717–722. [CrossRef]
- Goodford, P.J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. J. Med. Chem. 1985, 28, 849–857. [CrossRef]
- 50. Trott, O.; Olson, A.J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455. [CrossRef]
- Ortuso, F.; Langer, T.; Alcaro, S. GBPM: GRID-Based Pharmacophore Model: Concept and Application Studies to Protein-Protein Recognition. *Bioinformatics* 2006, 22, 1449–1455. [CrossRef] [PubMed]
- 52. Ortuso, F.; Mercatelli, D.; Guzzi, P.H.; Giorgi, F.M. Structural Genetics of Circulating Variants Affecting the SARS-CoV-2 Spike/Human ACE2 Complex. J. Biomol. Struct. Dyn. 2022, 40, 6545–6555. [CrossRef] [PubMed]
- Cesaro, E.; Pastore, A.; Polverino, A.; Manna, L.; Divisato, G.; Quintavalle, C.; di Sanzo, M.; Faniello, M.C.; Grosso, M.; Costanzo, P. ZNF224 Is a Mediator of TGF-β pro-Oncogenic Function in Melanoma. *Hum. Mol. Genet.* 2021, 30, 2100–2109. [CrossRef] [PubMed]



Article 3β-Corner Stability by Comparative Molecular Dynamics Simulations

Vladimir R. Rudnev^{1,2}, Kirill S. Nikolsky¹, Denis V. Petrovsky¹, Liudmila I. Kulikova^{1,2}, Anton M. Kargatov³, Kristina A. Malsagova^{1,*}, Alexander A. Stepanov¹, Arthur T. Kopylov¹, Anna L. Kaysheva¹, and Alexander V. Efimov³

- ¹ Biobanking Group, Branch of Institute of Biomedical Chemistry "Scientific and Education Center", 109028 Moscow, Russia
- ² Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, 142290 Pushchino, Russia
- ³ Institute of Protein Research, Russian Academy of Sciences, 142290 Pushchino, Russia
 - * Correspondence: kristina.malsagova86@gmail.com; Tel.: +7-(499)-764-98-78

Abstract: This study explored the mechanisms by which the stability of super-secondary structures of the 3 β -corner type autonomously outside the protein globule are maintained in an aqueous environment. A molecular dynamic (MD) study determined the behavioral diversity of a large set of non-homologous 3 β -corner structures of various origins. We focused on geometric parameters such as change in gyration radius, solvent-accessible area, major conformer lifetime and torsion angles, and the number of hydrogen bonds. Ultimately, a set of 3 β -corners from 330 structures was characterized by a root mean square deviation (RMSD) of less than 5 Å, a change in the gyration radius of no more than 5%, and the preservation of amino acid residues positioned within the allowed regions on the Ramachandran map. The studied structures retained their topologies throughout the MD experiments. Thus, the 3 β -corner structure was found to be rather stable per se in a water environment, i.e., without the rest of a protein molecule, and can act as the nucleus or "ready-made" building block in protein folding. The 3 β -corner can also be considered as an independent object for study in field of structural biology.

Keywords: super-secondary structure; 3β-corner; folding nuclei; structure stability

1. Introduction

Structural motifs (super-secondary structures, SSS) of globular proteins are defined as commonly occurring folding units composed of two or more elements of secondary structure that are adjacent along the polypeptide chain and are in close contact in threedimensional space. While many different structural motifs have been observed to recur within proteins, only some of motifs exhibit the definite handedness and a unique overall fold, irrespective of whether they occur in homologous or non-homologous proteins [1–4]. The high incidence of occurrence of structural motifs in unrelated proteins and the fact that many small proteins are merely composed of the motifs indicate their stability and ability to fold into unique structures per se. Such structural motifs are of particular interest since they can act as nuclei, or "ready-made" building blocks, in protein folding, or can be utilized as starting structures in protein modeling [4,5]. The structural motif with a unique overall fold occurring over all proteins of the structural group can be taken as the starting structure in modeling or as the root structure of structural tree. Larger protein structures are obtained by stepwise addition of α -helices and/or β -strands to the root motif in accordance with a restricted set of rules inferred from known principles of protein structures. Several structural trees for protein superfamilies, including several thousand of known proteins (their 3D structures were taken from PDB), have been constructed and are available at the following source: http://strees.protres.ru/ (accessed on 15 August 2022).

Citation: Rudnev, V.R.; Nikolsky, K.S.; Petrovsky, D.V.; Kulikova, L.I.; Kargatov, A.M.; Malsagova, K.A.; Stepanov, A.A.; Kopylov, A.T.; Kaysheva, A.L.; Efimov, A.V. 3β-Corner Stability by Comparative Molecular Dynamics Simulations. *Int. J. Mol. Sci.* **2022**, *23*, 11674. https:// doi.org/10.3390/ijms231911674

Academic Editor: Wajid Zaman

Received: 25 August 2022 Accepted: 27 September 2022 Published: 2 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). There is growing interest in structural motifs because they can be embryos in the process of protein folding. Over the past 30 years, significant research efforts have been made to identify and characterize the folding nuclei for the known protein structures [6,7]. In this scope, structural motifs were broadly discussed as a promising candidate [7–9]. The prominent representatives of structural motives are α - α -corner, β -hairpin, and Greek key motif [10]. Specification of these structures caused researchers to develop a catalog of autonomously folded protein motifs and archetypes.

This study is devoted to comparative analysis of variety molecular dynamics simulations of the stability of 3β -corner structural motif (the super-secondary structure).

The 3 β -corner can be represented as a triple-stranded β -sheet folded on to itself so that two $\beta\beta$ -hairpins are packed roughly orthogonally in different layers and the central strand bends by $\approx 90^{\circ}$ in right-handed direction when passing from one β -layer to the other [11] (Figure 1). All the 3 β -corners observed in proteins can be considered as Z-like β -sheets when viewed from their concave surfaces.



Figure 1. A schematic representation of the 3β -corner (**a**), ribbon and wireframe model of the 3β -corner, (**b**,**c**) beta-adrenergic receptor kinase 1 (bos taurus), PDB ID 1OMW region: A586-A625. Putative hydrophobic contacts are highlighted by dashed red lines, and hydrogen bonds by dashed blue lines.

The molecular dynamics experiment provides a range of tools for studying the properties of SSSs and their stability, and, consequently, it is crucial for the selection of candidate protein structures as folding nuclei [12].

2. Results

2.1. The 3β-Corner Motif as an Autonomous Structure

In the first stage, we examined the possibility of studying the SSS type 3β -corner as autonomously stable in an aqueous medium, that is, outside the protein globule. In this study, autonomous stability refers to the preservation of the structural topology of the 3β -corner in the MD experiment for the whole protein and separately from the protein environment in an aqueous medium. To do this, we removed parts of the structures corresponding to the 3β -corners from the standard files in the ".*pdb" format extracted from the PDB bank.

We carried out a comparative analysis of changes in the geometric characteristics of the 3β -corners in the composition of whole proteins and autonomously in an aqueous medium on the basis of the results of a 300 ns MD experiment. We performed this comparative analysis on two proteins, PDB ID 2E3H and PDB ID 2E3I.

Analyzing the results of the MD experiment revealed the strict preservation of the 3β -corner geometry in the protein and outside the protein (Table 1). The geometric characteristics were obtained by clustering the MD trajectory corresponding to the values for the conformer in the major cluster. By major cluster, we mean the conformer of the structure under study, which inhabits the trajectory for at least 80% of the total duration of the analysis. Table 1 presents the calculated values of the characteristics of the 3β -corner in the course of the MD experiments carried out with an individual SSS and the whole protein, where this 3β -corner is its constituent block. As an example, the table lists the characteristics of the two 3β -corners recognized in two proteins: 2E3H, chain A, region: 216–246 and PDB ID 2E3I, chain A, region: A62–A93 (Table S1).

Structure	SASA \pm SD (Ų)	$R_g \pm SD$	$B_{\rm H}\pm SD$	RMSD \pm SD (Å)	
PDB ID 2E3H, chain A, region: 216–246					
PDB structure *	3202.64	1.3	18	-	
MD: 3β-corner **	3061.3 ± 150.4	1.23 ± 0.3	14 ± 1	2.77 ± 0.5	
MD: whole protein ***	3606.075 ± 13.499	1.3 ± 0.04	16 ± 0	6.67 ± 0.72	
PDB ID 2E3I, chain A, region: A62–A93.					
PDB structure	3335.75	1.2	22	-	
MD: 3β-corner	3085.9 ± 130.3	1.19 ± 0.1	14 ± 1	2.1 ± 0.9	
MD: whole protein	3846.33 ± 4.59	1.23 ± 0.01	19 ± 2	6.86 ± 0.37	

Table 1. Comparative analysis of the geometric characteristics of the 3β -corner in a protein environment and autonomously from protein in an aqueous medium (three technical repetitions).

Designations: SASA is the area available for the solvent, $Å^2$; Rg is the gyration radius; B_H is the hydrogen bond; and PDB structure * is the experimental structure extracted from the PDB bank. Values are given for the conformer in the major cluster of the MD trajectory. Values are given for the structure of the 3 β -corner in the protein; MD: 3 β -corner **—results of MD experiment for autonomous 3 β -corner outside the protein globule; MD: whole protein ***—MD experiment of a whole protein containing a 3 β -corner; the results are shown for the whole protein (three technical repetitions).

The values of the solvent-accessible area, gyration radius, and number of hydrogen bonds of the 3β -corner structure were determined before the start of the experiment (experimental structure from the PDB bank), and their changes throughout the course of the experiment were analyzed. Table 1 illustrates the values of the studied parameters for the major cluster of the 3β -corner that participates in the experiment autonomously (outside the protein molecule) and as a part of the protein. The table lists the number of hydrogen bonds. For example, for 2E3I in the initial structure, 14 were found. Table 1 lists the RMSD values calculated for these structures during the MD experiment. For example, the following were observed for 2E3H:

- between the initial structure (PDB) and the structure outside the protein, it was equal to 2.8 Å;
- between the initial structure and the structure in the composition of the protein, it was equal to 5.6 Å.

These RMSD values were relatively small and favored the stability of the 3β -corner outside the protein molecule. The values of the studied characteristics 2E3I, chain A, region: A62–A93 also speak in favor of the stability of the 3β -corner outside the protein molecule.

To check the stability of the studied 3β -corners in the course of a computational experiment, we recorded the values of the torsion angles of each amino acid residue of the structure to analyze the change in conformation. According to the distribution of the limiting values of the angles φ and ψ on the Ramachandran map for all possible regions, a conformational description of all the remains of the studied structures was performed. Figure 2 shows the distributions of the values of the angles φ and ψ on the Ramachandran map of the studied amino acids. 3β -corners 2E3H and 2E3I were used during the MD experiment. The drawing contains two cards: Ramachandran plot for all amino acid motives, and for glycine. Green dots are the calculated angles for the initial structure (PDB), yellow dots are the calculated angles for the major cluster of the 3β -corner in the protein, and black dots are the angles for the major cluster of the 3β -corner that participated in the autonomous experiment. Since the amino acid sequence of the motif under study does not contain proline, the analysis of the allowed proline regions was excluded. From the figure, we see that the investigated amino acid residues in the 3β -corner did not leave the allowed areas. The angles of the studied amino acids for the original structure, the major cluster of the 3β -corner in the protein composition, and the major cluster of the 3β -corner that participated in the experiment autonomously were always located close to each other. All these peaks (as well as other parameters analyzed in the MD experiment) also favored the stability of this structure.



Figure 2. Distribution of angles φ and ψ on the Ramachandran map of the studied amino acid residues of the 3 β -corners of 2E3H, chain A, region: 216–246 and 2E3I, chain A, region: A62–A93 during the MD experiment. The calculated angles are highlighted: for the initial structure, in yellow; the angles for the major cluster of the 3 β -corner in the protein composition, in green; and the angles for the major cluster of the 3 β -corner that participated autonomously in the experiment, in black. (a) Ramachandran map for all amino acid residues (except for amino acids highlighted on a separate map) and glycine GLY for 2E3H. (b) Schematic representation of the 2E3I of the original motif (green), the major cluster of the 3 β -corner within the protein (red), and the major cluster of the 3 β -corner that participated autonomously in the experiment (grey). (c) Ramachandran map for all amino acid before proline; Pre-PRO, proline PRO for 2E3H. (d) Schematic representation of the 2B3I of the original motif (green), the major cluster of the 3 β -corner that participated autonomously in the experiment (grey). (c) Ramachandran map for all amino acid before proline; Pre-PRO, proline PRO for 2E3H. (d) Schematic representation of the 2B3I of the original motif (green), the major cluster of the 3 β -corner that participated autonomously in the experiment (grey). (c) Ramachandran map for all amino acid before proline; Pre-PRO, proline PRO for 2E3H. (d) Schematic representation of the 2B3I of the motif (green), the major cluster of the 3 β -corner within the protein (red), and the major cluster of the 3 β -corner that participated autonomously in the experiment (grey).

Figure 2b,d illustrates the overlay of images of the original motif, major cluster of the 3β -corner within the protein, and major cluster of the 3β -corner that participated in the experiment outside the protein (autonomously). We could clearly see the complete

coincidence of the elements of the secondary structure, β -strands. Minimal deviation of irregular areas (constrictions) was observed between the three analyzed images of the studied 3β -corner. However, the magnitude of the deviation, that is, the distance between the irregular sections of the structures, was insignificant. This fact, established during the experiment, proved that this super-secondary structure is a stable autonomous structure in aquatic environments. Ramachandran maps showed the distribution of the values of the angles φ and ψ for the studied amino acid residues of two types of the 3β -angle during the MD experiment. We can state that most of the studied amino acids for both studied structures, for major clusters of 3β -corners in the composition of the protein and for major clusters of autonomous 3β -corner structures is typical and not limited to the cases described.

In the second stage, we examined the possibility of reducing the duration of the MD experiment for autonomous 3β -corners in an aqueous medium. A comparative analysis was conducted on the behavior of the structures of the 3β -corners in a "long" trajectory (300 ns) and a "short" trajectory (20 ns). This methodology was guided by several arguments to reduce the duration of MD. Throughout most of our experiments, we observed that clustering the MD trajectory allowed us to clearly identify the conformer that inhabits the trajectory at least 80% of the time. In the MD experiment, we did not observe any significant changes in the values of the geometric characteristics for the 3β -angles between the "long" and "short" trajectories. A justified reduction in MD time makes it possible to significantly improve calculation results.

Results of "long" and "short" MD for some instances of 3β -corner structures excised from 1OMV (PDB ID), 2CX8 (PDB ID), 1V6Z (PDB ID), and 2BJQ (PDB ID) proteins (Figure 3) showed that the lower values of RMSD in both "long" and "short" trajectory dynamics corresponds to sections of the amino acid sequences identified as β -strand (in Figure 3, black lines at the bottom). On the contrary, unstructured areas (Figure 3, blank region on the bottom) were subject to greater variability in MD experiments, which was the most prominent for 2BJQ region, A32-A62 (Figure 3d). Results with similar values of RMSD were observed in the "long" and "short" MD experiments (Table S2).



Figure 3. Cont.



Figure 3. Comparison of changes in the values of the RMSD along the "long" (red color) and "short" (blue color) MDs. Unstructured sections of polypeptide chain are marked with omissions, whereas β -strands correspond to black lines at the bottom of each panel. Averages value of RMSD and associated standard deviations. Typical cases are presented for structures of the 3 β -corner type: 10MV: region A605–A634 (**a**), 2CX8: region A33–A60 (**b**), 1V6Z: region A33–A64 (**c**), and 2BJQ region: A32–A62 (**d**). All changes were made in three technical repetitions.

On the basis of the obtained results of the "long" and "short" MD, we assert that the 3β -corners in the composition of the studied set retained their stability in the aqueous environment. The optimal time for the MD experiments was 20 ns. Further reduction in the duration of MD was deemed inappropriate. The next section describes the MD results for the entire dataset.

2.2. Control Experiments

Control experiments were performed with the involvement of several types of protein structures close to 3β -corner structures in the length of amino acid sequence but were different in terms of spatial arrangement (Figure 4). It was demonstrated (Figure 4) that negative controls embodied structures that did not contain β -strands, as well as α -helix. The selected testing structures were characterized by radius of gyration and solvent-accessible areas comparable to the study set of 3β -corner structures (Table S3).



Figure 4. A set of overlapped protein structures used as a negative control. (**a**) Brain tumor protein (drosophila melanogaster) PDB ID 1Q7F region: A761–A789; (**b**) exosome complex component Rrp4 (archaeoglobus fulgidus) PDB ID 2BA0 region: A3–A21; (**c**) histone-lysine N-methyltransferase EHMT1 (homo sapiens) PDB ID 3HNA region: A993–A1016; (**d**) putative pectate lyase L (bacteroides thetaiotaomicron) PDB ID 5OLQ region: A58–A80. Color-scale legend: green color indicates structure in the protein composition (PDB data); blue, pink, and yellow colors define technical repetitions; **—chain A, region (PDB).

The MD experiments displayed changes in the geometry of negative control structures and structures in the 3β -corner dataset. The results of MD experiments in three technical repetitions (20 ns) revealed high variability of RMSD values and a low number of hydrogen bonds among the selected negative control structures, which defines them as unstructured tangles (Figure 4). Nevertheless, visualization before and after the MD experiment showed that the structures 1Q7F and 2BA0 arranged a small β -hairpin and one turn of the α -helix, respectively (according to STRIDE), only in one of the technical replications. Meanwhile, the gyration radius and solvent-accessible area were significantly smaller compared to those for the 3β -corner set (Table S3).

2.3. Results of the MD Experiment for 3β -Corners

Comparative analysis was managed for 330 experimental PDB structures under the study, and following the MD operation, the number of hydrogen bonds, alterations of permitted and prohibited areas of the Ramachandran map (Figure 5), and solvent-accessible areas (the number of water molecules in the contact or residual water in $Å^2$) were calculated.



Structures

Figure 5. Comparative analysis of MD geometric characteristics for 330 structures of 3β -corners of experimental structures (exp) versus structures after MD (data are shown in three technical replicates). Distribution of (**a**) solvent-accessible surface area (SASA, Å²); (**b**) gyration radius (Rg); (**c**) the number of hydrogen bonds (HB); (**d**) RMSD (nm). All the structures involved in the experiment were deposited along the OH axis. The order of structures along the OH axis was generated separately for each characteristic and arranged in incremental order (from maximum to minimum). Red lines indicate measurements for the structures of the control set (negative control); green lines indicate measurements of selected motifs within a whole protein molecule; black lines designate measurements of characteristics for 3 β -corners of the target dataset.

Results of MD experiments for all structures examined in three technical repetitions are summarized in Figure 5. Variation of solvent-accessible area (Figure 5a) of the experimental 3β -corner structures during molecular dynamics was estimated as

$$\Delta SASA = SASA (EXP) - SASA (MD)$$
(1)

Whether or not the value was equal to zero along the OY axis (marked with a black line) indicated the unaltered solvent-accessible area of the structure during the MD experiment;

thus, the indicator remained equal to that before the MD experiment (Equation (1). We noticed that most of the structures were localized close to a y~0 value, meaning a slight change of the measured parameter during MD modeling. However, even if the measured value was not equal to zero, most of structures lay along the selected axis (y = 0). The SASA values were calculated according to three technical specifications and represented by value spreads (Figure 5a, red lines), the largest of which were found among structures of the control set (negative control) suggesting instability of structures. At the same time, we can note that SASA had a minimal variation in the three techniques for the selected motifs that participated in the MD experiment as part of a protein molecule, which implied the stability of structures. Values of Δ SASA were characterized by a small spread for three technical repetitions among 3β -corner structures of the target dataset.

Similarly, most of the tested structures demonstrated a slight change of the radius of gyration (Equation (2)) through three replications of the MD experiment (Figure 5b).

$$\Delta Rg = Rg(EXP) - Rg(MD)$$
(2)

The ΔRg values of the control set structures were also characterized by a small spread, and the behavior of these structures for this indicator fell within the general scenario.

The number of hydrogen bonds (Equation (3)) is yet another important characteristic of the studied structures. The number of hydrogen bonds and their quantitative change were determined for each structure through the MD experiment (Figure 5c).

$$\Delta H_B = H_B (EXP) - H_B (MD)$$
(3)

The zero value on the OY axis (y = 0) means the lack of quantitative changes of hydrogen bonds during the experiment. Most of the structures lost the minimum number of hydrogen bonds during the MD experiment, suggesting the stability of structures. Moreover, there was minimal spread of Δ HB through over technical replicates of MD in the majority of tested structures assuming high stability of the 3 β -corners (Figure 5c). The amount of hydrogen bonds broken among the studied structures reached 10 and ranged from 0% to 50% of the total amount of bonds established before the experiment. The Δ HB spread was insignificant in the negative control set, probably because structures of the control set did not have a large number of hydrogen bonds initially; thence, such protein blocks did not have a hydrophobic core stabilizing the structure.

The mean RMSD value for three replicates of MD experiments varied roughly less than 5 Å among the structures of the target dataset in each single case. This confirmed the hypothesis that the 3 β -corner is a stable structure. On the contrary, structures of the negative control set showed wider spread of RMSD, significantly exceeding out of 5 Å and reaching up to 10 Å, suggesting low stability of the control set structures.

More detailed information regarding results of the MD testing under the studied structures can be approached in Tables S2 and S4. Specifically, the radius of gyration, solvent-accessible area, the number of hydrogen bonds, and retention of amino acid residue position at a certain zone of the Ramachandran map for each of the examined 330 structural motifs were calculated and collected (Table S4). These data were associated with coordinates of localization of motifs (3 β -corners) being found in tested proteins and with an average value of RMSD for such motifs. It was found that only 18 3 β -corner structures out of 330 bore the calculated average RMSD values for three MD techniques more than 5 Å, whereas 50 structures were characterized by RMSD values less than 2 Å, and the rest of the structures encompassed a range between 2 and 5 Å (Table S4). The narrow range of calculated values evidenced the stability of 3 β -corner turning a fresh look at this structure and to consider it as an independent object of research in the field of structural biology. On the contrary, structures of the control set were featured by greater variability of calculated parameters.

The beguiling result was achieved next to the analysis of probable contacts between amino acid residues involved in the organization and stabilization of the hydrophobic core of 3β -corners. Indeed, we observed that hydrophobic amino acid residues hold about 40% (Table S5) of the total amount of residues in the selected set of structures. Analysis of probable contacts between amino acid residues revealed a bimodal distribution of distances between interacting hydrophobic amino acids (Figure 6).



Figure 6. Distribution of distances between interacting hydrophobic amino acid residues in 3ßcorners: within one β -strand ("inside-strand") and between different β -strands ("cross-strand"): red color indicates contacts within one strand, blue color indicates contacts of hydrophobic amino acid residues between β -strands A and B, yellow color designates contacts between β -strands B and C, and green color specifies contacts between β -strands A and C.

The number of contacts arranged within one β -strand (inside-strand) was only a part of the likely contacts (Figure 6, red color), while most of the contacts were attributed to interactions of amino acid residues localized in different β -strands ("cross-strand").

There were two upper limits in the distribution of distances between interacting hydrophobic amino acid residues in the 3 β -corners that can be determined: (1) the "inside-strand" type interactions and (2) contacts between adjacent β -strands in the amino acid sequence (A \rightarrow B, B \rightarrow C, and A \rightarrow C). The maximum distribution of distances between contacting amino acids was at 6,8–7 Å and the second maximum was at 9.4–9.6 Å. It should be noticed that the distance 9.4–9.6 Å was characteristic of contacts between the first and the third β -strands (A \rightarrow C) of 3 β -corners since these elements of SSSs (by definition) were localized on different orthogonal planes, so the distances between them were greater. Thus, the 3 β -corner structures were distinguished by a high proportion of hydrophobic amino acid residues and "cross-strand"-type contacts, unique compact stacking, and autonomous stability in a molecular dynamic experiment.

3. Discussion

Research on SSSs is important due to unique and compact spatial packing of the polypeptide chain, which makes it possible to consider SSSs as possible nuclei for protein folding. Studies on the modeling of protein structures and stability of the obtained proteins have been conducted using molecular dynamics simulation experiments [13,14]. The present study analyzed the autonomous stability of the 3β -corner-type structural motif in an aqueous medium outside the protein environment. The MD study of the structural motif, which is small (in terms of the number of atoms) compared to the entire protein molecule, is characterized by high performance, relatively low computational requirements, and low experimental costs.

This study aimed to substantiate the possibility of studying the structure of the 3β -corner type outside the protein globule. To do so, we selected a set of 3β -corners of

330 structures extracted from the PDB bank. The generated structure database facilitated a comprehensive study of the characteristics of the motif when examining a larger set.

We conducted molecular dynamics experiments to establish the autonomous stability of a super-secondary structure of the 3 β -corner type outside the protein environment. The stability of conformational templates was analyzed on the basis of the distribution of the angles φ and ψ on the Ramachandran map during the experiment. A conformational description of the amino acid residues of the studied structures was performed according to the distribution of the limiting values of the angles φ and ψ on the Ramachandran map for all possible regions. The change in the conformation of the structure of the motif under the study was analyzed in the molecular dynamics experiment by recording the frequency of hitting the torsion angles of amino acid residues in certain areas.

In the study, we determined that the structures of 3β -corners behave in the same way in the MD experiment. Thus, analysis of the changes in geometric characteristics did not reveal significant differences between the experimental structures extracted from the PDB bank and the autonomous structures of the 3β -corners after MD operation. On the basis of the obtained results, we recommend an MD duration of 20 ns.

4. Materials and Methods

4.1. Dataset of 3β-Corner Structures

In this study, we focused on a simple type of SSS— 3β -corner. The aim of this study was to analyze the stability of 3β -corner as an autonomous unit or outside the protein globule in an aqueous environment. A set of 330 structures of the 3β -corner type was selected from the Protein Data Bank (PDB) (https://www.rcsb.org/, accessed on 5 August 2022) to perform MD experiments with the following analysis of standard deviation, change in gyration radius, solvent-accessible area, lifetime of the major conformer and torsion angles, and the number of hydrogen bonds reported to characterize the stability of proteins. Super-secondary 3β -corner structures are widely distributed, and there are a few small proteins consisting of only the 3β -corner and short irregular regions in nature [15,16].

The collected dataset was organized from β -proteins, most of which contained small β-barrels according to the SCOP [17] classification (https://scop.berkeley.edu/, accessed on 5 August 2022) folds b.34 (SH3-like; 21 superfamilies), b.43 (common domain of reductase/isomerase/elongation factor; three superfamilies), b.47 (trypsin-like serine proteases; one superfamily), b.55 (domain-like PH; one superfamily), etc. (see Table S6) [17–19]. Small stacks of β -barrel represent a closed structure, wherein the first and last β -weights were stabilized by hydrogen bonds [20]. The β -barrel structures can also be represented as two orthogonally arranged β -sheets [21]. Such β -barrel structures are characterized by a few β -strands (n = 3–5) and an elliptical type of cross-section with high shear values $(S\sim6-10)$, which provide a tight fit—a "flattened" ellipse. The 3 β -corner structural motifs ordered as " β -Strand \rightarrow Coil $\rightarrow \beta$ -Strand \rightarrow Coil $\rightarrow \beta$ -Strand" were extracted from the β -barrel folds. The correlation of amino acid residues within the selected structures to elements of the secondary structure was elaborated by using three STRIDE algorithms [22], DSSP [23], and iCn3D [24]. The results of convergence between three algorithms utilized for the identification of elements of the secondary structure are presented in Table S7. The match rate between STRIDE/DSSP, Stride/iCn3D, and DSSP/iCn3D was approximately 50% (Table S7). β -Strands in the 3 β -corner structure are usually short and consist of 4–6 to 10 amino acid residues, connected by loops. Short loops (from three to seven amino acid residues) shape a turn, while long loops (8-12 amino acid residues) are characterized by an unstructured shape. The studied dataset was represented by motifs extracted from proteins of various origins and containing different lengths of its constituent elements (Figure 7 and Table S4).



Figure 7. Origin of proteins in which 3β -corners were identified (**a**); length distribution of 3β -corners in the examined dataset (**b**).

SSSs were selected for homologous and non-homologous proteins of various origins (Figure 7a). Most of these proteins belong to humans, animals, and bacteria. There are also small groups of plant and viral proteins (Figure 7a). This observation is consistent with the variety of annotated protein structures in the PDB for different entities.

The analysis of lengths amongst selected 3β -corners showed that the majority of studied structures were 25–40 amino acids in length. Moreover, depending on the length, the maximum distribution of 3β -corners fell on 30–35 amino acid residues (Figure 7b), and only a few of the 3β -corners consisted of 45–55 amino acid residues.

The gathered 3β -corners were extracted from the PDB and SCOP and are presented in Table S6.

4.2. Molecular Dynamics Simulation

Simulations were performed using GROMACS software package (version 2021.4). The simulation process was identical for every simulation except simulation time (20 or 300 ns) and some parameters, depending on size and charge of every specific molecule, and was automatically calculated. Configuration and procedure already were used in our previous studies [25–27] but were modified for current research. Molecules were prepared with pdb2gmx tool (GROMACS, version 2021.4), which adds missing hydrogen atoms and forms the correct topology system using CHARMM36 force field converted for GROMACS. Molecules were placed into rectangular boxes for simulation. Box size was automatically configured (with gmx editconf tool) for each simulation depending on protein size, leaving at least 1.0 nm between molecule and box borders.

Boxes were also filled with water (gmx solvate tool, GROMACS, version 2021.4). Water was represented with SPC model selected as a compromise between simulation performance and realism of behavior in normal conditions (see [28] for details about water model comparison). The simulation system also was neutralized with Na+ or Cl⁻ ions. GROMACS automatically calculated the charge of the system and replaced some amount of water molecules in the box with ions to make the system neutral (using gmx genion tool). If the charge of the system was negative, it added positive ions (sodium), and if the system's charge was positive, it added negative (chlorine) ions.

The next preparation step was energy minimization (EM). EM is short specific simulation performed using the steepest decent algorithm. The simulation was performed until energy became lower than 1000.0 kJ/mol/nm or until 1000 simulation steps (0.002 ps per step [29]).

The final preparation was heating pre-run: the simulation system was heated from 5 to 311 K for 200 ps with 0.001 step using a v-rescale temperature coupling algorithm. The molecular dynamics simulation itself ran for 20 or 300 ns (different runs had different simulation times, which is mentioned in the corresponding sections of the study). Tem-

perature for simulation was kept at 311 K. The simulation was performed by GROMACS default leap-frog algorithm for integrating Newton's equations of motion. Simulation step was 0.002 ps.

Resulting trajectories were processed using GROMACS analysis utilities such as gmx rms, gmx gyrate, and gmx sasa. Clusterization was performed using the gmx clusters utilized. Cut-off value for clusterization was a varying value depending on the first clusterisation result with a cut-off = 0.3 nm: if the result contained > 12 clusters, clusterization was redone with a larger cut-off value until the cluster amount reached 20. Moreover, if there were < 8 clusters, clusterization was redone with a larger cutoff. The step for raising/lowering cutoff was set to 0.01 nm, and the limit for the clusterization attempt was set to 30.

RMSD changes were calculated for the backbone for all the trajectories after the heating step. These data in XG format were aggregated with our own specially written scripts. Intact values were taken from an initial PDB file or from the first step of the trajectory. MD resulting values were the average of values for trajectory parts where structure conformation matched the major cluster in the final clusterization result.

All the GROMACS configuration files are available on the following link https://github.com/protdb/beta-corner-stability/tree/master/GROMACS%20configuration (accessed on 5 August 2022).

A total of 330 structures were used in the MD experiments. In this study, the following MD experiments were carried out for SSS of the 3β -corner type:

- "long" dynamics, simulation duration 300 ns;
- "short" dynamics, simulation duration 20 ns;
- part of the whole protein, simulation duration of 300 ns.

According to the MD results, the changes in the following geometric characteristics of the SSS were analyzed:

- root-mean-square deviation (RMSD) and gyration radius change (Rg);
- solvent accessible surface area (SASA);
- lifetime of major conformer;
- change in torsion angles;
- change in the number of hydrogen bonds.

Data on the conditions of MD experiments are available on the following link https: //github.com/protdb/beta-corner-stability (accessed on 5 August 2022).

5. Conclusions

It can be concluded that in the water environment, the 3β -corner is rather stable per se, i.e., without the rest part of a protein molecule, and it can act as the nucleus or "ready-made" building block in protein folding. The 3β -corner can also be considered as an independent object for study in field of structural biology.

Understanding the protein architecture and recognizing and studying their individual structural blocks with unique and compact polypeptide chain folds provides an opportunity to gain insight regarding the structure, geometry, internal contacts, and patterns of organization of structural motifs of protein molecules. This knowledge provides a strong basis for addressing fundamental studies, such as prediction and modeling of three-dimensional protein structures, protein folding, and structural classification of proteins, as well as applied problems such as developing novel approaches for disease diagnosis and improving our understanding of pathogenesis, identifying drug targets, and designing proteins with desired properties (mimetics).

We annotated a set of 3β -corner SSSs in 330 structures. As a result of the MD experiments, we showed the autonomous stability of this type of super-secondary structure and demonstrated the possibility of analyzing them as independent objects. We observed the preservation of key geometric characteristics because of the MD of autonomous structures of 3β -corners in comparison with experimental structures extracted from the PDB. We

present an annotated set of 3β -corners of the research in the fields of structural biology and biomedicine.

Supplementary Materials: The following supporting information can be downloaded at https: //www.mdpi.com/article/10.3390/ijms231911674/s1.

Author Contributions: Conceptualization, A.L.K., L.I.K. and V.R.R.; methodology, K.S.N., D.V.P. and V.R.R.; software, K.S.N., A.A.S. and D.V.P.; validation, K.S.N., D.V.P. and A.V.E.; investigation, V.R.R., D.V.P. and A.T.K.; data curation, A.M.K. and L.I.K.; writing—original draft preparation, A.L.K., L.I.K. and K.S.N.; writing—review and editing, V.R.R., A.V.E. and K.A.M.; visualization, K.S.N., K.A.M. and D.V.P. All authors have read and agreed to the published version of the manuscript.

Funding: The work was conducted in the framework of the Russian Federation Fundamental Research Program for the long-term period for 2021–2030 (No. 122092200056-9).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to A. V. Veselovsky for helpful discussions and Amir Taldaev for organizing calculations at the computational facilities of the Tyumen State University (Tyumen, Russia). Simulations were performed using the equipment of the shared research facilities of the HPC computing resources at Lomonosov Moscow State University and the Joint Supercomputer Center of the Russian Academy of Sciences.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Rao, S.T.; Rossmann, M.G. Comparison of Super-Secondary Structures in Proteins. J. Mol. Biol. 1973, 76, 241–256. [CrossRef]
- 2. Levitt, M.; Chothia, C. Structural Patterns in Globular Proteins. *Nature* **1976**, *261*, 552–558. [CrossRef] [PubMed]
- 3. Efimov, A.V. Standard Structures in Proteins. Prog. Biophys. Mol. Biol. 1993, 60, 201–239. [CrossRef]
- 4. Efimov, A.V. Favoured Structural Motifs in Globular Proteins. Struct. Lond. Engl. 1993 1994, 2, 999–1002. [CrossRef]
- 5. Efimov, A.V. Structural Trees for Protein Superfamilies. Proteins Struct. Funct. Bioinform. 1997, 28, 241–260. [CrossRef]
- 6. Scholl, Z.N.; Yang, W.; Marszalek, P.E. Competing Pathways and Multiple Folding Nuclei in a Large Multidomain Protein, Luciferase. *Biophys. J.* 2017, 112, 1829–1840. [CrossRef]
- 7. Xia, X.; Longo, L.M.; Sutherland, M.A.; Blaber, M. Evolution of a Protein Folding Nucleus. *Protein Sci. Publ. Protein Soc.* 2016, 25, 1227–1240. [CrossRef]
- Du, D.; Zhu, Y.; Huang, C.-Y.; Gai, F. Understanding the Key Factors That Control the Rate of β-Hairpin Folding. *Proc. Natl. Acad. Sci. USA* 2004, 101, 15915–15920. [CrossRef]
- 9. Cruz, L.; Rao, J.S.; Teplow, D.B.; Urbanc, B. Dynamics of Metastable β-Hairpin Structures in the Folding Nucleus of Amyloid β-Protein. *J. Phys. Chem. B* 2012, *116*, 6311–6325. [CrossRef]
- 10. Rudnev, V.R.; Kulikova, L.I.; Nikolsky, K.S.; Malsagova, K.A.; Kopylov, A.T.; Kaysheva, A.L. Current Approaches in Supersecondary Structures Investigation. *Int. J. Mol. Sci.* **2021**, *22*, 11879. [CrossRef]
- 11. Efimov, A.V. A Novel Super-Secondary Structure of Proteins and the Relation between the Structure and the Amino Acid Sequence. *FEBS Lett.* **1984**, *166*, 33–38. [CrossRef]
- 12. Kmiecik, S.; Kolinski, A. Folding Pathway of the B1 Domain of Protein G Explored by Multiscale Modeling. *Biophys. J.* 2008, *94*, 726–736. [CrossRef] [PubMed]
- Arnittali, M.; Rissanou, A.N.; Amprazi, M.; Kokkinidis, M.; Harmandaris, V. Structure and Thermal Stability of WtRop and RM6 Proteins through All-Atom Molecular Dynamics Simulations and Experiments. *Int. J. Mol. Sci.* 2021, 22, 5931. [CrossRef] [PubMed]
- 14. Liwo, A.; Khalili, M.; Scheraga, H.A. Ab Initio Simulations of Protein-Folding Pathways by Molecular Dynamics with the United-Residue Model of Polypeptide Chains. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2362–2367. [CrossRef]
- 15. Hammarström, A.; Berndt, K.D.; Sillard, R.; Adermann, K.; Otting, G. Solution Structure of a Naturally-Occurring Zinc-Peptide Complex Demonstrates That the N-Terminal Zinc-Binding Module of the Lasp-1 LIM Domain Is an Independent Folding Unit. *Biochemistry* **1996**, *35*, 12723–12732. [CrossRef]
- 16. Pérez-Alvarado, G.C.; Miles, C.; Michelsen, J.W.; Louis, H.A.; Winge, D.R.; Beckerle, M.C.; Summers, M.F. Structure of the Carboxy-Terminal LIM Domain from the Cysteine Rich Protein CRP. *Nat. Struct. Biol.* **1994**, *1*, 388–398. [CrossRef]
- 17. Murzin, A.G.; Brenner, S.E.; Hubbard, T.; Chothia, C. SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. J. Mol. Biol. 1995, 247, 536–540. [CrossRef]

- Chandonia, J.-M.; Fox, N.K.; Brenner, S.E. SCOPe: Manual Curation and Artifact Removal in the Structural Classification of Proteins—Extended Database. J. Mol. Biol. 2017, 429, 348–355. [CrossRef]
- 19. Youkharibache, P.; Veretnik, S.; Li, Q.; Stanek, K.A.; Mura, C.; Bourne, P.E. The Small β-Barrel Domain: A Survey-Based Structural Analysis. *Struct. Lond. Engl.* 1993 **2019**, 27, 6–26. [CrossRef]
- 20. Murzin, A.G.; Lesk, A.M.; Chothia, C. Principles Determining the Structure of Beta-Sheet Barrels in Proteins. II. The Observed Structures. *J. Mol. Biol.* **1994**, *236*, 1382–1400. [CrossRef]
- Chothia, C.; Janin, J. Orthogonal Packing of Beta-Pleated Sheets in Proteins. *Biochemistry* 1982, 21, 3955–3965. [CrossRef] [PubMed]
 Heinig, M.; Frishman, D. STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. *Nucleic Acids Res.* 2004, 32 (Suppl. 2), W500–W502. [CrossRef] [PubMed]
- 23. Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577–2637. [CrossRef] [PubMed]
- 24. Xu, D.; Zhang, Y. Generating Triangulated Macromolecular Surfaces by Euclidean Distance Transform. *PLoS ONE* **2009**, *4*, e8140. [CrossRef] [PubMed]
- 25. Taldaev, A.; Rudnev, V.; Kulikova, L.; Nikolsky, K.; Efimov, A.; Malsagova, K.; Kaysheva, A. Molecular Dynamics Study of Citrullinated Proteins Associated with the Development of Rheumatoid Arthritis. *Proteomes* **2022**, *10*, 8. [CrossRef]
- 26. GROMACS Tutorial. A Guide to Molecular Dynamics Using GROMACS by Dimitar Dimitrov. Available online: https://static.igem.org/mediawiki/2021/d/d2/T--KCL_UK--UPDATEDGROMACSGUIDE.pdf (accessed on 15 August 2022).
- [Gmx-Users] Simulated Annealing Protocol. Available online: https://mailman-1.sys.kth.se/pipermail/gromacs.org_gmx-users/ 2012-March/069466.html (accessed on 15 August 2022).
- 28. Water Models. Available online: https://water.lsbu.ac.uk/water/water_models.html (accessed on 15 August 2022).
- 29. Molecular Dynamics Parameters (.mdp Options)—GROMACS 2022.3 Documentation. Available online: https://manual.gromacs. org/documentation/current/user-guide/mdp-options.html (accessed on 15 August 2022).





Article Deep-Learning to Predict BRCA Mutation and Survival from Digital H&E Slides of Epithelial Ovarian Cancer

Camilla Nero ^{1,*}, Luca Boldrini ², Jacopo Lenkowicz ², Maria Teresa Giudice ¹, Alessia Piermattei ³, Frediano Inzani ³, Tina Pasciuto ⁴, Angelo Minucci ⁵, Anna Fagotti ¹, Gianfranco Zannoni ³, Vincenzo Valentini ⁶ and Giovanni Scambia ¹

- ¹ Fondazione Policlinico Agostino Gemelli, IRCCS, Gynecology and Obstetrics, 00168 Rome, Italy
- ² Fondazione Policlinico Agostino Gemelli, IRCCS, Radiomics Core Facility, 00168 Rome, Italy
- ³ Fondazione Policlinico Agostino Gemelli, IRCCS, Pathology, 00168 Rome, Italy
- ⁴ Fondazione Policlinico Agostino Gemelli, IRCCS, Data Collection Core Facility, 00168 Rome, Italy
- ⁵ Fondazione Policlinico Agostino Gemelli, IRCCS, Genomics Core Facility, 00168 Rome, Italy
- ⁶ Fondazione Policlinico Agostino Gemelli, IRCCS, Radiation Oncology, 00168 Rome, Italy
- * Correspondence: camilla.nero@policlinicogemelli.it; Tel.: +39-06-30154979

Abstract: BRCA 1/2 genes mutation status can already determine the therapeutic algorithm of high grade serous ovarian cancer patients. Nevertheless, its assessment is not sufficient to identify all patients with genomic instability, since BRCA 1/2 mutations are only the most well-known mechanisms of homologous recombination deficiency (HR-d) pathway, and patients displaying HR-d behave similarly to BRCA mutated patients. HRd assessment can be challenging and is progressively overcoming BRCA testing not only for prognostic information but more importantly for drugs prescriptions. However, HR testing is not already integrated in clinical practice, it is quite expensive and it is not refundable in many countries. Selecting patients who are more likely to benefit from this assessment (BRCA 1/2 WT patients) at an early stage of the diagnostic process, would allow an optimization of genomic profiling resources. In this study, we sought to explore whether somatic BRCA1/2 genes status can be predicted using computational pathology from standard hematoxylin and eosin histology. In detail, we adopted a publicly available, deep-learning-based weakly supervised method that uses attention-based learning to automatically identify sub regions of high diagnostic value to accurately classify the whole slide (CLAM). The same model was also tested for progression free survival (PFS) prediction. The model was tested on a cohort of 664 (training set: n = 464, testing set: n = 132) ovarian cancer patients, of whom 233 (35.1%) had a somatic BRCA 1/2 mutation. An area under the curve of 0.7 and 0.55 was achieved in the training and testing set respectively. The model was then further refined by manually identifying areas of interest in half of the cases. 198 images were used for training (126/72) and 87 images for validation (55/32). The model reached a zero classification error on the training set, but the performance was 0.59 in terms of validation ROC AUC, with a 0.57 validation accuracy. Finally, when applied to predict PFS, the model achieved an AUC of 0.71, with a negative predictive value of 0.69, and a positive predictive value of 0.75. Based on these analyses, we have planned further steps of development such as proving a reference classification performance, exploring the hyperparameters space for training optimization, eventually tweaking the learning algorithms and the neural networks architecture for better suiting this specific task. These actions may allow the model to improve performances for all the considered outcomes.

Keywords: ovarian cancer; somatic BRCA mutational status; digital pathology; machine learning; artificial intelligence

1. Introduction

Epithelial ovarian cancer (EOC) is strongly dominated by copy number changes without a focal gene driving mutation. Approximately half of cases exhibits defective DNA

Citation: Nero, C.; Boldrini, L.; Lenkowicz, J.; Giudice, M.T.; Piermattei, A.; Inzani, F.; Pasciuto, T.; Minucci, A.; Fagotti, A.; Zannoni, G.; et al. Deep-Learning to Predict BRCA Mutation and Survival from Digital H&E Slides of Epithelial Ovarian Cancer. Int. J. Mol. Sci. 2022, 23, 11326. https://doi.org/10.3390/ ijms231911326

Academic Editor: Alvaro Galli

Received: 30 August 2022 Accepted: 22 September 2022 Published: 26 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). repair via homologous recombination (HR), frequently caused by inactivation of the breast cancer susceptibility (BRCA) genes (overall up to 30%) [1–4].

Defective HR (HR-d) reflects underlying genomic instability, which has significant therapeutic implications in EOC. It is in fact associated to a striking platinum sensitivity and can be targeted by poly-ADP ribose polymerase inhibitors (PARPi).

Although HR assessment is progressively overcoming BRCA testing, it is not yet integrated in clinical practice and can be challenging and expensive, being also still not refundable in many countries. Selecting patients who are more likely to benefit from it at an early stage of the diagnostic process (somatic BRCA 1/2 wild type [WT] patients), could allow an optimization of genomic profiling resources.

Histologic phenotypes have been recognized to somehow reflect genetic alterations in cancer tissues [5–10]. BRCA-mutated (BRCA-mut) EOC were found to have more frequent Solid, pseudo-Endometrioid, and Transitional cell carcinoma-like morphology (SET features) and higher mitotic indexes compared to BRCA WT EOC [11].

Since hematoxylin and eosin (H&E)-stained tissue slides are ubiquitously available for cancer patients, predicting mutations from tissue slides could be a time- and cost-effective method to characterize patients and address them to genomic profile.

Advances in digital pathology and artificial intelligence have presented the potential to analyze gigapixel whole-slide images (WSI), providing information on tissue microenvironment, integrative image-omic, resistence to treatments but also somatic genomic status straight away [12–33].

WSI is a complex domain with several unique challenges, that requires different deep-learning-based computational pathology approaches such as manual annotation of gigapixel WSIs in fully supervised settings or large datasets with slide-level labels in weakly supervised ones.

The delineation of a pixel, patch or region-of-interest (ROI)-level annotations has produced promising results even if it cannot be directly generalized, and if it may suffer from noisy training labels and reduced reproducibility on data from different sources and imaging devices [34–37].

On the other hand, weakly supervised approaches demonstrated an exceptional clinical-grade performance [38] but require thousands of WSIs to achieve performance comparable to fully supervised and ROI-level classifiers and may not be suitable for multiclass subtyping problems.

A recent paper proposed a clustering-constrained-attention multiple-instance learning (CLAM) as a high-throughput deep-learning framework that aims to address the key challenges outlined above [39].

The authors demonstrated that such approach can be used to localize well-known morphological features on WSIs without the need for spatial labels, overperforming standard weakly supervised classification algorithms and resulting adaptable to independent test cohorts, smartphone microscopy and different tissue content.

In this study, we aimed at identifying somatic BRCA 1/2 mutational status directly from H&E slides using a CLAM-based approach. In particular we planned to evaluate negative and positive predictive values of the CLAM-based model compared to genomic sequencing as the reference standard. As secondary endpoints, we aimed at evaluating CLAM-based model accuracy in prognosis prediction measured as progression free survival (PFS).

2. Results

From November 2016 to November 2020, 1265 consecutive patients underwent BRCA 1/2 testing in our institution.

A total of 664 patients was finally analyzed in the current study (see Figure 1). Regarding secondary endpoints, 8 patients were lost at follow-up and 656 patients were therefore lastly included.



Figure 1. Flowchart of the study.

Table 1 shows main clinic-pathological characteristics. Overall, median age of included patients was 61 years old. More than half of the patients had positive family history for cancers (mainly breast). The vast majority of the population had a serous histotype (95.9%), grade 3 (97.1%) and III or IV FIGO stage (92.4%).

Table 1. Clinical, pathological and surgical characteristics of the study population.

= 664
2.1
7)
.5
5.6)
4)
39.5)
3.9)
5.7)
26.2)
(3.3)
Э)
1))

Tabl	le 1.	Cont.

Characteristic	All Cases <i>n</i> = 664
Grading	
1	6/661 (0.9)
2	8/661 (1.2)
3	642/661 (97.1)
Not applicable	5/661 (0.8)
FIGO Stage	
IIB	14 (2.1)
IIIA1	6 (0.9)
IIIA2	4(0.6)
IIIB	26 (3.9)
IIIC IVA	390 (38.7)
IVB	186 (28)
Somatic BRCA	100 (20)
Wild type	431 (64.9)
Mutated	233 (35.1)
BRCA 1	120/233 (51.5)
BRCA 2	71/233 (30.5)
BRCA 1-2	1/233 (0.4)
VUS BRCA 1	19/233 (8.2)
VUS BRCA 2	22/233 (9.4)
Germline BRCA	
Wild type	118/202 (58.4)
BRCA 1	48/202 (23.8)
BRCA2	29/202 (14.4)
BKCA I and BKCA 2	0/202(0) 5/202(2,5)
VUS BRCA 2	2/202(2.3)
Type of mutation variant	2/202 (1)
Frameshift mutation	94/233 (40.4)
BRCA 1	56/94 (59.6)
BRCA 2	37/94 (39.5)
BRCA 1-2	1/94 (0.01)
Missense mutation	45/233 (19.4)
BRCA 1	7/45 (15.5)
BRCA 2	5/45 (11.1)
VUS BRCA 1	16/45 (35.6)
VUS BRCA 2	17/45 (37.8)
Nonsense mutation	60/233 (25.8)
DKCA 1 BBCA 2	39/60 (65) 20 /60 (22 2)
VUS BRCA 2	20/00 (33.3)
Splicing mutation	15/233(64)
BRCA 1	5/15 (33.3)
BRCA 2	5/15 (33.3)
VUS BRCA 1	2/15 (13.4)
VUS BRCA 2	3/15 (20)
Mutation in Copy Number Variation	14/233 (6)
BRCA 1	12/14 (85.7)
BRCA 2	2/14 (14.3)
Splicing/Missense mutation	2/233 (0.8)
BRCA 2	2/2 (100)
Mutation in 3' UTR	1/233 (0.4)
VUS BRCA 2	1/1 (100)
	1/233 (0.4)
VUS DKCA I	1/1(100) 1/232(0.4)
VUS BRCA 1	1/200 (0.4)
, CO DICCI I	1/1 (100)

Characteristic	All Cases $n = 664$	
Type of surgery		
PDS	294 (44.3)	
Diagnostic	370 (55.8)	
Further surgery		
None	88/664 (13.3)	
Attempt at IDS	28/664 (4.2)	
ÎDS	245/664 (36.9)	
Restaging	3/664 (0.5)	
Other	6/664 (0.9)	
LPS PIV at diagnosis §	8 (0–14)	
Residual tumor at debulking		
0	254/294 (86.4)	
$\leq 1 \text{ cm}$	27/294 (9.2)	
>1 cm	13/294 (4.4)	
CRS		
1	13/55 (23.6)	
2	26/55 (47.3)	
3	16/55 (29.1)	

Results are presented as n (%) except where indicated. BMI: Body Mass Index. FIGO: Federation of International of Gynecologists and Obstetricians. BRCA: BReast CAncer gene. VUS: Variants of Uncertain Significance. PDS: Primary Debulking Surgery. IDS: Interval Debulking Surgery. LPS: LaParoScopy. PIV: Peak Integral Value. * Information available for 642/664 patients. § Information available for 649/664 patients.

Regarding BRCA 1/2 status, 431 (64.9%) patients resulted WT while 233 (35.1%) mutated, 51.5% of which were BRCA 1 mutated. We have reported specific mutations analyzed in the Table 1; with a majority of frameshift mutations of 40.4%.

All mutated patients were addressed to genetic counselling and 86.6% were tested for germline BRCA 1/2 pathogenetic variants. A third (38.2%) had a germline alteration.

Table 2 shows treatment and oncological outcome data. Regarding therapeutic choices, 54.5% of these patients underwent neoadjuvant chemotherapy with a median number of cycles prior to interval debulking surgery of 4, after a laparoscopic assessment. Three hundred and forty out of 644 (52.8%) recurred, but only 27.7% died.

Table 2. Treatment and oncological outcome of the study population.

Characteristic	All Cases <i>n</i> = 664	
Chemotherapy		
Clinical Setting		
Adjuvant	223/510 (43.7)	
Neoadjuvant	278/510 (54.5)	
Not Applicable	9/510 (1.8)	
Chemotherapy regimen		
Platinum based	454/501 (90.6)	
Other	47/501 (9.4)	
Bevacizumab	163/164 (99.4)	
Number of cycles *	14.8 ± 8.1	
Chemotherapy experimental protocol	118/488 (24.2)	
Dose frequency		
Thrice-weekly	392/485 (80.8)	
Weekly	85/485 (17.5)	
Quatri-weekly	8/485 (1.6)	

Table 1. Cont.

Characteristic	All Cases n = 664
Number of pre-IDS cycles §	4 ± 1.5
Number of post-IDS cycles †	2.5 ± 1.0
Total number of cycles ‡	6.1 ± 1.8
Time to chemotherapy II	105 ± 1481.8
Time to chemotherapy post-IDS ¥	39.8 ± 13.9
Toxicity	204/439 (46.5)
Change of chemotherapy regimen during treatment	43/458 (9.4)
Post-CHT Ca125 Ω	130 ± 585.6
Recist Response	
Complete	60/145 (41.4)
Partial (at least 30% reduction)	49/145 (33.8)
Stability	20/145 (13.8)
Progression (at least 20% increase)	16/145 (11.0)
Serological Response (GCIG)	
Complete	190/268 (70.9)
Partial	63/268 (23.5)
Stable	7/268 (2.6)
Progression	8/268 (3.0)
Relapse/Progression	
Yes	427/656 (65.1)
No	229/656 (34.9)
Death	179/647 (27.7)
Status	
Alive	463/647 (71.6)
Dead for the Disease	172/647 (26.6)
Dead for other causes	7/647 (1.1)
Lost at Follow Up	5/647 (0.8)
Overall survival Ψ	24.7 ± 14.8
PFI	
<6 Months	109/427 (25.5)
6–12 Months	126/427 (29.5)
>12 Months	192/427 (45.0)

Table 2. Cont.

Results are presented as *n* (%) except where indicated. IDS: Interval Debulking Surgery. GCIG: Gynecologic Cancer InterGroup. PFI: Platinum-Free Interval. * Information available for 127/664 patients. § Information available for 243 out of 278 patients. † Information available for 175/278 patients. ‡ Information available for 398/664 patients. II Information available for 494/664 patients. ¥ Information available for 184/278 patients. Ω Information available for 359/664 patients. Ψ Information available for 427/664 patients.

The whole process is represented in Figure 2. The outcome was BRCA 1/2 mutated yes/no, in which VUS were considered as mutated patients.

2.1. Phase 0: Reference Standard for BRCA Status Prediction

A reference classification performance was established according to the classification accuracy and AUC ROC of an expert pathologist, based on available criteria [11]. On the whole 664 slides of the dataset, the reference performance was as follows: accuracy 0.629, specificity 0.879, sensitivity 0.167, negative predictive value 0.661, positive predictive value 0.428 (TN: 379, TP: 39, FN: 194, FP: 52).



Figure 2. Pipeline of analysis.

2.2. Phase 1: WSI for Somatic BRCA Status Prediction

The dataset was randomly split into into a development set, consisting of a training set and internal validation set, and an hold-out testing. The proportion between development and testing set set was 80% to 20%. Thus, we used 464 images for training (244/220), 132 images for validation (86/46), and 68 images were hold out for testing (44/24).

The performance on the training set was 0.7 in terms of AUC, but on the testing set the AUC was 0.55. In detail, for training set class zero (BRCA wild type) the model correctly identified 153 out of 244 images, while for class one (BRCA mutated) 139 out of 220. In the testing set class zero, the model correctly identified 49 out of 86, while in class one 24 out of 46.

2.3. Phase 2: ROI on WSI for Somatic BRCA Status Prediction

For this analysis, a subset of images were used, because of the time consuming process of manual ROI delineation. The process of dataset splitting was the same as before, only with slightly different proportions, so that we used 198 images for training (126/72), 87 images for validation (55/32), and only three images were hold out for testing (2/1), merely to create heatmaps of activation regions on unseen images.

The outcome was BRCA 1/2 mutated yes or no, and VUS were considered as mutated patients.

At the end of training, the model reached a zero classification error on the training set, but the performance of the predictive model on the validation set was 0.59 AUC ROC, with 0.57 validation accuracy (see Figures 3 and 4). The model correctly classified 39 out of 55 class zero (BRCA wild type) images, and 11 out of 32 class one (BRCA mutated) images on the validation set. All of the three held out images were correctly classified.

In particular, the model assigned to the BRCA 1/2 mutated held out image a probability of 98% of mutation, while a probability of 38% and <1% were assigned to the other two held out images (both WT), respectively.



Figure 3. Phase1: Performance of the model during training (300 epochs). Top: validation AUC ROC. Low left: negative predictive value (NPV). Low right: positive predictive value (PPV).



Figure 4. Phase2: Performance of the model during training (300 epochs). Top: validation AUC ROC. Low left: negative predictive value (NPV). Low right: positive predictive value (PPV).

2.4. Phase 3: Exploration of the Hyperparameters Space for Training Optimization

Given the results obtained in the previous two phases, we looked for chances of performance improvement by exploring the CLAM model hyperparameters space through a grid search. To do so, we let the following hyperparameters vary: patch level between zero and 2; attention branch single or multiple; bag loss function and clustering loss function one between support vector machine or cross entropy; the relative weight of the two loss in the overall loss function between 0.2 and 0.8 with 0.1 steps; the number of highest and lowest attention patches to be fed to the clustering algorithm within the set 4, 8, 16, 32, 64, 100, 500, 1000. The dataset splitting was the same as phase 1. Grid search type was random search with an early stopping criterion on validation loss decreasing. For each patch level, the best five experiments in terms of validation AUC were retained for performance assessment on the testing set. None of these hyperparameters combination led to significant or even relevant BRCA classification performance improvement neither in the validation nor in the testing set.

2.5. Phase 4: WSI for Predicting Relapse

For this analysis, slide resolution was taken to be fixed at the highest available value on the image (called "patch level zero" in CLAM framework). As in the previous steps, the dataset was split into a development set (training and validation) and a testing set. Grid search on hyperparameters was performed on the development set to select the five highest performing models, which were later assessed on the hold-out testing set for predictive performance.

The combination of parameters led to a grid search on 64 different models for 200 epochs training length with an early stopping criterion of 20 epochs non-decreasing loss for each outcome.

On a total of 656 images (229 class 0; 427 class 1), 394 images were assigned to training set, 131 images to validation set, 131 images were hold out for testing (46/85). The AUC on the testing set was 0.71 (see Figure 5).



Figure 5. Phase 4: classification on testing set with the highest test AUC model. Top: validation AUC ROC. Bottom left: ROC curve on BRCA mut cases. Bottom right: ROC curves on BRCA WT cases.

3. Discussion

Molecular profiling in cancer patients has been increasingly important to determine the optimal therapeutic strategy. The combination of the digitization of pathology WSI with deep learning to predict somatic mutations, could be a promising approach to achieve a time- and cost-effective complementary method for personalized treatment.

When applied on our dataset, the available morphological criteria (SET features) showed disappointing results (accuracy 0.629, negative predictive value 0.661). This suggests that phenotype and genotype may not be strongly related, as previously suggested [11].

In our phase 1 (high testing and validation errors), we focused on the specific features/patterns within the tissue images that the model recognized to make response predictions. Looking at the activation map (heatmap) of the highest and lowest prediction on the testing set (see Figure 6a–c), we found that the model identified tumor cells in the mutated cases and stroma in the wild type case which could reflect the morphological differences previously described namely solid phenotype and higher mitotic index [11].

However, given the performances, we hypothesize that the highest attention pattern should be focused on tumor areas on which reported differences might be more evident thus useful for outcome prediction. It was necessary to tweak parameters, starting from optimal tissue identification in the segmentation process.

Unfortunately, the process of manually delineating ROI by a dedicated pathologist did not improve the overall performances and neither did the exploration of the hyperparameters space for training optimization, even tweaking the learning algorithms and the neural networks architecture for better suiting the task of BRCA 1/2 status identification.

Several issues and limitations have been encountered during the analysis.

First, the retrospective nature of the study represents an unavoidable source of selection bias and imaging data inhomogeneity.

Although we collected only H&E slices of peritoneal tissue and checked for a minimum percentage of tumor cells in all specimens, patients were not divided in subgroups according to the type of surgery performed; thus small biopsies obtained from exploratory laparoscopy might have provided less representative specimens and images of lower quality compared to peritonectomies.

Second, the absence of an external validation does not allow us to draw any definitive conclusion on the replicability of the model, though the use of H&E slides for cancer diagnosis is spread all over the world. Moreover, we are well aware that there are concerns of between-center variation in imaging results which might significantly impact on the reproducibility of the model results.

Third, our BRCA 1/2 testing was mainly performed on fresh frozen ovarian cancer tissue. We assumed that all other areas of the disease within the same patient shared the same mutational status but this consideration may not be entirely correct. given the significant EOC heterogeneity.

Fourth, our patients were only screened for BRCA 1/2: no other genes involved in the HRD pathways were included in the analysis. Therefore, we cannot exclude the presence of mutations in the other genes whose mutational status could correlate with imaging features typical of mutated patients. Moreover, any type of BRCA 1/2 pathogenic variant was labeled as "mutated". There are not enough data to establish whether different mutations produce different downstream effects but we cannot rule out differences in phenotype. This might have affected our analysis.

Fifth, the analysis was carried out using open source pipelines such as the CLAM model which are not customized for the purpose of the study. Entirely in-house designed pipelines, tailored on genomic status identification, might improve final results.

Overall, the model could provide a critical information at the very beginning of the diagnostic process and, if proven effective, tailor further genomic testing (e.g., only BRCA testing or HRD testing) and optimizing genomic testing resources.



(a)



(b)

Figure 6. Cont.



(c)

Figure 6. (a). Heatmap of the held-out BRCA mutated image. Red area represent areas with higher model activation, i.e., areas where the model recognizes pattern associated to the somatic BRCA mutation. (b). Patch level top 5 highest attention patterns (WSI). (c). Patch level top 5 highest attention patterns (ROI on WSI).

Our Results in the Context of Other Observations

Preliminary but encouraging results have been published in the last 3 years on computational pathology.

In 2020, Jang and colleagues showed that APC, KRAS, PIK3CA, SMAD4, and TP53 mutations can be predicted from H&E pathology images using deep learning-based classifiers [40]. The AUCs for ROC curves ranged from 0.693 to 0.809 for frozen WSIs and from 0.645 to 0.783 for the FFPE WSIs.

Xu et al., developed a deep learning model to accurately classify chromosomal instability status on a cohort of 1010 patients with breast cancer (Training set: n = 858, Test set: n = 152) from The Cancer Genome Atlas achieving, an area under the curve of 0.822 with 81.2% sensitivity and 68.7% specificity in the test set. Patch-level predictions of chromosomal instability status suggested intra-tumor heterogeneity within slides [41].

Fu et al. in 2020 quantified histopathological patterns across 17,396 H&E stained histopathology slide images from 28 cancer types and successfully correlate these with matched genomic, transcriptomic and survival data [5].

Moreover, computational histopathology highlighted prognostically relevant areas, such as necrosis or lymphocytic aggregates. The authors underlined the remarkable potential of computer vision in characterizing the molecular basis of tumor histopathology.

Kiehl et al. developed a deep learning model from routine histological slides and/or clinical data to predict lymph node metastasis in colorectal cancer [42]. The deep learning model achieved an AUROC of 71.0%, the clinical classifier achieved an AUROC of 67.0% and a combination of the two classifiers yielded an improvement to 74.1%.

Finally, Wang et al. trained a deep convolutional neural network of ResNet on WSIs to predict the gBRCA mutation in breast cancer [43]. One hundred and sixty six images were combined from two different datasets. The model reached in the external validation dataset an AUC of 0.766 (0.763–0.769) at $40 \times$ magnification. The authors reported the role of histological grade on the accuracy of the prediction.

It has also to be acknowledged that new data are emerging regarding the relevance of BRCA status in the upfront surgical treatment. Not only BRCA WT OC patients seems to benefit more that BRCA mutated ones from hyperthermic intraperitoneal chemotherapy performed at primary debulking surgery but even a neoadjuvant chemotherapy approach has been supposed to be less detrimental in patients harboring BRCA mutation [44–46]. If these data are confirmed, the turnaround time of BRCA status acquisition will became

crucial. Artificial intelligence applied to digital pathology holds much promise in bringing innovative solutions to this possible future clinical unmet need.

4. Materials and Methods

4.1. Patients and Study Design

This is an observational study with patients retrospectively enrolled at the Fondazione Policlinico Universitario "Agostino Gemelli" IRCCS of Rome, Italy, from November 2016 to November 2020.

A weakly supervised deep learning-based model on H&E in EOC patients was set up for BRCA1/2 status prediction.

The retrospective data on BRCA1/2 testing performed on patients with NGS technique was considered as the reference standard of the computational pathology analysis. H&E slides were prepared by a technician and evaluated by a dedicated pathologist, according to current international indications.

In a second step, clinical and follow-up data including therapeutic regimens, progression free survival (PFS) and overall survival (OS), were considered as outcomes to be predicted.

This study was conducted in accordance with the declaration of Helsinki and was approved by the Ethical committee of Fondazione Policlinico Universitario Agostino Gemelli IRCCS (Prot.; 001134''3/21; ID: 3894, 25 March 2021), with the requirement for informed consent. The research was founded by the Italian Ministry of Health providing Institutional Financial Support 5×1000 (2020).

4.2. Study Population

Eligible population includes: (i) women affected by EOC, with known somatic BRCA 1/2 mutational profile and (ii) available Formalin-Fixed Paraffin-Embedded peritoneal tissue sample, collected at the time of first diagnosis of EOC with at least 30% of cancer cells.

For the second step only those patients for whom we had complete follow-up information (minimum follow up 18 months) were included.

The exclusion criteria were: (i) patients affected by recurrent ovarian cancer; (ii) samples collected after chemotherapy; (iii) patients with extra-ovarian tumors with metastases to ovaries; (iv) patients with history of other malignancies in the past 5 years; (v) patients who received any type of target therapy prior to EOC diagnosis.

All the enrolled women were required to sign written informed consent.

Standardized procedures according to previously published workflows were observed to achieve somatic BRCA 1/2 genes mutational status [47–49].

4.3. Deep Learning Approach (CLAM)

CLAM is a deep-learning-based weakly supervised method that uses attention-based learning to automatically identify sub regions of high diagnostic value to accurately classify the whole slide, while also enabling the use of instance-level clustering over the representative regions identified to constrain and refine the feature space.

CLAM is publicly available as a Python package over GitHub (https://github.com/ mahmoodlab/CLAM, accessed on 29 August 2022) [50].

For whole-slide-level learning without annotation, CLAM uses an attention-based pooling function [51] to aggregate patch-level features into slide-level representations for classification. At a high level, during both training and inference, the model examines and ranks all patches in the tissue regions of a WSI, assigning an attention score to each patch, which informs its contribution or importance to the collective slide-level representation for a specific class.

This interpretation of the attention score is reflected in the slide-level aggregation rule of attention-based pooling, which computes the slide-level representation as the average of all patches in the slide weighted by their respective attention score. Unlike the standard MIL algorithm [45,46], which was designed and widely used for weakly

supervised positive/negative binary classification (for example, cancer versus normal), CLAM is designed to solve generic multi-class classification problems. A CLAM model has N parallel attention branches that together calculate N unique slide-level representations, where each representation is determined from a different set of highly attended regions in the image viewed by the network as strong positive evidence for the one of N classes in a multi-class diagnostic task. Each class-specific slide representation is then examined by a classification layer to obtain the final probability score predictions for the whole slide.

The slide-level ground-truth label and the attention scores predicted by the network can be used to generate pseudo labels for both highly and weakly attended patches as a technique to increase the supervisory signals for learning a separable patch-level feature space. During training, the network learns from an additional supervised learning task of separating the most- and least-attended patches of each class into distinct clusters. In addition, it is possible to incorporate domain knowledge into the instance-level clustering to add further supervision.

The pipeline provided by Lu et al., first automatically segments the tissue region of each slide and divides it into many smaller patches, so that they can serve as direct inputs to a CNN. Next, using a CNN for feature extraction, the tool converts all tissue patches into sets of low-dimensional feature embeddings. Following this feature extraction, both training and inference can occur in the low-dimensional feature space instead of the high-dimensional pixel space. The volume of the data space is decreased nearly 200-fold, drastically reducing the subsequent computation required to train supervised deep-learning models.

4.4. WSI Processing

4.4.1. Scanning

Clinical slides were reviewed with the supervision of a dedicated pathologist and selected hematoxylin and eosin-stained slides containing tumor were scanned using the C13220-31 NanoZoomer S360 (Hamamatsu, Japan).

Each slide was scanned using a $40 \times$ objective lens (scanning resolution 0.23 μ M/pixel) of the NanoZoomer and the slide code details, the scanning area and the number of focus points for each slide were determined by the user. The number of focal points was approximately 15 focus points per slide.

Place the glass slides in the cassettes and set them in the holder of the machine, each slide was automatically handled and scanned. The approximate time taken to scan the image of the whole slide per case was up to 1 min.

Scanned images in their proprietary NDP Image (NDPI) file format were checked for the whole and details of the tissues using the NDP.view2 image viewing software for NanoZoomer on a desktop computer with a high-definition resolution screen (1920 \times 1080 pixels). NDPI stores an image pyramid in TIFF directory entries.

Images and data were stored and exported to an external storage device.

4.4.2. Segmentation

The first step is an automated segmentation of the tissue regions. The first step focuses on segmenting the tissue and excluding any holes. The segmentation of specific slides can be adjusted by tuning the individual parameters. The pipeline input is digitized whole slide image data in well-known standard formats (.svs, .ndpi, .tiff etc.). The WSI is read into memory at a down sampled resolution, converted from RGB to the HSV color space. A binary mask for the tissue regions (foreground) is computed based on thresholding the saturation channel of the image after median blurring to smooth the edges and is followed by additional morphological closing to fill small gaps and holes [52]. The approximate contours of the detected foreground objects are then filtered based on an area threshold and stored for downstream processing while the segmentation mask for each slide is made available for optional visual inspection. A human-readable text-file is also automatically generated, which includes the list of files processed along with editable fields containing the set of key segmentation parameters used.

4.4.3. Patching

After segmentation, the background is removed from images for each slide and the remaining pixels are grouped into a grid of smaller images (256×256 patches) from within the segmented foreground contours at the user-specified magnification and stores stacks of image patches along with their coordinates and the slide metadata using the Hierarchical Data Format version 5 (HDF5).

This is not a computationally intensive process and is dependent on resolution level.

The number of patches extracted from each slide can range from hundreds (biopsy slide patched at $\times 20$ magnification) to hundreds of thousands (large resection slide patched at $\times 40$ magnification). Output is a representation of images through patches in a high dimensional feature (HDF) space generated by a pre-trained convolutional neural network.

4.4.4. Feature Extraction

Following patching, we used the pre-trained ResNet50 model [53] already embedded in CLAM to compute a low-dimensional feature representation for each image patch of each slide. Features extraction from patches is a computationally intensive step. Requiring about one minute per whole slide image on NVIDIA Quadro RTX 5000.

4.4.5. Attention Branch

Each patch feature vector then enters an attention network which is trained to recognize patterns often associated with a particular Slide-level label (over-simplification).

At the end of training, the overall model should be able to identify characteristic regions of activation and make classification at the whole slide image level.

4.5. Statistical Analysis

The sample size available for the analysis consisted of 664 histological images. For computational reasons, the pipeline was first applied to about 30% of available data (171 images) and then incremented to 298 in order to measure the performance gain due to increased sample size.

The dataset was split into a training and validation set, respectively 80% and 20% of the considered number of samples.

Classification performance of the predictive model was monitored during training with ROC AUC, negative and positive predictive value on the validation set. The final model was then applied to held out images to generate activation maps on unseen BRCA mutated images. The model was trained for 300 epochs on a NVDIA Quadro RTX 5000. All statistical analysis was performed in Python version 3.7.4.

5. Conclusions

Our results confirm that models applied to H&E slides cannot yet match the performance level of the gold standards thus their use in current clinical practice cannot be advocated. Nevertheless, its potentiality as a screening tool for personalization and optimization of genomic testing warrants further investigation.

From a clinical point of view, the information obtained directly from frozen section H&E slides, may give clinicians a crucial information at an early stage of the therapeutic decision making process that could be integrated with already validated clinical scores or multiomics translational approaches [54–56].

For all these reasons, we believe that further developments are well worth carrying out. We have already planned to enlarge the dataset collecting cases from December 2020 to August 2022. Moreover, we are working on improving data input for CLAM analysis by using a recently released segmentation model code (https://github.com/MSKCC-Computational-Pathology/DMMN-ovary, accessed on 29 August 2022) [57]. On the other
hand, we are exploring in collaboration with other groups, new approaches such as the development of a persistent homology-based model on the same dataset [58].

Future studies must include data across multiple centers not used in the model training to demonstrate high accuracy and reproducibility. A deeper involvement of pathologists should be pursued in order to achieve the finest tuning possible according to well recognized features.

Author Contributions: C.N., L.B., J.L., M.T.G., A.P., A.F., G.Z., V.V. and G.S. were responsible for the conceptualization of the study design. C.N., M.T.G. and A.P. were responsible for sample collection and C.N., M.T.G., A.P., F.I., T.P. and A.M. for data collection. C.N., J.L. and M.T.G. were responsible for drafting of the manuscript. C.N. and J.L. were responsible for the formal data analysis. The underlying data reported in the manuscript has been accessed and verified by multiple authors (C.N., L.B., J.L., M.T.G., F.I., A.P. and A.M.). All authors have read and agreed to the published version of the manuscript.

Funding: The research was founded by the Italian Ministry of Health providing Institutional Financial Support 5×1000 (2020).

Institutional Review Board Statement: Prot.; 001134"3/21; ID: 3894, 25 March 2021.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: We would like to acknowledge the contribution of Nikon for the free trial of Nanozoomer S360, G-STeP Immunohistochemistry, Data collection and Radiomics core facilities (https://gstep.policlinicogemelli.it/#/, accessed on 29 August 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Konstantinopoulos, P.A.; Ceccaldi, R.; Shapiro, G.I.; D'Andrea, A.D. Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer. *Cancer Discov.* **2015**, *5*, 1137–1154. [PubMed]
- Alsop, K.; Fereday, S.; Meldrum, C.; deFazio, A.; Emmanuel, C.; George, J.; Dobrovic, A.; Birrer, M.J.; Webb, P.M.; Stewart, C.; et al. BRCA mutation frequency and patterns of treatment response in BRCA mutation-positive women with ovarian cancer: A report from the Australian Ovarian Cancer Study Group. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 2012, 30, 2654–2663. [CrossRef] [PubMed]
- Pal, T.; Permuth-Wey, J.; Betts, J.A.; Krischer, J.P.; Fiorica, J.; Arango, H.; LaPolla, J.; Hoffman, M.; Martino, M.A.; Wakeley, K.; et al. BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. *Cancer* 2005, 104, 2807–2816. [CrossRef] [PubMed]
- 4. Hennessy, B.T.; Timms, K.M.; Carey, M.S.; Gutin, A.; Meyer, L.A.; Flake II, D.D.; Abkevich, V.; Potter, J.; Pruss, D.; Glenn, P.; et al. Somatic mutations in BRCA1 and BRCA2 could expand the number of patients that benefit from poly (ADP ribose) polymerase inhibitors in ovarian cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 2010, 28, 3570–3576. [CrossRef]
- Fu, Y.; Jung, A.W.; Torne, R.V.; Gonzalez, S.; Vöhringer, H.; Shmatko, A.; Yates, L.; Jimenez-Linan, M.; Moore, L.; Gerstung, M. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer* 2020, *1*, 800–810. [CrossRef]
- 6. Ninomiya, H.; Hiramatsu, M.; Inamura, K.; Nomura, K.; Okui, M.; Miyoshi, T.; Okumura, S.; Satoh, Y.; Nakagawa, K.; Nishio, M.; et al. Correlation between morphology and EGFR mutations in lung adenocarcinomas: Significance of the micropapillary pattern and the hobnail cell type. *Lung Cancer* **2009**, *63*, 235–240. [CrossRef]
- Warth, A.; Penzel, R.; Lindenmaier, H.; Brandt, R.; Stenzinger, A.; Herpel, E.; Goeppert, B.; Thomas, M.; Herth, F.J.F.; Dienemann, H.; et al. EGFR, KRAS, BRAF and ALK gene alterations in lung adenocarcinomas: Patient outcome, interplay with morphology and immunophenotype. *Eur. Respir. J.* 2014, *43*, 872–883. [CrossRef]
- 8. Mosquera, J.M.; Perner, S.; Demichelis, F.; Kim, R.; Hofer, M.D.; Mertz, K.D.; Paris, P.L.; Simko, J.; Collins, C.; Bismar, T.A.; et al. Morphological features of TMPRSS2-ERG gene fusion prostate cancer. *J. Pathol.* **2007**, *212*, 91–101. [CrossRef]
- Hakimi, A.A.; Tickoo, S.K.; Jacobsen, A.; Sarungbam, J.; Sfakianos, J.P.; Sato, Y.; Morikawa, T.; Kume, H.; Fukayama, M.; Homma, Y.; et al. TCEB1-mutated renal cell carcinoma: A distinct genomic and morphological subtype. *Mod. Pathol.* 2015, 28, 845–853. [CrossRef]
- Weisman, P.S.; Ng, C.K.; Brogi, E.; Eisenberg, R.E.; Won, H.H.; Piscuoglio, S.; De Filippo, M.R.; Ioris, R.; Akram, M.; Norton, L.; et al. Genetic alterations of triple negative breast cancer by targeted next-generation sequencing and correlation with tumor morphology. *Mod. Pathol.* 2016, 29, 476–488. [CrossRef]
- 11. Soslow, R.A.; Han, G.; Park, K.J.; Garg, K.; Olvera, N.; Spriggs, D.R.; Kauff, N.D.; Levine, D.A. Morphologic patterns associated with BRCA1 and BRCA2 genotype in ovarian carcinoma. *Mod. Pathol.* **2012**, *25*, 625–636. [CrossRef] [PubMed]

- 12. Bera, K.; Schalper, K.A.; Madabhushi, A. Artificial intelligence in digital pathology-new tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **2019**, *16*, 703–715. [CrossRef] [PubMed]
- 13. Niazi, M.K.K.; Parwani, A.V.; Gurcan, M.N. Digital pathology and artificial intelligence. *Lancet Oncol.* **2019**, *20*, e253–e261. [CrossRef]
- Hollon, T.C.; Pandian, B.; Adapa, A.R.; Urias, E.; Save, A.V.; Khalsa, S.S.S.; Eichberg, D.G.; D'Amico, R.S.; Farooq, Z.U.; Lewis, S.; et al. Near real-time intraoperative brain tumor diagnosis using stimulated raman histology and deep neural networks. *Nat. Med.* 2020, 26, 52–58. [CrossRef]
- Kather, J.N.; Pearson, A.T.; Halama, N.; Jager, D.; Krause, J.; Loosen, S.H.; Marx, A.; Boor, P.; Tacke, F.; Neumann, U.P.; et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* 2019, 25, 1054–1056. [CrossRef] [PubMed]
- Bulten, W.; Pinckaers, H.; van Boven, H.; Vink, R.; de Bel, T.; van Ginneken, B.; van der Laak, J.; Hulsbergen-van de Kaa, C.A.; Litjens, G. Automated deep-learning system for gleason grading of prostate cancer using biopsies: A diagnostic study. *Lancet* Oncol. 2020, 21, 233–241. [CrossRef]
- Ström, P.; Kartasalo, K.; Olsson, H.; Solorzano, L.; Delahunt, B.; Berney, D.M.; Bostwick, D.G.; Evans, A.; Grignon, D.; Humphrey, P.A.; et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: A population-based, diagnostic study. *Lancet Oncol.* 2020, 21, 222–232. [CrossRef]
- Schapiro, D.; Jackson, H.W.; Raghuraman, S.; Fischer, J.R.; Zanotelli, V.R.T.; Schulz, D.; Giesen, C.; Catena, R.; Varga, Z.; Bodenmiller, B. HistoCAT: Analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods* 2017, 14, 873–876. [CrossRef]
- 19. Moen, E.; Bannon, D.; Kudo, T.; Graf, W.; Covert, M.; van Valen, M. Deep learning for cellular image analysis. *Nat. Methods* **2019**, *16*, 1233–1246. [CrossRef]
- 20. Mahmood, F.; Borders, D.; Chen, R.; McKay, G.N.; Salimian, K.J.; Baras, A.; Durr, N.J. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Trans. Med. Imaging* **2019**, *39*, 3257–3267. [CrossRef]
- 21. Graham, S.; Dang Vu, S.; Ahmed Raza, S.E.; Azam, A.; Tsang, Y.W.; Kwak, J.T.; Rajpoot, N. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* **2019**, *58*, 101563. [CrossRef] [PubMed]
- Saltz, J.; Gupta, R.; Hou, L.; Kurk, T.; Singh, P.; Nguyen, V.; Samaras, D.; Shroyer, K.R.; Zhao, T.; Batiste, R.; et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* 2018, 23, 181–193. [CrossRef] [PubMed]
- Javed, S.; Mahmood, A.; Fraz, M.M.; Koohbanani, N.A.; Benes, K.; Tsang, Y.W.; Hewitt, K.; Epstein, D.; Snead, D.; Rajpoot, N. Cellular community detection for tissue phenotyping in colorectal cancer histology images. *Med. Image Anal.* 2020, 63, 101696. [CrossRef] [PubMed]
- Mobadersany, P.; Yousefia, S.; Amgada, M.; Gutmanb, D.A.; Barnholtz-Sloanc, J.S.; Velázquez Vega, J.E.; Brat, D.J.; Cooper, L.A.D. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. USA* 2018, 115, E2970–E2979. [CrossRef]
- Heindl, A.; Khan, A.M.; Rodrigues, D.N.; Eason, K.; Sadanandam, A.; Orbegoso Aguilar, C.; Punta, M.; Sottoriva, A.; Lise, S.; Banerjee, S. Microenvironmental niche divergence shapes brca1-dysregulated ovarian cancer morphological plasticity. *Nat. Commun.* 2018, *9*, 3917. [CrossRef]
- Yuan, Y.; Failmezger, H.; Rueda, O.M.; Ali, H.R.; Gräf, S.; Chin, S.F.; Schwarz, R.F.; Curtis, C.; Dunning, M.J.; Bardwell, H.; et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci. Transl. Med.* 2012, *4*, 157ra143. [CrossRef]
- Lazar, A.J.; McLellan, M.D.; Bailey, M.H.; Miller, C.A.; Appelbaum, E.L.; Cordes, M.G.; Fronick, C.C.; Fulton, L.A.; Fulton, R.S.; Mardis, E.R.; et al. Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* 2017, 171, 950–965.
- Kather, J.N.; Heij, L.R.; Grabsch, H.I.; Kooreman, L.F.S.; Loeffler, C.; Echle, A.; Krause, J.; Muti, H.S.; Niehues, J.M.; Sommer, K.A.J.; et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Cancer* 2020, 1, 789–799. [CrossRef]
- 29. Chen, R.J.; Lu, M.Y.; Wang, J.; Williamson, D.F.K.; Rodig, S.J.; Lindeman, N.I.; Mahmood, F. Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans. Med. Imaging* **2020**, 41, 757–770. [CrossRef]
- Beck, A.H.; Sangoi, A.R.; Leung, S.; Marinelli, R.J.; Nielsen, T.O.; van de Vijver, M.J.; West, R.B.; van de Rijn, M.; Koller, D. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. Transl. Med.* 2011, 3, 108ra113. [CrossRef]
- Yamamoto, Y.; Tsuzuki, T.; Akatsuka, J.; Ueki, M.; Morikawa, H.; Numata, Y.; Takahara, T.; Tsuyuki, T.; Tsutsumi, K.; Nakazawa, R. Automated acquisition of explainable knowledge from unannotated histopathology images. *Nat. Commun.* 2019, 10, 5642. [CrossRef] [PubMed]
- 32. Pell, R.; Oien, K.; Robinson, M.; Pitman, H.; Rajpoot, N.; Rittscher, J.; Snead, D.; Verrill, C. The use of digital pathology and image analysis in clinical trials. *J. Pathol. Clin. Res.* **2019**, *5*, 81–90. [CrossRef] [PubMed]
- Bejnordi, B.E.; Veta, M.; van Diest, P.J.; van Ginneken, B.; Karssemeijer, N.; Litjens, G.; van der Laak, J.A.W.M. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017, 318, 2199–2210. [CrossRef] [PubMed]

- Chen, P.H.C.; Gadepalli, K.; MacDonald, R.; Liu, Y.; Kadowaki, S.; Nagpal, K.; Kohlberger, T.; Dean, J.; Corrado, G.S.; Hipp, J.D.; et al. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. *Nat. Med.* 2019, 25, 1453–1457. [CrossRef]
- Nagpal, K.; Foote, F.; Tan, F.; Liu, Y.; Chen, P.H.C.; Steiner, D.F.; Manoj, N.; Olson, N.; Smith, J.L.; Mohtashamian, A.; et al. Development and validation of a deep learning algorithm for improving gleason scoring of prostate cancer. *npj Digit. Med.* 2019, 2, 48. [CrossRef]
- Wang, S.; Zhu, Y.; Yu, L.; Chen, H.; Lin, H.; Wan, X.; Fan, X.; Heng, P.-A. RMDL: Recalibrated multi-instance deep learning for whole slide gastric image classification. *Med. Image Anal.* 2019, 58, 101549. [CrossRef]
- Coudray, N.; Ocampo, P.S.; Sakellaropoulos, T.; Navneet, N.; Matija, S.; Fenyo, D.; Moreira, A.L.; Narges, R.; Razavian, N.; Tsirigos, A. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* 2018, 24, 1559–1567. [CrossRef]
- Campanella, G.; Hanna, M.G.; Geneslaw, L.; Miraflor, A.P.; Werneck Krauss Silva, V.; Busam, K.J.; Brogi, E.; Reuter, V.; Klimstra, D.S.; Fuchs, T.J. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* 2019, 25, 1301–1309. [CrossRef]
- 39. Ming, Y.L.; Williamson, D.F.K.; Chen, T.Y.; Chen, R.J.; Barbieri, M.; Mahmood, F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **2021**, *5*, 555–570.
- 40. Jang, H.J.; Lee, A.; Kang, J.; Song, I.H.; Lee, S.H. Prediction of clinically actionable genetic alterations from colorectal cancer histopathology images using deep learning. *World J. Gastroenterol.* **2020**, *26*, 6207–6223. [CrossRef]
- 41. Xu, Z.; Verma, A.; Naveed, U.; Bakhoum, S.F.; Khosravi, P.; Elemento, O. Deep learning predicts chromosomal instability from histopathology images. *iScience* 2021, 24, 102394. [CrossRef] [PubMed]
- Kiehl, L.; Kuntz, S.; Höhn, J.; Jutzi, T.; Krieghoff-Henning, E.; Kather, J.N.; Holland-Letz, T.; Kopp-Schneider, A.; Chang-Claude, J.; Brobeil, A.; et al. Deep learning can predict lymph node status directly from histology in colorectal cancer. *Eur. J. Cancer* 2021, 157, 464–473. [CrossRef] [PubMed]
- Wang, X.; Zou, C.; Zhang, Y.; Li, X.; Wang, C.; Ke, F.; Chen, J.; Wang, W.; Wang, D.; Xu, X.; et al. Prediction of BRCA Gene Mutation in Breast Cancer Based on Deep Learning and Histopathology Images. *Front. Genet.* 2021, 12, 661109. [CrossRef] [PubMed]
- 44. Ghirardi, V.; De Felice, F.; D'Indinosante, M.; Bernardini, F.; Giudice, M.T.; Fagotti, A.; Scambia, G. Hyperthermic intraperitoneal chemotherapy (HIPEC) after primary debulking surgery in advanced epithelial ovarian cancer: Is BRCA mutational status making the difference? *Cancer Treat. Res. Commun.* **2022**, *31*, 100518. [CrossRef] [PubMed]
- Petrillo, M.; Marchetti, C.; De Leo, R.; Musella, A.; Capoluongo, E.; Paris, I.; Benedetti Panici, P.; Scambia, G.; Fagotti, A. BRCA mutational status, initial disease presentation, and clinical outcome in high-grade serous advanced ovarian cancer: A multicenter study. *Am. J. Obstet. Gynecol.* 2017, 217, 334.e1–334.e9. [CrossRef]
- Mahdi, H.; Gockley, A.; Esselen, K.; Marquard, J.; Nutter, B.; Yang, B.; Hinchcliff, E.; Horowitz, N.; Rose, P.G. Outcome of neoadjuvant chemotherapy in BRCA1/2 mutation positive women with advanced-stage Müllerian cancer. *Gynecol. Oncol.* 2015, 139, 407–412. [CrossRef]
- Marchetti, C.; Minucci, A.; D'Indinosante, M.; Ergasti, R.; Arcieri, M.; Capoluongo, E.D.; Pietragalla, A.; Caricato, C.; Scambia, G.; Fagotti, A. Feasibility of tumor testing for BRCA status in high-grade serous ovarian cancer using fresh-frozen tissue based approach. *Gynecol. Oncol.* 2020, 158, 740–746. [CrossRef]
- 48. Costella, A.; De Leo, R.; Guarino, D.; D'Indinosante, M.; Concolino, P.; Mazzuccato, G.; Urbani, A.; Scambia, G.; Capoluongo, E.; Fagotti, A.; et al. High-resolution melting analysis coupled with next-generation sequencing as a simple tool for the identification of a novel somatic BRCA2 variant: A case report. *Hum. Genome Var.* 2018, *5*, 10. [CrossRef]
- Concolino, P.; Rizza, R.; Mignone, F.; Costella, A.; Guarino, D.; Carboni, I.; Capoluongo, E.; Santonocito, C.; Urbani, A.; Minucci, A. A comprehensive BRCA1/2 NGS pipeline for an immediate Copy Number Variation (CNV) detection in breast and ovarian cancer molecular diagnosis. *Clin. Chim. Acta* 2018, 480, 173–179. [CrossRef]
- 50. Data-Efficient and Weakly Supervised Computational Pathology on Whole Slide Images. Available online: https://github.com/ mahmoodlab/CLAM (accessed on 29 August 2022).
- 51. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual attention network for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
- 52. Maragos, P. Tutorial on advances in morphological image processing and analysis. Opt. Eng. 1987, 26.7, 623–632. [CrossRef]
- 53. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
- De Maria Marchiano, R.; Di Sante, G.; Piro, G.; Carbone, C.; Tortora, G.; Boldrini, L.; Pietragalla, A.; Daniele, G.; Tredicine, M.; Cesario, A.; et al. Translational Research in the Era of Precision Medicine: Where We Are and Where We Will Go. *J. Pers. Med.* 2021, 11, 216. [CrossRef] [PubMed]
- 55. Fagotti, A.; Ferrandina, G.; Fanfani, F.; Ercoli, A.; Lorusso, D.; Rossi, M.; Scambia, G. A laparoscopy-based score to predict surgical outcome in patients with advanced ovarian carcinoma: A pilot study. *Ann. Surg. Oncol.* **2006**, *13*, 1156–1161. [CrossRef] [PubMed]
- Vizzielli, G.; Costantini, B.; Tortorella, L.; Pitruzzella, I.; Gallotta, V.; Fanfani, F.; Gueli Alletti, S.; Cosentino, F.; Nero, C.; Scambia, G.; et al. A laparoscopic risk-adjusted model to predict major complications after primary debulking surgery in ovarian cancer: A single-institution assessment. *Gynecol. Oncol.* 2016, 142, 19–24. [CrossRef] [PubMed]

- Ovarian Cancer Segmentation Using Deep Multi-Magnification Network. Available online: https://github.com/MSKCC-57. Computational-Pathology/DMMN-ovary (accessed on 29 August 2022). Herbert, E.; Harer, J.L. *Computational Topology: An Introduction;* American Mathematical Society: Providence, RI, USA, 2010;
- 58. ISBN 978-0-8218-4925-5.





Article **Target of Rapamycin Regulates Photosynthesis and Cell Growth in** *Auxenochlorella pyrenoidosa*

Tingting Zhu^{1,2,†}, Linxuan Li^{1,2,3,†}, Huimin Chang³, Jiasui Zhan^{4,*} and Maozhi Ren^{1,2,3,*}

- ¹ Institute of Urban Agriculture, Chinese Academy of Agricultural Sciences, Chengdu National Agricultural Science and Technology Center, Chengdu 610213, China
- ² Hainan Yazhou Bay Seed Laboratory, Sanya 572025, China
- ³ School of Agricultural Science, Zhengzhou University, Zhengzhou 450001, China
- ⁴ Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden
- * Correspondence: jiasui.zhan@slu.se (J.Z.); renmaozhi01@caas.cn (M.R.)
- + These authors contributed equally to this work.

Abstract: Auxenochlorella pyrenoidosa is an efficient photosynthetic microalga with autotrophic growth and reproduction, which has the advantages of rich nutrition and high protein content. Target of rapamycin (TOR) is a conserved protein kinase in eukaryotes both structurally and functionally, but little is known about the TOR signalling in Auxenochlorella pyrenoidosa. Here, we found a conserved ApTOR protein in Auxenochlorella pyrenoidosa, and the key components of TOR complex 1 (TORC1) were present, while the components RICTOR and SIN1 of the TORC2 were absent in Auxenochlorella pyrenoidosa. Drug sensitivity experiments showed that AZD8055 could effectively inhibit the growth of Auxenochlorella pyrenoidosa, whereas rapamycin, Torin1 and KU0063794 had no obvious effect on the growth of Auxenochlorella pyrenoidosa. Transcriptome data results indicated that Auxenochlorella pyrenoidosa TOR (ApTOR) regulates various intracellular metabolism and signaling pathways in Auxenochlorella pyrenoidosa. Most genes related to chloroplast development and photosynthesis were significantly down-regulated under ApTOR inhibition by AZD8055. In addition, ApTOR was involved in regulating protein synthesis and catabolism by multiple metabolic pathways in Auxenochlorella pyrenoidosa. Importantly, the inhibition of ApTOR by AZD8055 disrupted the normal carbon and nitrogen metabolism, protein and fatty acid metabolism, and TCA cycle of Auxenochlorella pyrenoidosa cells, thus inhibiting the growth of Auxenochlorella pyrenoidosa. These RNA-seq results indicated that ApTOR plays important roles in photosynthesis, intracellular metabolism and cell growth, and provided some insights into the function of ApTOR in Auxenochlorella pyrenoidosa.

Keywords: TOR; photosynthesis; cell growth; AZD8055; Auxenochlorella pyrenoidosa

1. Introduction

Chlorella is a unicellular eukaryotic green alga that emerged 2 billion years ago and is a high-efficiency primary producer in ecosystems [1]. *Chlorella* can be cultured in a natural environment or in controllable closed systems, with higher productivity than most plants. For a long time, *Chlorella* has been deemed as a source of protein and fat, and it is used as a raw material for human food and animal feed [2,3]. Like land plants, *Chlorella* also performs photosynthesis via chloroplast, converting solar energy into chemical energy that is vital to its development and generates oxygen. *Chlorella* contains many high-value substances such as protein, pigment, antioxidants, vitamins, minerals and cell growth factor, and has been referred to as "the best genetic food in the 21st century" by the World Health Organization [4]. At present, over 10 species of *Chlorella* have been identified in the world [5,6], among which *Auxenochlorella pyrenoidosa* (*A. pyrenoidosa*, formerly *Chlorella pyrenoidosa*) has attracted much attention because it is edible and its protein content can account for more than 50% of dry weight [7]. The genome size of

Citation: Zhu, T.; Li, L.; Chang, H.; Zhan, J.; Ren, M. Target of Rapamycin Regulates Photosynthesis and Cell Growth in *Auxenochlorella pyrenoidosa. Int. J. Mol. Sci.* 2022, 23, 11309. https://doi.org/10.3390/ ijms231911309

Academic Editor: Wajid Zaman

Received: 2 August 2022 Accepted: 21 September 2022 Published: 25 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). *A. pyrenoidosa* FACHB-9 is 56.6 Mbp, including 10,284 genes [8]. An analysis of the genome structure provides a foundation for improving *A. pyrenoidosa* production as food and fuels. Furthermore, *A. pyrenoidosa* has been widely used in wastewater treatment, especially in high-concentration inorganic industrial wastewater [9]. Utilizing high ammonium salts in industrial wastewater can produce up to 56.7% (dry weight) protein in *A. pyrenoidosa*, and 95% of ammonium salt could be consumed [10]. Additionally, *A. pyrenoidosa* also utilizes organic carbon source and nitrogen source for high-density heterotrophic growth, with production efficiency being over ten times higher than that of autotrophic growth [8]. However, chloroplast was degraded, lipid content was increased, and protein synthesis was inhibited in heterotrophic *A. pyrenoidosa* [8,11,12]. Genomic and transcriptomic sequencing results showed that heterotrophic to photoautotrophic of *A. pyrenoidosa* resulted in global metabolic reprogramming [8].

Target of rapamycin (TOR) is a core regulatory factor for eukaryotic growth and development, which coordinates cell proliferation, growth and metabolism [13,14]. TOR protein has highly conserved structures, including N-terminal HEAT repeats, FAT, FRB, catalytic and C-terminal FATC domains [15]. TOR protein and other proteins form TOR complex 1 (TORC1) and TORC2 in yeast and mammals [16,17]. However, there was only the conserved and functional TORC1 in plants, and TORC1 was composed of TOR, regulatory associate the protein target of rapamycin (RAPTOR) and lethal with sec-13 protein 8 (LST8) [18,19]. TORC2 core proteins RICTOR and SIN1 seem to be missing in photosynthetic eukaryotes, including plants and green algae [20,21].

Rapamycin is a macrolide immunosuppressant from the bacterium *Streptomyces hygroscopicus*. It binds to the 12 kDa FK506 binding protein (FKBP12) and the FRB domain of TOR, thereby inhibiting the activity of TOR protein [22]. Loss of FKBP12 function prevents rapamycin from inhibiting the TOR protein in most plants [23–26]. Fortunately, TOR kinase inhibitors such as Torin1, AZD8055 and KU0063794 from mammals were developed and applied in plants, and have also been proved to specifically and efficiently inhibit TOR kinase activity in plants [15,27,28]. With the help of TOR inhibitors and various omics research methods, animal and plant conserved and plant-specific TOR signaling pathways have been revealed [18,29]. In plants, TOR regulates cell division and elongation, protein synthesis, nutrient and metabolism, and stress response by integrating multiple exogenous environmental signals and endogenous physiological signals [15,30–32]. TOR affects plant growth and development from embryogenesis to photomorphogenesis, root and leaf development, flowering, and senescence in plants [13,18,33].

Genomic analysis of some algal species revealed that TORC1 components are widely conserved in algae [34]. Different from other microalgae, the functions of TOR have been comprehensively studied in the model green alga Chlamydomonas reinhardtii (C. reinhardtii) [35]. Previous studies have shown that C. reinhardtii is sensitive to rapamycin, and the rapamycin sensitive TORC1 signaling regulates cell growth, protein synthesis, autophagy, and key metabolism processes in C. reinhardtii [35–37]. In addition, a recent study has shown that TOR controls the carotenoid production by phosphorylating lycopene beta/epsilon cyclase in *C. reinhardtii* [38]. This is the first evidence that TOR directly regulates the biosynthesis of secondary metabolite carotenoid in algae. As an industrial production alga, A. pyrenoidosa has fast growth rate, rich nutrition and high protein. However, the TOR signaling pathway of A. pyrenoidosa has not been reported, and whether TOR signaling regulates the cell growth and protein synthesis of A. pyrenoidosa remains unknown. In this study, homologous sequence alignment revealed that there was only the conserved TORC1 signaling pathway in A. pyrenoidosa. Drug sensitivity experiments showed that AZD8055 could effectively inhibit the growth of A. pyrenoidosa, while rapamycin, Torin1 and KU0063794 had no effect on the growth of A. pyrenoidosa. RNA-seq results showed that most genes involved in photosynthesis were significantly down-regulated in A. pyrenoidosa treated with AZD8055, indicating that ApTOR had an important effect on photosynthesis of A. pyrenoidosa. In addition, DEGs involved in the regulation of autophagy and ubiquitin mediated proteolysis were almost all up-regulated, suggesting that ApTOR was also

involved in regulating autophagy and protein catabolic process of *A. pyrenoidosa*. These results suggested that ApTOR plays major roles in regulating photosynthesis and cellular metabolism in *A. pyrenoidosa*.

2. Results

2.1. Conserved TOR Signaling Pathway in Auxenochlorella Pyrenoidosa

The conserved TORC1 signal regulates intracellular metabolism, nutrient, and energy in C. reinhardtii [35–37]. To investigate the conserved TOR signaling pathway in A. pyrenoidosa, BLASTp analysis was performed on the public transcriptome data of A. pyrenoidosa (NCBI accession number: PRJNA730327). Only one conserved TOR protein (ApTOR) was found in A. pyrenoidosa, with a maximum similarity of 69% to CrTOR protein. ApTOR contains N-terminal HEAT repeats, FAT, FRB, catalytic and FATC domains at C-terminal (Figure 1A). Homologous sequence alignment revealed that the catalytic domain of ApTOR was the most conserved with the highest similarity among species, while the FAT domain had the lowest similarity among species (Figure 1A,C). Phylogenetic tree showed that ApTOR and CrTOR were the most conserved in evolution but had the most distant evolutionary relationship with TpTOR and PtTOR (Figure 1B). Meanwhile, sequence alignment found that RAPTOR and LST8, the key proteins of TORC1, were also present in A. pyrenoidosa (Table 1), whereas RICTOR and SIN1, the key proteins of TORC2, were not found in A. pyrenoidosa. In addition, TORC1 signaling downstream components also existed in A. pyrenoidosa (Table 1). These results showed that there was a conserved TORC1 signaling pathway in A. pyrenoidosa.



Figure 1. A structure and sequence analysis of ApTOR. (**A**) An analysis of the conserved domains of ApTOR protein and homologs from other species. The number denotes the identity (%) of ApTOR domain with homologs from other species. *Chlamydomonas reinhardtii* (Cr) (Chlorophyta), *Arabidopsis thaliana* (At) (Plantae, Magnoliophyta), *Homo sapiens* (Hs) (Animalia, Chordata), *Saccharomyces cerevisiae* (Sc) (Fungi, Ascomycota), *Phaeodactylum tricornutum* (Pt) (Bacillariophyta), *Thalassiosira pseudonana* (Tp) (Bacillariophyta). (**B**) The phylogenetic tree of ApTOR protein and homologs from other species. The phylogenetic tree was constructed by MEGA 4 software using the Neighbor-Joining method. Numbers represent bootstrap percentages (1000 of bootstrap replicates). (**C**) Sequence alignment of the catalytic domains of ApTOR protein and homologs from other species. Red represents identical amino acid sequences, and blue represents more than 75% identical amino acid sequences.

Protein Name	Chlamydomonas reinhardtii	Auxenochlorella pyrenoidosa	Identity (%)
Target of rapamycin (TOR)	CrTOR	ApTOR like	58
Regulatory associate protein of TOR (RAPTOR)	CrRAPTOR	ApRAPTOR like	37
Lethal with SEC-13 protein 8 (LST8)	CrLST8	ApLST8 like	81
FK506-binding protein 12 kDa (FKBP12)	CrFKBP12	ApFKBP12 like	43
Ribosomal protein S6 kinase (S6K)	CrS6K	ApS6K like	45
Ribosome protein small subunit 6 (RPS6)	CrRPS6	ApRPS6 like	78
Transcription factor E2F alpha (E2FA)	CrE2FA	ApE2FA like	39
Translation initiation factor 2 alpha subunit (eIF2 α)	CreIF2a	ApeIF2α like	72
Type 2A phosphatase associated protein 46 (TAP46)	CrTAP46	ApTAP46 like	40
Autophagy protein 1 (ATG1)	CrATG1	ApATG1 like	39
Autophagy protein 13 (ATG13)	CrATG13	ApATG13 like	32

Table 1. The putative components of TOR signaling pathway in Auxenochlorella pyrenoidosa.

2.2. Effects of TOR Inhibitors on the Growth of Auxenochlorella Pyrenoidosa

In order to elucidate function of TOR signal in A. pyrenoidosa, A. pyrenoidosa was treated with rapamycin, a specific inhibitor of TOR protein. The results showed that rapamycin has no obvious effect on the growth of A. pyrenoidosa, even at a higher concentration of rapamycin (Figure 2A), indicating that A. pyrenoidosa is insensitive to rapamycin. Previous studies have shown that rapamycin inhibits the TOR activity by forming a ternary complex with FKBP12 and the FRB domain of TOR [39]. A ApFKBP12 sequence with 43% similarity to CrFKBP12 was found in the transcriptome data of A. pyrenoidosa, encoding 159 amino acids (Table 1). Interestingly, ApFKBP12 protein is evolutionarily closer to rapamycin-sensitive species (Figure 2B). Some amino acids involved in the formation of the FKBP12-rapamycin complex are required for inhibiting TOR activity [40]. We found that the ApFKBP12 protein sequence contains the conserved amino acids required for FKBP12 binding to rapamycin, including Tyr26, Asp38 and Gly89 (numbered according to human HsFKBP12) (Figure 2C). However, there was an additional sequence of 52 amino acids at the N-terminal of ApFKBP12 protein compared with other species (Figure 2C), which may change the function of ApFKBP12 protein and failure in binding to rapamycin. Rapamycin also interacts with the FRB domain of TOR by binding to aromatic residues [40], and sequence alignment revealed that these key amino acids were highly conserved in A. pyrenoidosa and other species (Figure 2D). As the FRB domain of ApTOR is highly conserved, the resistance of *A. pyrenoidosa* to rapamycin may be due to the loss of ApFKBP12 function.

Furthermore, TOR kinase inhibitors AZD8055, Torin1 and KU0063794 were used to treat *A. pyrenoidosa*. The results showed that AZD8055 could effectively inhibit the growth of *A. pyrenoidosa*, while Torin1 and KU0063794 had no effect on the growth of *A. pyrenoidosa* even at higher concentrations (Figure 3). The 50% inhibitory concentration (IC50) of AZD8055 on the growth of *A. pyrenoidosa* was about 1 μ M. When the concentration of AZD8055 reached 5 μ M, the growth of *A. pyrenoidosa* was completely inhibited, indicating that the lethal concentration of AZD8055 may be 5 μ M (Figure 3A,B). However, when AZD8055 was removed from the medium, the inhibited cells resumed growth (Supplementary Figure S1), indicating that inhibition of ApTOR kinase activity by AZD8055 can be applied to elucidate the function of ApTOR in *A. pyrenoidosa*.



Figure 2. *Auxenochlorella pyrenoidosa* is resistant to rapamycin. (**A**) The phenotype of *A. pyrenoidosa* treated with different concentrations of rapamycin for 0, 2, 4, and 6 days. The numbers denote the corresponding OD680nm values. (**B**) The phylogenetic tree of ApFKBP12 protein and homologs from other species. Phylogenetic tree was constructed by MEGA 4 software using the Neighbor-Joining method. Numbers represent bootstrap percentages (1000 of bootstrap replicates). *Schizosaccharomyces pombe* (Sp) (Fungi, Ascomycota), *Oryza sativa* (Os) (Plantae, Tracheophyta), *Solanum tuberosum* (St) (Plantae, Tracheophyta). (**C**) Sequence alignment of the ApFKBP12 protein and homologs from other species. The red rectangle denotes the amino acid required for FKBP12 binding to rapamycin. (**D**) Sequence alignment of the FRB domains of the ApTOR protein and homologs from other species. The red rectangle denotes the amino acid required for the FRB domain binding to rapamycin.



Figure 3. The effects of TOR protein inhibitors on the growth of *Auxenochlorella pyrenoidosa*. (A) AZD8055 inhibits the growth of *A. pyrenoidosa* in a dose-dependent manner. (B) Change curves of

OD680nm values of *A. pyrenoidosa* treated with 1, 5 and 10 μ M AZD8055 for 0, 2, 4 and 6 days. (C) Phenotype of *A. pyrenoidosa* treated with 1, 5 and 10 μ M KU0063794 for 0, 2, 4, and 6 days. (D) Change curves of OD680nm values as described in (C). (E) Phenotype of *A. pyrenoidosa* treated with 1, 10, 20 μ M Torin1 for 0, 2, 4, and 6 days. (F) Change curves of OD680nm values as described in (E).

2.3. Analysis of Transcriptome Sequencing under ApTOR Inhibition

To further clarify the roles of ApTOR signaling pathway in *A. pyrenoidosa*, the transcriptome sequencing was performed in *A. pyrenoidosa* treated with AZD8055. The growth curve showed that *A. pyrenoidosa* was in the logarithmic phase after incubation for 4 days (Figure 3); we therefore cultured the algal cells for 4 days before AZD8055 treatment. In addition, we found that the OD_{680nm} value, chlorophyll content, protein content, and starch content of *A. pyrenoidosa* were significantly changed during *A. pyrenoidosa* treated with 5 μ M AZD8055 for 24 h (Figure 4); thus, *A. pyrenoidosa* treated with 5 μ M AZD8055 for 24 h was used for transcriptome sequencing.



Figure 4. TOR regulates the biosynthesis of major intracellular substances in *Auxenochlorella pyrenoidosa*. (**A**) Phenotypes of *A. pyrenoidosa* treated with AZD8055 for 0, 12, 24, 36, and 48 h. *A. a pyrenoidosa* was cultured in 50 mL BG11 liquid medium for 4 days. Then, final-concentration 5 μ M AZD8055 or equivalent DMSO was added into the alga solution for 0, 12, 24, 36, and 48 h. (**B**) Change curves of OD680nm values of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**C**) Total chlorophyll content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**C**) Total chlorophyll content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) Starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) starch content of *A. pyrenoidosa* treated with 5 μ M AZD8055 for 0, 12, 24, 36, and 48 h. (**E**) starch content of *A. pyrenoidosa* (the pyrenoidosa) starch contents, respectively. The data represents the mean \pm SD of n = 3 independe

A. pyrenoidosa was cultured for 4 days, and then final-concentration 5 μM AZD8055 or equivalent DMSO was added into the algal solution for 24 h. Subsequently, AZD8055treated algal cells were used for transcriptome sequencing. After filtering the raw data, clean reads for subsequent analysis were obtained, and the data summary is as shown in Supplementary Table S1. A total of 2823 differentially expressed genes (DEGs) were found between AZD8055 treatment and DMSO control, of which 1205 DEGs were up-regulated and 1618 were down-regulated (Figure 5A). To verify the reliability of transcriptome data, 10 DEGs were randomly selected from the transcriptome data for qRT-PCR. The qRT-PCR results showed the same trend as the transcriptome data (Supplementary Figure S2), indicating that the transcriptome data were valid and reliable.



Figure 5. The transcriptome data analysis of AZD8055-treated *Auxenochlorella pyrenoidosa*. (**A**) Number of up- and down-regulated DEGs between AZD8055 and DMSO treatment. (**B**) Cluster analysis of DEGs between AZD8055 and DMSO treatment. The color represents the FPKM value of the gene by Z-score. Red denotes high gene expression and green denotes low gene expression. (**C**) The top 20 enriched gene ontology in down-regulated DEGs. (**D**) The top 20 enriched gene ontology in up-regulated DEGs. (**E**) The top 20 enriched KEGG pathways in down-regulated DEGs. (**F**) The top 20 enriched KEGG pathways in up-regulated DEGs.

A cluster analysis of DEGs showed that the transcription levels of many genes were changed in *A. pyrenoidosa* treated with AZD8055 compared to the DMSO control (Figure 5B). To clarify the functions of DEGs, GO functional enrichment analysis was conducted, and a total of 121 down-regulated GO terms and 124 up-regulated GO terms were enriched in the transcriptome data. Among the down-regulated GO terms, thylakoid (GO:0009579) and photosynthesis (GO:0015979) were most significant enrichment (Figure 5C). Among the up-regulated GO terms, the cellular protein modification process (GO:0006464) and autophagy (GO:0006914) were most significant enrichment (Figure 5D). These results suggested that ApTOR regulates multiple biological processes in *A. pyrenoidosa*. A KEGG pathway enrichment analysis showed that photosynthesis, carbon metabolism and carbon fixation in photosynthetic organisms were most significant enrichment in the down-regulated DEGs (Figure 5E). Among the up-regulated KEGG pathways, ubiquitin mediated proteolysis and circadian rhythm-plant were most significantly enriched (Figure 5F). These results suggested that ApTOR controls various intracellular metabolism and signaling pathways in *A. pyrenoidosa*.

2.4. DEGs Involved in Regulating Chloroplast Development and Photosynthesis of Auxenochlorella Pyrenoidosa

Chloroplasts containing chlorophyll are necessary for photosynthesis in plants [41]. Previous studies showed that TOR has the function of regulating chloroplast development and photosynthesis in *Arabidopsis* [42]. Down-regulated GO terms related to photosynthesis and chloroplast development were enriched in the transcriptome data (Figure 5C). Meanwhile, metabolic pathways related to plant photosynthesis were also found in the KEGG pathways, such as photosynthesis, carbon fixation in photosynthetic organisms, and porphyrin and chlorophyll metabolism (Figure 5E). These results indicated that ApTOR has important effects on chloroplast development and photosynthesis in *A. pyrenoidosa*.

A total of 62 DEGs were associated with photosynthesis under ApTOR inhibition, among which 29, 9 and 24 DEGs were enriched in the "Photosynthesis", "Photosynthesisantenna proteins" and "Carbon fixation in photosynthetic organisms" KEGG pathways, respectively (Table 2). Most DEGs related to photosynthesis were significantly down-regulated, and all DEGs of photosystem I, photosystem II and chlorophyll a/b binding protein pathways were down-regulated under ApTOR inhibition (Table 2 and Supplementary Figure S3). The highest down-regulated gene was *Chlorophyll a-b binding protein 4 (Cluster-495.6691)* gene with 131.60-fold decrease. In the dark reaction of photosynthesis, 22 down-regulated DEGs and 2 up-regulated DEGs were involved in carbon fixation. In addition, all 16 chlorophyll synthesis genes involved in "Porphyrin and chlorophyll biosynthesis" pathway were down-regulated from 2.27- to 27.67-fold, and 44 down-regulated DEGs and 6 up-regulated DEGs were involved in the "Thylakoid" pathway (Supplementary Table S2). These results suggested that ApTOR positively regulates chloroplast development and photosynthesis in *A. pyrenoidosa*.

Table 2. Differentially expressed genes in the photosynthetic process.

Gene ID	Log ₂ (Fold Change)	P-adj	KO Name	KO Description
Photosynthesis				
Cluster-495.6756	-5.9733	$5.56 imes 10^{-90}$	PETE	Plastocyanin
Cluster-498.0	-5.6390	$3.47 imes 10^{-5}$	PSBE	Photosystem II cytochrome b559 subunit α
Cluster-495.7678	-5.6230	$6.53 imes 10^{-124}$	PSAN	Photosystem I subunit psan
Cluster-495.7002	-5.4544	$9.90 imes 10^{-104}$	PSAH	Photosystem I subunit VI
Cluster-495.6324	-5.3253	$2.42 imes 10^{-70}$	PSBP	Photosystem II oxygen-evolving enhancer 2
Cluster-495.8198	-5.2049	$9.45 imes10^{-3}$	ATPF0B	F-type H+ transporting ATPase subunit b
Cluster-495.827	-4.6442	$6.89 imes10^{-4}$	PETA	Apocytochrome f
Cluster-495.10228	-4.6211	$4.34 imes10^{-2}$	PSBJ	Photosystem II psbj protein
Cluster-495.5726	-4.0208	$1.05 imes10^{-31}$	PSBS	Photosystem II 22kda protein
Cluster-495.5249	-4.0063	$1.02 imes 10^{-46}$	PSB27	Photosystem II Psb27 protein
Cluster-495.4105	-3.9952	$2.49 imes 10^{-2}$	ATPF1A	F-type H+-transporting atpase subunit alpha
Cluster-495.6958	-3.8390	5.22×10^{-43}	PSAF	Photosystem I subunit III
Cluster-495.7332	-3.7857	$2.68 imes 10^{-3}$	PSBC	Photosystem II CP43 chlorophyll apoprotein
Cluster-495.6505	-3.4964	$1.36 imes 10^{-28}$	PSAK	Photosystem I subunit X
Cluster-495.6979	-3.3834	1.87×10^{-23}	PSBR	Photosystem II 10kda protein
Cluster-495.8002	-3.2884	$4.79 imes10^{-6}$	ATPF1B	F-type H+-transporting atpase subunit beta
Cluster-495.6124	-3.1305	$3.85 imes 10^{-3}$	PETB	Cytochrome b6
Cluster-495.7894	-3.0592	$3.40 imes10^{-17}$	PSAB	Photosystem I P700 chlorophyll a apoprotein
Cluster-495.1037	-2.9414	$3.35 imes 10^{-20}$	PSBP	Photosystem II oxygen-evolving enhancer 2
Cluster-495.4143	-2.8547	$8.51 imes 10^{-45}$	PSAD	Photosystem I subunit II
Cluster-495.6035	-2.8402	$3.44 imes10^{-36}$	PSAL	Photosystem I subunit XI
Cluster-495.4194	-2.7017	$7.15 imes10^{-18}$	PSAO	Photosystem I subunit psao
Cluster-495.4418	-2.3110	$2.12 imes 10^{-21}$	PSAG	Photosystem I subunit V
Cluster-495.5609	-2.2904	$4.52 imes10^{-18}$	PSBO	Photosystem II oxygen-evolving enhancer 1
Cluster-495.4835	-1.7041	$5.65 imes 10^{-12}$	PETJ	Cytochrome c6
Cluster-495.6263	-1.5457	$9.25 imes 10^{-13}$	PETH	Ferredoxin–NADP+ reductase
Cluster-495.7190	-1.4895	$5.90 imes 10^{-11}$	PSBY	Photosystem II psby protein
Cluster-495.4933	-1.2359	$4.99 imes10^{-8}$	PSB28	Photosystem II 13kda protein
Cluster-495.7871	1.1214	$1.83 imes 10^{-5}$	PETF	Ferredoxin

Gene ID	Log ₂ (Fold Change)	P-adj	KO Name	KO Description
Photosynthesis-ante	nna proteins			
Cluster-495.6691	-7.0421	$1.03 imes 10^{-108}$	LHCA4	Photosystem I chlorophyll a/b binding protein 4
Cluster-495.3640	-5.9832	$4.43 imes 10^{-122}$	LHCA3	Photosystem I chlorophyll a/b binding protein 3
Cluster-495.6564	-5.2717	$1.81 imes 10^{-78}$	LHCA1	Photosystem I chlorophyll a/b binding protein 1
Cluster-495.5144	-4.5671	$6.62 imes 10^{-78}$	LHCB4	Photosystem II chlorophyll a/b binding protein 4
Cluster-495.6386	-4.1987	$1.03 imes10^{-51}$	LHCB1	Photosystem II chlorophyll a/b binding protein 1
Cluster-495.5553	-3.9840	$1.96 imes10^{-76}$	LHCA4	Photosystem I chlorophyll a/b binding protein 4
Cluster-495.6485	-3.9329	$1.45 imes 10^{-40}$	LHCB5	Photosystem II chlorophyll a/b binding protein 5
Cluster-495.6349	-3.4326	$3.63 imes10^{-33}$	LHCB1	Photosystem II chlorophyll a/b binding protein 1
Cluster-495.8808	-2.5721	2.26×10^{-2}	LHCB2	Photosystem II chlorophyll a/b binding protein 2
Carbon fixation in pl	hotosynthetic organisms			, ,, ,, ,,
Cluster-495.5099	-5.0936	$5.78 imes 10^{-121}$	PGK	Phosphoglycerate kinase
Cluster-495.2276	-4.3041	$1.39 imes10^{-43}$	ALDO	Fructose-bisphosphate aldolase, class I
Cluster-495.7895	-4.0597	$1.13 imes 10^{-9}$	GOT2	Aspartate aminotransferase
Cluster-495.2903	-3.5833	$1.72 imes 10^{-36}$	MDH2	Malate dehydrogenase
Cluster-495.5815	-3.0446	$6.82 imes 10^{-42}$	PRKB	Phosphoribulokinase
Cluster-495.5677	-2.9728	$3.99 imes 10^{-29}$	PPDK	Pyruvate, orthophosphate dikinase
Cluster-495.5005	-2.8583	$3.63 imes10^{-42}$	TPI	Triosephosphate isomerase (TIM)
Cluster-495.5217	-2.7963	$1.37 imes 10^{-29}$	ALDO	Fructose-bisphosphate aldolase, class I
Cluster-495.4546	-2.7251	$3.02 imes 10^{-38}$	TPI	Triosephosphate isomerase (TIM)
Cluster-495.5985	-2.5229	$6.25 imes 10^{-42}$	MDH1	Malate dehydrogenase
Cluster-495.3332	-2.5073	$2.29 imes10^{-19}$	RPIA	Ribose 5-phosphate isomerase A
Cluster-495.1259	-2.5056	$8.33 imes 10^{-6}$	PCKA	Phosphoenolpyruvate carboxykinase
Cluster-495.4145	-1.8990	$6.39 imes10^{-16}$	MDH2	Malate dehydrogenase
Cluster-495.4967	-1.7549	$3.89 imes 10^{-11}$	MAEB	Malate dehydrogenase (NADP+)
Cluster-495.3492	-1.6252	$7.78 imes 10^{-7}$	MAEB	Malate dehydrogenase (NADP+)
Cluster-495.5038	-1.6083	$3.40 imes10^{-17}$	TKTA	Transketolase
Cluster-495.6521	-1.6040	$1.68 imes 10^{-14}$	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
Cluster-495.6227	-1.2768	$9.08 imes 10^{-9}$	GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
Cluster-495.3372	-1.1984	$1.25 imes 10^{-4}$	PPC	Phosphoenolpyruvate carboxylase
Cluster-495.5601	-1.0961	2.63×10^{-9}	E3.1.3.37	Sedoheptulose-bisphosphatase
Cluster-495.2869	-1.0914	3.29×10^{-7}	FBP	Fructose-1,6-bisphosphatase I
Cluster-495.6190	-1.0435	5.51×10^{-5}	GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
Cluster-495.5207	1.1296	$7.49 imes 10^{-7}$	E1.1.1.39	Malate dehydrogenase (decarboxylating)
Cluster-495.7215	1.4168	$6.87 imes 10^{-8}$	GGAT	Glutamate-glyoxylate aminotransferase

Table 2. Cont.

2.5. DEGs Involved in Regulating Protein Synthesis and Catabolism of Auxenochlorella Pyrenoidosa

A. pyrenoidosa has a high protein content, but whether ApTOR regulates the protein synthesis of *A. pyrenoidosa* remains unknown. Previous studies have shown that ribosomes are responsible for protein synthesis in all organisms, and TOR plays an essential role in regulating ribosome synthesis [43–45]. In this study, genes involved in the "Ribosome biogenesis" pathway were significantly changed in *A. pyrenoidosa*, including ribosomal proteins and U3 small nucleolar ribonucleoprotein proteins. A total of 40 DEGs were enriched in the KEGG "Ribosome biogenesis" pathway, including 26 down-regulated DEGs and 14 up-regulated DEGs, and most of DEGs were ribosome proteins (Supplementary Table S3). Importantly, most of DEGs associated with ribosomal proteins were significantly down-regulated, and the most down-regulated gene was *50S ribosomal protein L2 (Cluster-829.0)* with 26.17-fold decrease (Supplementary Table S3). These results indicated that ApTOR inhibition leads to dysfunction of ribosomes, especially changes in ribosomal protein-related genes, further indicating that ApTOR controls protein synthesis by ribosomes.

Autophagy plays a central role in protein degradation, and previous studies showed that TOR negatively regulates autophagy [46–48]. In this study, transcriptome analysis showed that autophagy related DEGs were significantly enriched in GO terms and KEGG pathways (Figure 5D,F). Total 8 DEGs were assigned to the "Regulation of autophagy" pathway, of which 7 genes were up-regulated including *SnRK1* α and *ATG* genes, and 1 gene was down-regulated in the RNA-seq data (Table 3). These results suggested that ApTOR negatively regulates autophagy in *A. pyrenoidosa*. Ubiquitin (Ub)/26S proteasome system (UPS) is the main pathway of protein degradation in cells. Ub is sequentially covalently

linked to the target protein by ubiquitin activase (E1), ubiquitin binding enzyme (E2), and ubiquitin protein ligase (E3), and then the target protein is degraded by the proteasome [49,50]. The "Ubiquitin mediated proteolysis" KEGG pathway was influenced by AZD8055 (Figure 5F). Total 19 DEGs were assigned to the "Ubiquitin mediated proteolysis" pathway, including 15 up-regulated genes and 4 down-regulated genes (Table 3). Four genes encoding E1 activating enzyme were up-regulated from 2.28- to 7.84-fold under ApTOR inhibition. In addition, some important E3 ubiquitin ligase genes, including *Cullin 1, Cullin 3,* and *Cullin 4,* were significantly up-regulated (Table 3). These results showed that ApTOR inhibition activated protein catabolism in *A. pyrenoidosa*.

Gene ID	Log ₂ (Fold Change)	P-adj	KO Name	KO Description
Regulation of autop	hagy			
Cluster-495.165	1.6278	3.68×10^{-2}	SnRK1a	SNF1-related protein kinase 1 α subunit
Cluster-495.6489	1.5520	$9.00 imes10^{-13}$	ATG7	Autophagy-related protein 7
Cluster-495.4583	1.3266	$5.06 imes 10^{-5}$	ATG3	Autophagy-related protein 3
Cluster-495.7507	1.2287	$1.39 imes10^{-6}$	ATG16L1	Autophagy-related protein 16-1
Cluster-495.6822	1.1679	$3.45 imes10^{-6}$	SnRK1a	SNF1-related protein kinase 1 α subunit
Cluster-495.7877	1.1533	$1.91 imes 10^{-7}$	ATG101	Autophagy-related protein 101
Cluster-495.4705	1.0000	$4.53 imes10^{-6}$	ATG11	Autophagy-related protein 11
Cluster-495.2673	-1.6051	$8.07 imes10^{-6}$	ATG12	Autophagy-related protein 12
Ubiquitin mediated	proteolysis			
Cluster-495.5469	2.9686	$4.17 imes 10^{-21}$	UBLE1B	Ubiquitin-like 1-activating enzyme E1 B
Cluster-495.3804	2.7399	$2.91 imes 10^{-53}$	UBE2A	Ubiquitin-conjugating enzyme E2 A
Cluster-495.4516	2.3622	$8.08 imes 10^{-26}$	ERCC8	DNA excision repair protein ERCC8
Cluster-495.4258	1.7941	$1.47 imes 10^{-13}$	CDH1	Cell division cycle 20-like protein 1
Cluster-495.3283	1.7552	$2.29 imes10^{-19}$	CUL3	Cullin 3
Cluster-495.5596	1.7515	$2.80 imes 10^{-2}$	UBLE1B	Ubiquitin-like 1-activating enzyme E1 B
Cluster-495.8038	1.4688	$1.57 imes10^{-16}$	CUL1	Cullin 1
Cluster-495.4275	1.4562	$6.98 imes10^{-11}$	SKP1	S-phase kinase-associated protein 1
Cluster-495.6597	1.4529	$1.99 imes10^{-6}$	UBE2E	Ubiquitin-conjugating enzyme E2 E
Cluster-495.4573	1.2755	$1.57 imes 10^{-9}$	UBE1	Ubiquitin-activating enzyme E1
Cluster-495.4611	1.2633	$8.88 imes10^{-9}$	CUL4	Cullin 4
Cluster-495.2951	1.1905	$1.28 imes 10^{-8}$	UBE1C	Ubiquitin-activating enzyme E1 C
Cluster-495.8466	1.1285	$6.54 imes10^{-4}$	RBX1	RING-box protein 1
Cluster-495.1824	1.0291	$3.12 imes 10^{-5}$	PPIL2	Peptidyl-prolyl cis-trans isomerase-like 2
Cluster-495.4193	1.0217	1.66×10^{-3}	RBX1	RING-box protein 1
Cluster-495.1695	-1.1161	1.39×10^{-3}	FANCL	E3 ubiquitin-protein ligase FANCL
Cluster-495.1057	-1.2830	$1.16 imes 10^{-4}$	UBE3A	Ubiquitin-protein ligase E3 A
Cluster-495.7130	-1.6390	$1.07 imes 10^{-10}$	UBLE1A	Ubiquitin-like 1-activating enzyme E1 A
Cluster-495.9644	-1.7193	$5.31 imes 10^{-3}$	UBE2S	Ubiquitin-conjugating enzyme E2 S

Table 3. Differentially expressed genes in protein catabolism.

2.6. DEGs Involved in Regulating the Cell Growth of Auxenochlorella Pyrenoidosa

Carbon and nitrogen metabolism, protein and fat synthesis are important limiting factors of cell growth and proliferation [51,52]. In this study, the genes associated with carbon metabolism, amino acid metabolism and fatty acid metabolism were significantly changed under ApTOR inhibition (Supplementary Table S4). DEGs of carbon metabolism and biosynthesis of amino acids and fatty acid pathways were significantly enriched in the down-regulated KEGG pathways (Figure 5E). A total 65 DEGs were assigned to the "carbon metabolism" pathway, including 56 down-regulated genes and 9 up-regulated genes. Some rate-limiting enzyme genes in the "carbon metabolism" pathway such as fructose bisphosphate aldolase and pyruvate kinase were significantly down-regulated. A total 45 DEGs were assigned to the "biosynthesis of amino acids" pathway, including 41 down-regulated genes and 4 up-regulated genes. In addition, all 10 DEGs assigned to the "fatty acid biosynthesis" pathway were down-regulated from 2.53- to 28.44-fold (Supplementary Table S4), indicating that AZD8055 inhibited the biosynthesis of fatty

acids in *A. pyrenoidosa*. These results suggested that ApTOR inhibition affects a variety of intracellular metabolic processes, especially carbon and nitrogen metabolism and fatty acid metabolism. The disruption of metabolic homeostasis by AZD8055 may help to inhibit the growth of *A. pyrenoidosa* cells. Consistent with the growth phenotype of *A. pyrenoidosa* treated with AZD8055, all 14 DEGs related to tricarboxylic acid (TCA) cycle were down-regulated from 2.18- to 11.98-fold in the transcriptome data, including rate-limiting enzymes isocitrate dehydrogenase, α -oxoglutarate dehydrogenase and pyruvate dehydrogenase (Table 4), implying that AZD8055 inhibited cell growth of *A. pyrenoidosa* by inhibiting TCA cycle and reducing energy supply.

Gene ID	Log2 (Fold Change)	P-adj	KO Name	KO Description
Cluster-495.2903	-3.5833	1.72×10^{-36}	MDH2	Malate dehydrogenase
Cluster-495.5985	-2.5229	$6.25 imes 10^{-42}$	MDH1	Malate dehydrogenase
Cluster-495.1259	-2.5056	$8.33 imes10^{-6}$	PCKA	Phosphoenolpyruvate carboxykinase
Cluster-495.4145	-1.8990	$6.39 imes10^{-16}$	MDH2	Malate dehydrogenase
Cluster-495.5734	-1.6082	$5.41 imes10^{-16}$	LSC1	Succinyl-CoA synthetase alpha subunit
Cluster-495.5974	-1.5814	$3.82 imes 10^{-15}$	ACO	Aconitate hydratase
Cluster-495.4348	-1.4680	$3.68 imes 10^{-8}$	PDHD	Dihydrolipoamide dehydrogenase
Cluster-495.4853	-1.2993	$1.45 imes10^{-10}$	DLST	α-oxoglutarate dehydrogenase E2
Cluster-495.6949	-1.2720	$1.91 imes 10^{-5}$	PDHC	Pyruvate dehydrogenase E2
Cluster-495.6523	-1.2587	$7.65 imes 10^{-7}$	FUMC	Fumarate hydratase, class II
Cluster-495.5492	-1.2187	$2.33 imes10^{-8}$	IDH1	Isocitrate dehydrogenase
Cluster-495.6105	-1.2067	$8.64 imes10^{-11}$	LSC2	Succinyl-CoA synthetase beta subunit
Cluster-495.2778	-1.1371	2.54×10^{-6}	SDHD	Succinate dehydrogenase subunit
Cluster-495.5250	-1.1254	$1.39 imes10^{-6}$	OGDH	α-oxoglutarate dehydrogenase E1

Table 4. Differentially expressed genes in the TCA cycle.

3. Discussion

TOR regulates protein synthesis, intracellular metabolism and cell proliferation by integrating nutrients, energy and environmental signals [13,14,53]. In this study, we provide some new insights into how ApTOR controls multiple cellular processes to regulate cell growth of *A. pyrenoidosa*. Only TORC1 is found in higher plants and the green algae *C. reinhardtii*, which contains key proteins TOR, RAPTOR and LST8. TORC1 activity is regulated by nutrients and environmental stresses and responds to different environmental conditions by controlling intracellular metabolic processes [21,35]. Consistent with the results of higher plants and *C. reinhardtii*, only one conserved ApTOR protein was found in *A. pyrenoidosa* (Figure 1 and Table 1). The key components RAPTOR and LST8 of TORC1 were present, while the components RICTOR and SIN1 of the TORC2 were absent in *A. pyrenoidosa*, implying that the conserved TORC1 pathway exists in *A. pyrenoidosa*.

Studies have shown that *C. reinhardtii* is sensitive to rapamycin [54]. Unexpectedly, we found that rapamycin had no obvious effect on the growth of *A. pyrenoidosa*, even at a higher concentration of rapamycin (20 μ M) (Figure 2), showing that *A. pyrenoidosa* is insensitive to rapamycin. Phylogenetic tree analysis and amino acid sequence alignment showed that the resistance of *A. pyrenoidosa* to rapamycin may be caused by the loss of ApFKBP12 function. In addition, we found that AZD8055 could effectively inhibit the growth of *A. pyrenoidosa*, while Torin1 and KU0063794 had no effect on the growth of *A. pyrenoidosa* even at higher concentrations, implying that Torin1 and KU0063794 could not act on the kinase domain of ApTOR protein due to amino acid variation.

Photosynthesis is a plant-specific physiological activity, providing energy and sugars for plants autotrophic growth, which is the biggest difference from animals [55,56]. Previous studies have shown that TOR signaling is closely related to chloroplast development and photosynthesis in plants [33,57,58]. Photosynthetic absorption of CO₂ increased TOR activity, which in turn the enhanced TOR activity further promoted photosynthesis in *C. reinhardtii* [58]. Most DEGs involving chloroplast development and photosynthesis,

such as thylakoid, porphyrin and chlorophyll biosynthesis, and photosynthesis, were down-regulated under ApTOR inhibition by AZD8055 (Table 2), showing that ApTOR had important effects on chloroplast development and photosynthesis of *A. pyrenoidosa*.

Protein degradation is mainly mediated by the ubiquitin/26S proteasome pathway and autophagy [59,60]. In this study, we found that ApTOR inhibition activates autophagy and ubiquitin mediated proteolysis pathway in *A. pyrenoidosa* (Table 3), promoting catabolism of protein. However, genes related to ribosome synthesis were significantly down-regulated in the RNA-seq data, thus inhibiting protein synthesis. These results indicated that ApTOR is involved in regulating protein synthesis and catabolism by multiple metabolic pathways in *A. pyrenoidosa*. Furthermore, the transcriptome data showed that ApTOR controls various intracellular metabolism and signaling pathways in *A. pyrenoidosa*. Inhibition of ApTOR activity resulted in disorders of carbon and nitrogen metabolism, protein and fatty acid metabolism and TCA cycle, which further inhibited the cell growth of *A. pyrenoidosa*.

4. Materials and Methods

4.1. Algae and Growth Condition

The strain of *A. pyrenoidosa* (FACHB-9) used in this study was purchased from the Institute of Hydrobiology, Chinese Academy of Sciences (Wuhan, China). *A. pyrenoidosa* was cultured in BG11 liquid medium supplemented with 20 g.L⁻¹ glucose under 28 °C, 2000 lux continuous light, and 180 rpm.

4.2. Treatment of Auxenochlorella Pyrenoidosa by TOR Inhibitors

The *A. pyrenoidosa* cells was inoculated into a 50 mL BG11 liquid medium supplemented with different concentrations of TOR inhibitors (rapamycin, AZD8055, KU0063794, Torin1) and incubated at 28 °C, 2000 lux continuous light, and 180 rpm. The cell density at 680 nm optical density (OD680) was measured with a Microplate Reader (Biotek EpochTM2, Winooski, VT, USA) at 0, 2, 4 and 6 days.

To test whether *A. pyrenoidosa* cells were killed by high concentrations of AZD8055. *A. pyrenoidosa* cells were treated with 1, 5 and 10 μ M AZD8055 for 4 days, and AZD8055 was removed from the medium, then the pellet was resuspended with BG11 and adjusted to the same OD value. Meanwhile, the removed supernatant containing different concentrations of AZD8055 was added into fresh *A. pyrenoidosa* cells. The phenotype was observed after culturing with or without AZD8055 for 4 days.

4.3. Phylogenetic Tree Analysis

Homologous sequences from different species were aligned by ClustalX software. Phylogenetic tree was generated from the Neighbor-Joining method by MEGA 4 software, and Poisson correction model was used to compute genetic distance. TpTOR (XP_002293107.1), CrTOR (XP_042921379.1), PtTOR (XP_002181617.1), AtTOR (NP_175425.2), HsTOR (NP_001373429.1), ScTOR1 (NP_012600.1), ScTOR2 (NP_012719.2), CrFKBP12 (XP_001693615.1), AtFKBP12 (NP_201240.1), StFKBP12 (XP_006351741.1), OsFKBP12 (XP_015625368.1), SpFKBP12 (NP_595257.1), HsFKBP12 (NP_000792.1) and ScFKBP12 (NP_014264.1) protein sequences were download from NCBI database.

4.4. Construction of the RNA-seq Library and Transcriptome Sequencing

A. pyrenoidosa was cultured in 50 mL BG11 liquid medium supplemented with 20 g·L⁻¹ glucose at 28 °C, 2000 lux continuous light, and 180 rpm for 4 days. Then, final-concentration 5 μ M AZD8055 and equivalent DMSO were added into the alga solution for 24 h, and algal cells were precipitated by centrifugation and collected. Three independent biological replicates were performed for each treatment. Total RNA of *A. pyrenoidosa* treated with AZD8055 or DMSO was extracted by Plant RNA extraction kit (TIANGEN, Beijing, China). The RNA library was constructed using NEBNext[®] Ultra TMRNA Library Prep Kit (NEB, Boston, MA, USA) by Tianjin Novogene Bioinformatics Technology Co., Ltd. Qualified

library was sequenced on an Illumina Novaseq 6000 platform and 150 bp paired-end reads were generated. Clean reads were obtained by filtering the raw data.

4.5. Transcriptome Assembly, Annotation and Differential Expression Analysis

After obtaining clean reads, the Trinity software (V2.6.6, Marlborough, MA, USA) [61] was used to spliced clean reads to obtain reference sequences for subsequent analysis. Diamond software (V0.9.13.114, Tübingen, Germany) [62] was used to match the gene sequence into the protein database for functional annotation. Using gene function annotations information from major databases, including NR, GO, KEGG, Pfam, KOG/COG, and Swiss-prot databases, the spliced genes were annotated. DESeq2 R package (1.20.0) [63] was used to analyze the differentially expressed genes (DEGs) between AZD8055 treatment and DMSO control. *P*-adj < 0.05 and | Log2 (Fold change) | > 1 were set as the threshold values of gene differential expression. GO and KEGG plant databases were used to predict the function of genes and describe the gene products, and the annotation information related to plants was selected for GO and KEGG pathway enrichment. Goseq (V1.10.0, Parkville, Australia) and KOBAS (V2.0.12, Beijing, China) software were used for GO and KEGG pathway enrichment analysis of DEGs, respectively [64,65].

4.6. Quantitative Real-Time PCR (qRT-PCR) Validation

To verify reliability of transcriptome data, qRT-PCR was used to quantify the expression levels of 10 randomly selected genes. CDS sequences of the genes were derived from transcriptome sequencing data, and the corresponding specific primers were presented in Supplementary Table S5. *ApActin (Cluster-495.7101)* was used as a reference gene. RNA from *A. pyrenoidosa* that was processed in the same batch as transcriptome sequencing was selected for qRT-PCR. Relative expression levels of genes were assayed by two-step RT-PCR analysis using the Bio-Rad CFX96 Manager software (BIO-RAD, Hercules, CA, USA). Reaction was performed in a final volume of 20 µL containing 10 µL of $2 \times$ SYBR Green PCR Mastermix (Solarbio, Beijing, China). The relative RNA products of the genes were analyzed using the formula $2^{-\Delta\Delta CT}$.

5. Conclusions

In conclusion, this study revealed the conserved ApTOR signaling in *A. pyrenoidosa* and elucidated the effects of TOR inhibitors on the growth of *A. pyrenoidosa*. Transcriptome data results showed that ApTOR is involved in regulating chloroplast development, photosynthesis and intracellular metabolism in *A. pyrenoidosa*, and ApTOR promotes the cell growth of *A. pyrenoidosa* by regulating various signaling pathways and intracellular metabolic processes. This study provides some insights into the function of ApTOR in *A. pyrenoidosa*.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms231911309/s1.

Author Contributions: M.R. and T.Z. designed the experiments. T.Z. and L.L. performed the experiments. M.R., T.Z. and L.L. analyzed the data. T.Z., L.L., J.Z. and H.C. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grants from the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences (34-IUA-02), the "open competition mechanism to select the best candidates project" of Hainan Yazhou Bay Seed Laboratory (B21HJ0203) and the special project of Nanfan Research Institute of Chinese Academy of Agricultural Sciences (YBXM12).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The transcriptome data have been deposited in the NCBI Sequence Read Archive under accession number PRJNA841794.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Heckman, D.S.; Geiser, D.M.; Eidell, B.R.; Stauffer, R.L.; Kardos, N.L.; Hedges, S.B. Molecular evidence for the early colonization of land by fungi and plants. *Science* 2001, *293*, 1129–1133. [CrossRef]
- Spolaore, P.; Joannis-Cassan, C.; Duran, E.; Isambert, A. Commercial applications of microalgae. J. Biosci. Bioeng. 2006, 101, 87–96. [CrossRef]
- 3. Barkia, I.; Saari, N.; Manning, S.R. Microalgae for High-Value Products towards Human Health and Nutrition. *Mar. Drugs* **2019**, *17*, 304. [CrossRef] [PubMed]
- 4. Torres-Tiji, Y.; Fields, F.J.; Mayfield, S.P. Microalgae as a future food source. Biotechnol. Adv. 2020, 41, 107536. [CrossRef]
- 5. Yang, B.; Liu, J.; Jiang, Y.; Chen, F. Chlorella species as hosts for genetic engineering and expression of heterologous proteins: Progress, challenge and perspective. *Biotechnol. J.* **2016**, *11*, 1244–1261. [CrossRef]
- 6. Krienitz, L.; Huss, V.A.; Bock, C. Chlorella: 125 years of the green survivalist. *Trends Plant Sci.* 2015, 20, 67–69. [CrossRef] [PubMed]
- Song, X.; Wang, J.; Wang, Y.; Feng, Y.; Cui, Q.; Lu, Y. Artificial creation of Chlorella pyrenoidosa mutants for economic sustainable food production. *Bioresour. Technol.* 2018, 268, 340–345. [CrossRef] [PubMed]
- 8. Fan, J.; Ning, K.; Zeng, X.; Luo, Y.; Wang, D.; Hu, J.; Li, J.; Xu, H.; Huang, J.; Wan, M.; et al. Genomic Foundation of Starch-to-Lipid Switch in Oleaginous *Chlorella* spp. *Plant Physiol.* **2015**, *169*, 2444–2461. [CrossRef] [PubMed]
- Zhao, X.C.; Tan, X.B.; Yang, L.B.; Liao, J.Y.; Li, X.Y. Cultivation of Chlorella pyrenoidosa in anaerobic wastewater: The coupled effects of ammonium, temperature and pH conditions on lipids compositions. *Bioresour. Technol.* 2019, 284, 90–97. [CrossRef] [PubMed]
- 10. Wang, Q.; Yu, Z.; Wei, D. High-yield production of biomass, protein and pigments by mixotrophic Chlorella pyrenoidosa through the bioconversion of high ammonium in wastewater. *Bioresour. Technol.* **2020**, *313*, 123499. [CrossRef]
- 11. Sachdeva, N.; Kumar, G.D.; Gupta, R.P.; Mathur, A.S.; Manikandan, B.; Basu, B.; Tuli, D.K. Kinetic modeling of growth and lipid body induction in Chlorella pyrenoidosa under heterotrophic conditions. *Bioresour. Technol.* **2016**, *218*, 934–943. [CrossRef]
- 12. Liu, Z.H.; Li, T.; He, Q.Y.; Sun, Z.; Jiang, Y. Role of Mitochondria in Regulating Lutein and Chlorophyll Biosynthesis in Chlorella pyrenoidosa under Heterotrophic Conditions. *Mar. Drugs* **2018**, *16*, 354. [CrossRef]
- 13. Burkart, G.M.; Brandizzi, F. A Tour of TOR Complex Signaling in Plants. *Trends Biochem. Sci.* **2021**, *46*, 417–428. [CrossRef] [PubMed]
- 14. Artins, A.; Caldana, C. The metabolic homeostaTOR: The balance of holding on or letting grow. *Curr. Opin. Plant Biol.* **2022**, *66*, 102196. [CrossRef]
- 15. Shi, L.; Wu, Y.; Sheen, J. TOR signaling in plants: Conservation and innovation. *Development* **2018**, 145, dev160887. [CrossRef] [PubMed]
- 16. Szwed, A.; Kim, E.; Jacinto, E. Regulation and metabolic functions of mTORC1 and mTORC2. *Physiol. Rev.* **2021**, *101*, 1371–1426. [CrossRef] [PubMed]
- 17. Tafur, L.; Kefauver, J.; Loewith, R. Structural Insights into TOR Signaling. Genes 2020, 11, 885. [CrossRef] [PubMed]
- 18. Liu, Y.; Xiong, Y. Plant target of rapamycin signaling network: Complexes, conservations, and specificities. *J. Integr. Plant Biol.* **2022**, *64*, 342–370.
- 19. Xiong, Y.; Sheen, J. Novel links in the plant TOR kinase signaling network. *Curr. Opin. Plant Biol.* **2015**, *28*, 83–91. [CrossRef] [PubMed]
- 20. van Dam, T.J.; Zwartkruis, F.J.; Bos, J.L.; Snel, B. Evolution of the TOR pathway. J. Mol. Evol. 2011, 73, 209–220. [CrossRef]
- 21. Dobrenel, T.; Caldana, C.; Hanson, J.; Robaglia, C.; Vincentz, M.; Veit, B.; Meyer, C. TOR Signaling and Nutrient Sensing. *Annu. Rev. Plant Biol.* **2016**, *67*, 261–285. [CrossRef]
- 22. Crespo, J.L.; Hall, M.N. Elucidating TOR signaling and rapamycin action: Lessons from Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.* 2002, *66*, 579–591. [CrossRef]
- 23. Xu, Q.; Liang, S.; Kudla, J.; Luan, S. Molecular characterization of a plant FKBP12 that does not mediate action of FK506 and rapamycin. *Plant. J.* **1998**, *15*, 511–519. [CrossRef] [PubMed]
- Sormani, R.; Yao, L.; Menand, B.; Ennar, N.; Lecampion, C.; Meyer, C.; Robaglia, C. Saccharomyces cerevisiae FKBP12 binds Arabidopsis thaliana TOR and its expression in plants leads to rapamycin susceptibility. *BMC Plant Biol.* 2007, 7, 26. [CrossRef] [PubMed]
- 25. Imamura, S.; Ishiwata, A.; Watanabe, S.; Yoshikawa, H.; Tanaka, K. Expression of budding yeast FKBP12 confers rapamycin susceptibility to the unicellular red alga Cyanidioschyzon merolae. *Biochem. Biophys. Res. Commun.* **2013**, 439, 264–269. [CrossRef]
- 26. Ren, M.; Venglat, P.; Qiu, S.; Feng, L.; Cao, Y.; Wang, E.; Xiang, D.; Wang, J.; Alexander, D.; Chalivendra, S.; et al. Target of rapamycin signaling regulates metabolism, growth, and life span in Arabidopsis. *Plant Cell* **2012**, *24*, 4850–4874. [CrossRef]
- 27. Montané, M.H.; Menand, B. TOR inhibitors: From mammalian outcomes to pharmacogenetics in plants and algae. *J. Exp. Bot.* **2019**, *70*, 2297–2312. [CrossRef] [PubMed]
- 28. Montané, M.H.; Menand, B. ATP-competitive mTOR kinase inhibitors delay plant growth by triggering early differentiation of meristematic cells but no developmental patterning change. *J. Exp. Bot.* **2013**, *64*, 4361–4374. [CrossRef]
- Van Leene, J.; Han, C.; Gadeyne, A.; Eeckhout, D.; Matthijs, C.; Cannoot, B.; De Winne, N.; Persiau, G.; Van De Slijke, E.; Van de Cotte, B.; et al. Capturing the phosphorylation and protein interaction landscape of the plant TOR kinase. *Nat. Plants* 2019, *5*, 316–327. [CrossRef]

- 30. Li, L.; Liu, K.H.; Sheen, J. Dynamic Nutrient Signaling Networks in Plants. Annu. Rev. Cell Dev. Biol. 2021, 37, 341–367. [CrossRef]
- McCready, K.; Spencer, V.; Kim, M. The Importance of TOR Kinase in Plant Development. Front. Plant Sci. 2020, 11, 16. [CrossRef] [PubMed]
- 32. Fu, L.; Wang, P.; Xiong, Y. Target of Rapamycin Signaling in Plant Stress Responses. Plant Physiol. 2020, 182, 1613–1623. [CrossRef]
- Song, Y.; Alyafei, M.S.; Masmoudi, K.; Jaleel, A.; Ren, M. Contributions of TOR Signaling on Photosynthesis. *Int. J. Mol. Sci.* 2021, 22, 8959. [CrossRef] [PubMed]
- 34. Shemi, A.; Ben-Dor, S.; Vardi, A. Elucidating the composition and conservation of the autophagy pathway in photosynthetic eukaryotes. *Autophagy* **2015**, *11*, 701–715. [CrossRef]
- 35. Pérez-Pérez, M.E.; Couso, I.; Crespo, J.L. The TOR Signaling Network in the Model Unicellular Green Alga Chlamydomonas reinhardtii. *Biomolecules* **2017**, *7*, 54. [CrossRef]
- 36. Mubeen, U.; Jüppner, J.; Alpers, J.; Hincha, D.K.; Giavalisco, P. Target of Rapamycin Inhibition in Chlamydomonas reinhardtii Triggers de Novo Amino Acid Synthesis by Enhancing Nitrogen Assimilation. *Plant Cell* **2018**, *30*, 2240–2254. [CrossRef]
- Jüppner, J.; Mubeen, U.; Leisse, A.; Caldana, C.; Wiszniewski, A.; Steinhauser, D.; Giavalisco, P. The target of rapamycin kinase affects biomass accumulation and cell cycle progression by altering carbon/nitrogen balance in synchronized Chlamydomonas reinhardtii cells. *Plant J.* 2018, 93, 355–376. [CrossRef]
- Werth, E.G.; McConnell, E.W.; Couso Lianez, I.; Perrine, Z.; Crespo, J.L.; Umen, J.G.; Hicks, L.M. Investigating the effect of target of rapamycin kinase inhibition on the Chlamydomonas reinhardtii phosphoproteome: From known homologs to new targets. *New Phytol.* 2019, 221, 247–260. [CrossRef]
- 39. Benjamin, D.; Colombi, M.; Moroni, C.; Hall, M.N. Rapamycin passes the torch: A new generation of mTOR inhibitors. *Nat Rev Drug Discov.* **2011**, *10*, 868–880. [CrossRef]
- 40. Choi, J.; Chen, J.; Schreiber, S.L.; Clardy, J. Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. *Science* **1996**, 273, 239–242. [CrossRef]
- 41. Kaiser, E.; Correa Galvis, V.; Armbruster, U. Efficient photosynthesis in dynamic light environments: A chloroplast's perspective. *Biochem. J.* 2019, 476, 2725–2741. [CrossRef] [PubMed]
- Dong, P.; Xiong, F.; Que, Y.; Wang, K.; Yu, L.; Li, Z.; Ren, M. Expression profiling and functional analysis reveals that TOR is a key player in regulating photosynthesis and phytohormone signaling pathways in Arabidopsis. *Front. Plant Sci.* 2015, 6, 677. [CrossRef] [PubMed]
- 43. Brunkard, J.O. Exaptive Evolution of Target of Rapamycin Signaling in Multicellular Eukaryotes. *Dev. Cell* **2020**, *54*, 142–155. [CrossRef] [PubMed]
- 44. Saba, J.A.; Liakath-Ali, K.; Green, R.; Watt, F.M. Translational control of stem cell function. *Nat. Rev. Mol. Cell Biol.* 2021, 22, 671–690. [CrossRef]
- 45. Mak, T.; Jones, A.W.; Nurse, P. The TOR-dependent phosphoproteome and regulation of cellular protein synthesis. *Embo J.* **2021**, 40, e107911. [CrossRef]
- 46. Ganley, I.G.; Lam du, H.; Wang, J.; Ding, X.; Chen, S.; Jiang, X. ULK1.ATG13.FIP200 complex mediates mTOR signaling and is essential for autophagy. *J. Biol. Chem.* **2009**, *284*, 12297–12305. [CrossRef]
- 47. Kim, J.; Kundu, M.; Viollet, B.; Guan, K.L. AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. *Nat. Cell Biol.* **2011**, *13*, 132–141. [CrossRef]
- 48. Pu, Y.; Luo, X.; Bassham, D.C. TOR-Dependent and -Independent Pathways Regulate Autophagy in Arabidopsis thaliana. *Front. Plant Sci.* **2017**, *8*, 1204. [CrossRef]
- 49. Cappadocia, L.; Lima, C.D. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem. Rev.* **2018**, *118*, 889–918. [CrossRef]
- 50. Mehrtash, A.B.; Hochstrasser, M. Ubiquitin-dependent protein degradation at the endoplasmic reticulum and nuclear envelope. *Semin. Cell Dev. Biol.* **2019**, *93*, 111–124. [CrossRef]
- Yuan, H.X.; Xiong, Y.; Guan, K.L. Nutrient sensing, metabolism, and cell growth control. *Mol. Cell* 2013, 49, 379–387. [CrossRef] [PubMed]
- 52. Baslam, M.; Mitsui, T.; Sueyoshi, K.; Ohyama, T. Recent Advances in Carbon and Nitrogen Metabolism in C3 Plants. *Int. J. Mol. Sci.* **2020**, *22*, 318. [CrossRef] [PubMed]
- Caldana, C.; Li, Y.; Leisse, A.; Zhang, Y.; Bartholomaeus, L.; Fernie, A.R.; Willmitzer, L.; Giavalisco, P. Systemic analysis of inducible target of rapamycin mutants reveal a general metabolic switch controlling growth in Arabidopsis thaliana. *Plant J.* 2013, 73, 897–909. [CrossRef] [PubMed]
- 54. Pérez-Pérez, M.E.; Florencio, F.J.; Crespo, J.L. Inhibition of target of rapamycin signaling and stress activate autophagy in Chlamydomonas reinhardtii. *Plant Physiol.* **2010**, *152*, 1874–1888. [CrossRef]
- 55. Martin, W.F.; Bryant, D.A.; Beatty, J.T. A physiological perspective on the origin and evolution of photosynthesis. *FEMS Microbiol. Rev.* **2018**, *42*, 205–231. [CrossRef]
- 56. Stirbet, A.; Lazár, D.; Guo, Y.; Govindjee, G. Photosynthesis: Basics, history and modelling. *Ann. Bot.* 2020, 126, 511–537. [CrossRef]
- 57. Li, L.; Song, Y.; Wang, K.; Dong, P.; Zhang, X.; Li, F.; Li, Z.; Ren, M. TOR-inhibitor insensitive-1 (TRIN1) regulates cotyledons greening in Arabidopsis. *Front. Plant Sci.* 2015, *6*, 861. [CrossRef]

- 58. Mallén-Ponce, M.J.; Pérez-Pérez, M.E.; Crespo, J.L. Photosynthetic assimilation of CO₂ regulates TOR activity. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2115261119. [CrossRef]
- 59. Vierstra, R.D. The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Bio.l* 2009, 10, 385–397. [CrossRef]
- 60. Siva Sankar, D.; Dengjel, J. Protein complexes and neighborhoods driving autophagy. Autophagy 2021, 17, 2689–2705. [CrossRef]
- 61. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef]
- Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 2015, 12, 59–60. [CrossRef]
- Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014, 15, 550. [CrossRef]
- 64. Mao, X.; Cai, T.; Olyarchuk, J.G.; Wei, L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **2005**, *21*, 3787–3793. [CrossRef]
- 65. Young, M.D.; Wakefield, M.J.; Smyth, G.K.; Oshlack, A. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* **2010**, *11*, R14. [CrossRef] [PubMed]





Article Atomic Simulation of the Binding of JAK1 and JAK2 with the Selective Inhibitor Ruxolitinib

Maxim Kondratyev ¹^(D), Vladimir R. Rudnev ^{2,3}, Kirill S. Nikolsky ², Alexander A. Stepanov ²^(D), Denis V. Petrovsky ², Liudmila I. Kulikova ^{2,3,4}, Arthur T. Kopylov ²^(D), Kristina A. Malsagova ^{2,*}^(D) and Anna L. Kaysheva ²^(D)

- ¹ Institute of Cell Biophysics, Russian Academy of Sciences, 142290 Pushchino, Russia
- ² Biobanking Group, Branch of Institute of Biomedical Chemistry "Scientific and Education Center", 109028 Moscow, Russia
- ³ Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, 142290 Pushchino, Russia
- ⁴ Institute of Mathematical Problems of Biology RAS—The Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, 142290 Pushchino, Russia
- * Correspondence: kristina.malsagova86@gmail.com; Tel.: +7-(499)-764-98-78

Abstract: Rheumatoid arthritis belongs to the group of chronic systemic autoimmune diseases characterized by the development of destructive synovitis and extra-articular manifestations. Cytokines regulate a wide range of inflammatory processes involved in the pathogenesis of rheumatoid arthritis and contribute to the induction of autoimmunity and chronic inflammation. Janus-associated kinase (JAK) and signal transducer and activator of transcription (STAT) proteins mediate cell signaling from cytokine receptors, and are involved in the pathogenesis of autoimmune and inflammatory diseases. Targeted small-molecule drugs that inhibit the functional activity of JAK proteins are used in clinical practice for the treatment of rheumatoid arthritis. In our study, we modeled the interactions of the small-molecule drug ruxolitinib with JAK1 and JAK2 isoforms and determined the binding selectivity using molecular docking. Molecular modeling data show that ruxolitinib selectively binds the JAK1 and JAK2 isoforms with a binding affinity of -8.3 and -8.0 kcal/mol, respectively. The stabilization of ligands in the cavity of kinases occurs primarily through hydrophobic interactions. The amino acid residues of the protein globules of kinases that are responsible for the correct positioning of the drug ruxolitinib and its retention have been determined.

Keywords: ruxolitinib; JAK inhibitor; rheumatoid arthritis; molecular modeling

1. Introduction

Rheumatoid arthritis (RA) can lead to severe disability; therefore, timely and effective treatment is important to reduce the negative effects of functional damage (deformation and destruction) to the joints, which occurs at an early stage of RA development [1]. There are no specific RA serological markers that allow a high-precision diagnosis of RA at an early stage before the onset of clinical symptoms in current clinical practice [2]. A deep study of the molecular basis of the disease's development seems to be relevant to the development of new approaches to RA prevention and cure. The most attractive approaches are non-therapeutic ones (physical activity, diet) that, on the one hand, improve the general state of human health, and, on the other hand, reduce the negative symptoms of RA development (inflammation, pain, joint damage). The literature presents the results of studies that confirm the positive effect of long-term, high intensity exercise on the general health of patients with rheumatoid arthritis and knee osteoarthritis [3-6]. The authors of the studies agree that the deterioration in health status in RA is exacerbated by the avoidance of exercise [5,6]. The authors of a number of studies note cases of combined progression in patients with RA and intolerance to certain food allergens [2,7–9]. Studies show a positive effect of short-term fasting, a vegan diet, and an elimination diet on improving the health

Citation: Kondratyev, M.; Rudnev, V.R.; Nikolsky, K.S.; Stepanov, A.A.; Petrovsky, D.V.; Kulikova, L.I.; Kopylov, A.T.; Malsagova, K.A.; Kaysheva, A.L. Atomic Simulation of the Binding of JAK1 and JAK2 with the Selective Inhibitor Ruxolitinib. *Int. J. Mol. Sci.* 2022, 23, 10466. https://doi.org/10.3390/ ijms231810466

Academic Editor: Wajid Zaman

Received: 12 August 2022 Accepted: 6 September 2022 Published: 9 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). of RA patients, which is likely due to a reduction in immune reactivity to certain food antigens in the gastrointestinal tract that are eliminated by changing the diet [10,11].

The activation of lymphocytes and synthesis of pro-inflammatory cytokines by macrophages, primarily interleukin 1 beta (IL1 β) and tumor necrosis factor alpha (TNF α), play a key role in the inflammation of the synovial membrane in RA. These interleukins contribute to the persistence of the inflammatory process in the synovial membrane and the destruction of cartilage and bone tissue, because of the direct effect on synovial fibroblasts, chondrocytes, and osteoclasts. Similar to the inflammatory process, cyclooxygenase-2 (COX-2) is activated, leading to an increase in the synthesis of prostaglandins. Disease-modifying antirheumatic drugs (DMARDs), which are immunosuppressive and immunomodulatory agents, are currently widely used in the clinical treatment of RA [12]. Janus-associated kinase (JAK) inhibitors are a new class of targeted synthetic DMARDs (tsDMARDs) used to treat RA [13]. tsDMARDs provide targeted inhibition of JAKs, which plays a key role in the pathogenesis of RA [1]. JAK family enzymes are represented by four isoforms: JAK1, JAK2, JAK3, and tyrosine kinase 2 (TYK2), which are important signaling molecules involved in cytokine and growth factor signaling [14,15]. The cytokine receptors signal through the JAK-STAT pathway [15]. JAK molecules act together in cytokine signaling (the majority of cytokine receptors use three JAK combinations), but under certain conditions, they exhibit selectivity for one isoform (Figure 1).



Figure 1. Cytokines that signal through the JAK family proteins. IL-6 signaling, which plays an important role in the pathogenesis of RA, leads to the activation of JAK1, JAK2, and TYK2. Erythropoietin receptor (EpoR) signaling is mediated by JAK2 and is important in the development and deployment of reticulocytes and erythrocytes. JAK3 in combination with JAK1 is an important signaling component for cytokine receptors that share a common gamma chain (γ c), such as IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21.

Isoforms of the JAK family exhibit cross-activity. JAK1 plays an important role in the signaling of several pro-inflammatory cytokines, often in collaboration with other members of the JAK family. Thus, JAK2 is used mainly by receptors for hematopoietic growth factors, such as erythropoietin and thrombopoietin. JAK3 is thought to play a major role in the mediated activation of immune function, while TYK2 functions in association with JAK2 or JAK3 to signal cytokines such as interleukin-12 (IL-12) [16].

Timely and effective treatment of RA at an early stage is essential to control joint damage. Small molecule JAK inhibitors represent a new class of drugs for the treatment of rheumatoid arthritis.

Ruxolitinib (KEGG ID D09960) is a tsDMARD with the International Union of Pure and Applied Chemistry (IUPAC) name (3R)-3-cyclopentyl-3-[4-(7H-pyrrolo [2,3-d]pyrimidin-4-yl) pyrazol-1-yl] propanenitrile and a molecular weight of 306.4 Da [17]. Ruxolitinib is a pyrazole, substituted at position 1 by a 2-cyano-1-cyclopentylethyl group, and at position 3 by a pyrrolo [2,3-d]pyrimidin-4-yl group. Ruxolitinib (INCB018424) is the first potent and selective JAK1/2 inhibitor to be approved for medical use, with a half maximal inhibitory concentration (IC50) of 3.3 nM/2.8 nM in cell-free assays and >130-fold selectivity for JAK1/2 versus JAK3 [18].

Ruxolitinib is a class I molecule of the Biopharmaceutical Classification System (BCS) with high permeability, high solubility, and rapid disintegration. Ruxolitinib is primarily metabolized by the CYP3A4 cytochrome P450 family proteins (>50%) via CYP2C9. At clinically significant concentrations, the drug does not inhibit CYP1A2, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, or CYP3A4, and is not a potent inducer of CYP1A2, CYP2B6, or CYP3A4. The drug is used in myeloproliferative neoplasms and autoimmune diseases, which are known to be associated with the dysregulation of JAK1 and JAK2. This dysregulation is believed to be due to the high levels of circulating cytokines, which are associated with JAK-STAT pathway activation, JAK2V617F mutations, and the down-regulation of negative regulatory mechanisms. Ruxolitinib inhibits JAK-STAT signaling and cell proliferation in cytokine-dependent cell models of hematologic malignancies, and Ba/F3 cells become cytokine-independent by the expression of JAK2V617F mutant protein, with IC50 ranging from 80 to 320 nM [19]. Ruxolitinib has two chiral forms: S2902 S-Ruxolitinib is the S form and S1378 Ruxolitinib is the D form. One of the carbons in this molecule is asymmetric, rendering the two molecules mirror images of each other. It is probable that the biological activities of these two molecules is different. Numerous studies show that ruxolitinib, a JAK2 inhibitor, suppresses proliferation and induces apoptosis of mutated JAK2V617F cell lines [19–22].

This study was performed to model the interactions of ruxolitinib with JAK1 and JAK2 isoforms, and to determine the binding selectivity using molecular docking.

2. Results

2.1. Characterization of the "Grotto" of Ligand Binding for JAK Isoforms

Calculations show that the studied proteins and ligands bind very specifically. The van der Waals surface of the globule of each protein has a niche, or "grotto", in which each of the studied ligands is completely sterically placed. Externally, the topology of the binding proteins and ligands appears very similar in the "grotto" (Figure 2).

The "grotto" for the JAK1 isoform is formed by a fragment of the amino acid sequence 865–1154 of the JH1 domain, which exhibits tyrosine kinase activity [20]. The secondary structure of the "grotto" is formed by 13 β -strands and 17 α -helices (according to structural identification (STRIDE)), which form the fold of Scope ID d1t46a. The protein belongs to structural class d: alpha and beta proteins (a + b). The "grotto" for the JAK2 isoform is formed by a fragment of the amino acid sequence 842–1130, also of the JH1 domain. The secondary structure of the "grotto" is rich in β -strands (12) and contains 17 α -helices that form the fold of Scope ID d1u46a. This protein also belongs to structural class d: alpha and beta proteins (a + b).

The physicochemical and structural parameters of the "grotto" are similar for both the isoforms. Thus, the area accessible to the solvent "grotto" of JAK1 and JAK2 is 29,255.34 Å² and 31,252.36 Å², respectively (Table 1).



Figure 2. Topology of the "grotto" involved in the binding of low molecular weight ligands (view in two projections) in the globule of the JAK1 protein (locus 865-1154 a.a.) (**a**), JAK2 (locus 842-1130 a.a.) (**b**).

Table 1. Structural and physicochemical properties of "grottoes" for JAK1 and JAK2.

"Grotto"	SASA, Å ²	HB*	Nonpolar a.a. (%)	Polar a.a. (%)	Charged a.a. (%)	Nonpolar "Cross-Elements" Contacts	Nonpolar "Inside-Elements" Contacts	Distance ** of "Cross-Elements" Contacts ($ m \AA \pm$ SD)	Distance of "Inside-Elements" Contacts (Å ± SD)
JAK1 (865-1154)	29,255.336	189	36.8	30.9	31.6	310	43	8.28 ± 1.37	7.80 ± 1.54
JAK2 (842-1130)	31,252.359	198	38.4	30.4	30.4	333	38	8.25 ± 1.40	7.67 ± 1.62

HB *—number of hydrogen bonds; distance **—between nonpolar a.a.

Interesting results were obtained from the analysis of possible contacts between amino acid residues involved in the formation and stabilization of the "grotto" structures. Indeed, we observe that in selected regions of amino acid sequences, approximately 40% of the sequence is made up of hydrophobic amino acid residues. An analysis of possible contacts between amino acid residues reveals a bimodal distribution of distances between the interacting hydrophobic amino acids. The number of contacts formed within one element of the secondary structure (inside-elements) is only a small fraction of probable contacts (Table 1, column "Nonpolar "inside-elements" contacts"), most of which can be attributed to the interactions of amino acid residues localized in different elements of the secondary structures (Table 1, column "Nonpolar "cross-elements" contacts"). Of note is the compactness of the structural motifs that form the "grottoes", which is due to the presence of two maxima in the distribution of distances between the interacting hydrophobic amino acid residues between the interacting hydrophobic amino acid residues between the interacting hydrophobic amino acid residues is observed within 7.7–7.8 Å, and the second maximum for contacts between elements of secondary structures, is observed for distances of 8.3 Å.

2.2. Results of Molecular Docking

Calculations of flexible docking make it possible to estimate the binding affinity, or "affinity". This value is the total energy indicated in kcal/mol with a negative sign, as these complexes are stable. According to our data, the affinity energy of decernotinib (known as ligand-467 from the experimentally solved 4YTH complex in PDB) for the JAK1 protein is -8.7 kcal/mol, and the same value is obtained for the KEV ligand. Ruxolitinib shows an identical binding site with a comparable affinity value of -8.3 kcal/mol (Figure 3a–c). Characterizing the binding of ligands and the JAK2 protein, we found that decernotinib demonstrates the highest affinity (-8.7 kcal/mol), while the KEV ligand binds with an affinity of -8.67 kcal/mol, and ruxolitinib with an energy of -8.0 kcal/mol. These structures are shown in Figure 3d–f.



Figure 3. Ligand binding site and JAK1: (**a**) decernotinib, (**b**) KEV, and (**c**) ruxolitinib; ligands and JAK2: (**d**) decernotinib, (**e**) KEV, and (**f**) ruxolitinib (Supplementary Materials, Figures S1–S6).

Amino acid residues that are important for binding to the studied ligands are detected using the LIGPLOT+ package, and the results are given below. Initially, during visual analysis of the calculated structures, we assumed that ligands with a pronounced heterocyclic nature, as well as halogen substituents, would be bound by hydrophobic interactions (residues LEU881, GLY884, and GLY962 for JAK1) or hydrogen bonds (ARG879 for JAK1) (Supplementary materials, Table S1). This would achieve a reproducible position of the ligands in the "grotto" of the protein globule. For the JAK2 protein, a similar analysis suggests that the LEU855, GLY858, and LEU983 residues are the source of hydrophobic interactions, whereas the ARG980 residue probably acts as a hydrogen bonding agent. Calculations using the LIGPLOT+ package make it possible to confirm that the main type of interactions that stabilizes the studied complexes are hydrophobic interactions (Figure 4a–c,d–f for JAK1 and JAK2, respectively).

The analysis of the presented data makes it possible to supplement the list of residues responsible for positioning the ligand in the hydrophobic pocket of the JAK1 protein. The hydrogen bond with decernotinib is formed by the GLU966 residue, while the amino acid residue LEU959 interacts with the KEV ligand and ruxolitinib. In JAK2 kinase, LEU855, LEU983, and ASP994 residues are involved in the interactions, while in the case of decernotinib and ruxolitinib ligands, they form only hydrophobic interactions. In the KEV ligand, aspartic acid residue forms a hydrogen bond. When analyzing the structure of these homologous proteins, a question arises about the distribution of amino acid residues over identical sites of the globule, especially over the inner surface of the niche, the "grotto", where the binding of the studied set of ligands occurs. Therefore, we calculated the alignment of the amino acid sequences of these two kinases (JAK1 and JAK2) using Clustal 0 version 1.2.4 (Conway Institute UCD, Dublin, Ireland) [23]. The results are shown in Figure 5.

It is worth paying attention to the fact that the globule in the PDB file of the JAK1 protein begins with the VAL865 residue and ends with LYS1154 (290 amino acids in total), while, in the JAK2 protein PDB file, it starts with the THR842 residue and ends with MET1130 (289 residues in total). The alignment results show an average homology (approximately 54%) of the amino acid sequences forming the "grotto" of the two isoforms. Despite this, the identified amino acid substitutions are predominantly homologous; for example, the replacement of a hydrophobic amino acid with another hydrophobic one (such as Val–Ile, Ile–Ala, and Leu–Ile), or a negatively or positively charged amino acid with an amino acid of the same charge character (Asp–Glu and Arg–Lys). This determines the high conservatism of the binding of the studied ligands in the hydrophobic pocket of the globule of each of the studied kinases.







Figure 4. Interactions diagram of JAK1 kinase interactions with decernotinib (**a**), with KEV ligand (**b**), and with ruxolitinib (**c**); JAK2 kinases with decernotinib (**d**), with KEV ligand (**e**), and with ruxolitinib (**f**) (Supplementary materials, Figures S7–S12).

CLUSTAL O(1.2.4) multiple sequence alignment

6N7A_1 Chains 4YTH_1 Chain	GSGDIVSEKKPATEVDPTHFEKRFLKRIRDLGEGHFGKVELCRYDPEGDNTGEQVAVKSL 60 GSHNMTQFEERHLKFLQQLGKGNFGSVEMCRYDPLQDNTGEVVAVKKL 48 . : *:**:*.** :::**:*******************
6N7A_1 Chains 4YTH_1 Chain	KPESGGNHIADLKKEIEILRNLYHENIVKYKGICTEDGGNGIKLIMEFLPSGSLKEYLPK 120 QH-STEEHLRDFEREIEILKSLQHDNIVKYKGVCYSAGRRNLKLIMEYLPYGSLRDYLQK 107 : * :*: *:::*****::* *:******:* . *:*****:** ***::** ***::** *
6N7A_1 Chains 4YTH_1 Chain	NKNKINLKQQLKYAVQICKGMDYLGSRQYVHRDLAARNVLVESEHQVKIGDFGLTKAIET 180 HKERIDHIKLLQYTSQICKGMEYLGTKRYIHRDLATRNILVENENRVKIGDFGLTKVLPQ 167 :*::*: : *:*: ******:***::*****:********
6N7A_1 Chains 4YTH_1 Chain	DKEYYTVKDDRDSPVFWYAPECLMQSKFYIASDWSFGVTLHELLTYCDSDSSPMALFLK 240 DKEYYKVKEPGESPIFWYAPESLTESKFSVASDWSFGVVLYELFTYIEKSKSPPAEFMR 227 *****.**: :**:******.* :*** :**********
6N7A_1 Chains 4YTH_1 Chain	MIGPT-HGQMTVTRLVNTLKEGKRLPCPPNCPDEVYQLMRKCWEFQPSNRTSFQNLIEGF 299 MIGNDKQGQMIVFHLIELLKNNGRLPRPDGCPDEIYMIMTECWNNNVNQRPSFRDLALRV 287 *** :*** * :*:: **:. *** * .****:* :* :**: : .:* **::* .
6N7A_1 Chains 4YTH_1 Chain	EALLK 304 DQIRDNMAG 296 : : .

Figure 5. Amino acid sequence alignment (FASTA) results for JAK1 and JAK2 kinases. The symbol (*) under the sequences indicate identity, conservative substitutions (":") and semi-conservative substitutions (".").

3. Discussion

The study of the pathological processes molecular basis development in the human body has been a promising area of biomedical research for several decades. Even though often clinical practice, diagnosis, and treatment planning do not require knowledge of the etiology of the disease or pathogenesis, awareness of events at the molecular and cellular level is required. This allows us to propose new approaches for determining risk groups among conditionally healthy people and search for attractive therapeutic targets. The present study contributes to understanding the architecture of ligand binding to JAK family proteins, and contributes to understanding the molecular basis of RA development.

The high conservation of JAK1 and JAK2 amino acid sequences in the kinase activity domain (JH1) determines the structural and physicochemical similarities of most of the targeted kinase inhibitors (ligands). Understanding the molecular basis of ligand and target specificity is important for identifying new drugs and inhibitors of various types of kinases [24]. In the present study, a structural analysis of JAK1 and JAK2 isoform motifs involved in the selective binding of ruxolitinib was performed.

There are few studies focusing on the analysis of the structural features of ruxolitinib binding to kinases. Duan (2014) showed, for the first time, the structure of the active domain pair of chicken c-Src kinase (residues 251-533)-ruxolitinib, at a resolution of 2.26 Å; ruxolitinib has a lower selectivity for c-Src compared to JAK1 [24]. The identity between the amino acid sequences of the kinase domains of JAK1 and c-Src is 34%. The authors observed that the pyrrolopyrimidine rings of ruxolitinib are oriented towards the hinge region. The ligand in the JAK1 cavity is stabilized by two hydrogen bonds with Glu957 and Leu959. Simultaneously, the cyclopentane ring is oriented towards the N-terminus of JAK1, while the propanenitrile group of ruxolitinib interacts with JAK1 [23]. In a recent study by Babu (2022), JAK1 binding selectivity was screened for a set of 52 C-2 methyl/hydroxyethylimidazopyrrolopyridine derivatives, including ruxolitinib [25]. Ruxolitinib forms hydrogen bonds at Leu959 and Glu957 in the JAK1 isoform.

The results obtained in the present study show that the pyrazole derivative ruxolitinib binds with a high affinity to the JAK1 and JAK2 isoforms (Table 2).

Ligand	JAK1 (6N7A), kcal/mol	Locus of "Grotto"	HB *	Hydrophobic Bonds **	JAK2 (4YTH), kcal/mol	Locus of "Grotto"	HB	Hydrophobic Bonds
Decernotinib	-8.7	ARG879, LEU881, GLY884, GLY962	1	10	-8.7	LEU855, GLY858, ARG980, LEU983	0	13
Ligand KEV	-8.7	ARG879, LEU881, GLY884, GLY962	1	15	-8.6	LEU855, GLY858, ARG980, LEU983	1	12
Ruxolitinib	-8.3	ARG879, LEU881, GLY884, GLY962	0	16	-8.0	LEU855, GLY858, ARG980, LEU983	0	11

Table 2. Steric and energetic parameters of kinase binding to ligands.

*—number of hydrogen bonds between the target and the ligand; **—number of hydrophobic bonds with distances between the contacting amino acid residues of the target and the ligand less than 10 Å.

The drug ruxolitinib shows similar selectivity for both the JAK isoforms: -8.3 and -8.0 kcal/mol for JAK1 and JAK2, respectively. This result is indirectly confirmed by the literature data (IC50 3.3 ± 1.2 nM for JAK1 and 2.8 ± 1.2 nM for JAK2). The stabilization of the ligands studied in this work, characterized by similar physicochemical properties, was performed primarily by hydrophobic interactions in the JAK "grotto". In the case of decernotinib and KEV ligands, one hydrogen bond was involved in stabilizing the ligand–target complex.

In our study, we refined the localization of ruxolitinib in the JAK1 and JAK2 "grottoes" with a structural resolution of 1.33 Å and 2.04 Å, respectively. In both the isoforms, ruxolitinib is completely located in the "grotto". For the JAK1 isoform, the pyrrolopyrimidine and pyrazole rings form a single plane, which is oriented towards the N-terminus of JAK1 and is parallel to the plane of the Leu875–Gly884 β -hairpin in the "grotto" β -barrel. The plane of the cyclopentane ring is approximately perpendicular to the plane of the pyrrolopyrimidine and pyrazole rings, and probably forms a hydrophobic interaction with the Gly884 amino acid residue in the same β -hairpin. The propanenitrile group forms the third plane, which passes through the line of intersection of the first two planes, the plane of the pyrrolopyrimidine ring and the plane of the cyclopentane ring, and is oriented towards the Asn1008–Ile1019 β -hairpin.

In the "grotto" of the JAK2 isoform, the pyrrolopyrimidine and pyrazole rings are located in intersecting planes. The plane of the pyrrolopyrimidine ring, similar to the JAK1 isoform, is oriented towards the N-terminus of JAK2 and is parallel to the plane of the Asn859–Tyr868 β -hairpin in the "grotto" β -barrel. The pyrrolopyrimidine ring forms a single plane with the cyclopentane ring, which is oriented towards the Tyr913–Glu930 β -hairpin. The propanenitrile group forms a third plane oriented towards the Ile982–Ile992 β -hairpin.

For both the JAK isoforms, we observe that the studied complexes are stabilized by hydrophobic interactions. Hydrogen interactions between the ligand and target are not revealed in our study.

4. Materials and Methods

4.1. Objects of Study

Flexible molecular docking was performed for the ligands: (a) decernotinib (PDB ID 4YTH), (b) KEV with the chemical name N-[3-(5-chloro-2-methoxyphenyl)-1-methyl-1H-pyrazol-4-yl]-2-methyl-2H-pyrazolo [4,3-c]pyridine-7-carboxamide (PDB ID 6N7A) [26], and (c) ruxolitinib, a selective antitumor inhibitor of JAK kinase (Figure 6).



Figure 6. Structural formula of the ligand of (**a**) decernotinib (ligand-467), (**b**) KEV (ligand is drug 39 from), (**c**) ruxolitinib.

Decernotinib (VX-509) is a JAK3 inhibitor (Table 3). This experimental drug is characterized by high selectivity for the JAK3 isoform.

The main objective of our study was to identify the ruxolitinib ligand, an antitumor agent and a selective inhibitor of the JAK1 and JAK2 isoforms. These kinases facilitate the signaling of numerous cytokines and growth factors that play an important role in hematopoiesis and immune system function (Table 3).

Ligand	IUPAC (Formula)	Class	Mw, Da	Donor HB *	Acceptor HB	TPSA **, Å ²	SC ³ *	Selectivity	IC50, Nm ⁴ *
Decernotinib	(2R)-2-methyl-2-[[2-(1H-pyrrolo [2,3-b]pyridin-3-yl)pyrimidin-4- yl]amino]-N-(2,2,2- trifluoroethyl)butanamide	Pyrimidine	392.4	3	8	95.6	0	JAK3	2.5 ±0.7
KEV	N-[3-(5-chloro-2-methoxyphenyl)-1- methyl-1H-pyrazol-4-yl]-2-methyl- 2H-pyrazolo [4,3-c]pyridine-7- carboxamide(C19H17CIN6O2)	Pyrazole	396.8	1	5	86.9	0	JAK1	N/A
Ruxolitinib	(3R)-3-cyclopentyl-3-[4-(7H-pyrrolo [2,3-d]pyrimidin-4-yl)pyrazol-1- yl]propanenitrile	Pyrazole	306.4	1	4	83.2	1	JAK1 and JAK2	$\begin{array}{c} 3.3 \pm 1.2 \\ (JAK1) 2.8 \pm \\ 1.2 \ (JAK2) \end{array}$

Table 3. Physicochemical properties of the ligands used in the work.

*-hydrogen bond; **-topological polar surface area; ^{3*}-stereocenter; ^{4*}-cell-free assay.

As shown in Table 3, the physicochemical properties of the ligands are similar. Two isoforms were chosen as targets for modeling: JAK1 (PDB ID 6N7A) and JAK2 (PDB ID 4YTH) (Table 4).

Table 4. Structures of JAK1 and JAK2 proteins used for redocking with ruxolitinib.

Target	PDB ID	Uniprot ID	Resolution, Å	Ligand	Method	Sequence Length, a.a.	Native Ligand	Binding Affinity with Native Ligand, nM
JAK1	6N7A	P23458	1.33	Ruxolitinib	X-ray	304	KEV	4.3
JAK2	4YTH	O60674	2.04	Ruxolitinib	X-ray	296	Decernotinib	2

4.2. Molecular Docking

Data preparation for docking calculations was carried out according to standard protocols [27–29]. Ligand–target binding and binding energy rankings were performed using the AutoDock VINA 1.1.2 software package [30]. The results were visualized using AutoDockTools 1.5.6. RC3 [18]. The projection schemes of the ligand–receptor interactions were drawn using LIGPLOT 1.3.6 [28].

The spatial structures of the target protein models obtained from PDB were purified from the solvent, ligand, and buffer molecules. In the MGLTools package (CCSB, San Diego, CA, USA), charges were placed on the proteins, which made them a receptor, a target for searching for optimal binding sites of the studied ligands [31]. The structures of the ligands were created using the HyperChem molecular constructor (http://www.hypercubeusa.com/) and sequentially optimized, first by using the Assisted Model Building with Energy Refinement (AMBER) force field [32], and then quantum-chemically optimized using parametric method 3 (PM3) [33]. The arrangement of charges on the ligand molecule and its protonation/deprotonation were carried out automatically using the MGLTools 1.5.6 package [31]. The calculations were performed on a server using IntelXEON processors (40 cores).

The task of analyzing the docking results was significant. The LIGPLOT+ package was used to obtain "sweeps" of interactions [34]. From the three-dimensional coordinates of the atoms of the protein–ligand complex, interaction diagrams were generated that depict the schemes of hydrogen bonds, as well as hydrophobic contacts between the ligand and elements of the main or side chain of the protein.

5. Conclusions

Molecular modeling data shows that the drug ruxolitinib selectively binds the JAK1 and JAK2 isoforms with a binding affinity of -8.3 and -8.0 kcal/mol, respectively. In both the JAK isoforms, ruxolitinib resides entirely in the "grotto" or binding cavity. The stabilization of the ligand in the "grotto" of kinases is performed primarily by hydrophobic interactions, including 18 JAK1–ligand contacts and 11 JAK2–ligand contacts. For the first time, we detail the amino acid residues of the JH1 domains of protein globules, which are responsible for the correct positioning of the ligand and its retention. Agr879, Leu881, Gly884, and Gly962 are involved in the formation of the JAK1–ligand complex; hydrophobic amino acid residues Leu855, Gly858, Arg980, and Leu983 are also involved in the formation of the JAK2–ligand complex.

Supplementary Materials: The supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms231810466/s1.

Author Contributions: Conceptualization, A.L.K.; methodology, M.K. and D.V.P.; software, A.A.S.; validation, K.S.N. and A.T.K.; formal analysis, V.R.R. and K.A.M.; investigation, V.R.R. and L.I.K.; data curation, A.L.K.; writing—original draft preparation, A.L.K. and L.I.K.; writing—K.A.M.; project administration, A.L.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Russian Science Foundation, grant number 21-14-00381.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Sacquin-Mora, S. Fold and Flexibility: What Can Proteins' Mechanical Properties Tell Us about Their Folding Nucleus? J. R. Soc. Interface 2015, 12, 20150876. [CrossRef] [PubMed]
- Khanna, S.; Jaiswal, K.S.; Gupta, B. Managing Rheumatoid Arthritis with Dietary Interventions. Front. Nutr. 2017, 4, 52. [CrossRef] [PubMed]
- 3. Malorgio, A.; Malorgio, M.; Benedetti, M.; Casarosa, S.; Cannataro, R. High Intensity Resistance Training as Intervention Method to Knee Osteoarthritis. *Sports Med. Health Sci.* 2021, *3*, 46–48. [CrossRef]
- 4. Sandstad, J.; Stensvold, D.; Hoff, M.; Nes, B.M.; Arbo, I.; Bye, A. The Effects of High Intensity Interval Training in Women with Rheumatic Disease: A Pilot Study. *Eur. J. Appl. Physiol.* **2015**, *115*, 2081–2089. [CrossRef]
- de Jong, Z.; Munneke, M.; Zwinderman, A.H.; Kroon, H.M.; Ronday, K.H.; Lems, W.F.; Dijkmans, B.A.C.; Breedveld, F.C.; Vlieland, T.P.M.V.; Hazes, J.M.W.; et al. Long Term High Intensity Exercise and Damage of Small Joints in Rheumatoid Arthritis. *Ann. Rheum. Dis.* 2004, 63, 1399–1405. [CrossRef] [PubMed]
- Bartlett, D.B.; Willis, L.H.; Slentz, C.A.; Hoselton, A.; Kelly, L.; Huebner, J.L.; Kraus, V.B.; Moss, J.; Muehlbauer, M.J.; Spielmann, G.; et al. Ten Weeks of High-Intensity Interval Walk Training Is Associated with Reduced Disease Activity and Improved Innate Immune Function in Older Adults with Rheumatoid Arthritis: A Pilot Study. *Arthritis Res. Ther.* 2018, 20, 127. [CrossRef]
- Fraser, D.A.; Thoen, J.; Reseland, J.E.; Førre, O.; Kjeldsen-Kragh, J. Decreased CD4+ Lymphocyte Activation and Increased Interleukin-4 Production in Peripheral Blood of Rheumatoid Arthritis Patients after Acute Starvation. *Clin. Rheumatol.* 1999, 18, 394–401. [CrossRef]
- 8. Kjeldsen-Kragh, J.; Haugen, M.; Borchgrevink, C.F.; Laerum, E.; Eek, M.; Mowinkel, P.; Hovi, K.; Førre, O. Controlled Trial of Fasting and One-Year Vegetarian Diet in Rheumatoid Arthritis. *Lancet Lond. Engl.* **1991**, *338*, 899–902. [CrossRef]
- 9. Michalsen, A.; Riegert, M.; Lüdtke, R.; Bäcker, M.; Langhorst, J.; Schwickert, M.; Dobos, G.J. Mediterranean Diet or Extended Fasting's Influence on Changing the Intestinal Microflora, Immunoglobulin A Secretion and Clinical Outcome in Patients with Rheumatoid Arthritis and Fibromyalgia: An Observational Study. *BMC Complement. Altern. Med.* **2005**, *5*, 22. [CrossRef]
- 10. Hafström, I.; Ringertz, B.; Spångberg, A.; von Zweigbergk, L.; Brannemark, S.; Nylander, I.; Rönnelid, J.; Laasonen, L.; Klareskog, L. A Vegan Diet Free of Gluten Improves the Signs and Symptoms of Rheumatoid Arthritis: The Effects on Arthritis Correlate with a Reduction in Antibodies to Food Antigens. *Rheumatol. Oxf. Engl.* **2001**, *40*, 1175–1179. [CrossRef]
- 11. McDougall, J.; Bruce, B.; Spiller, G.; Westerdahl, J.; McDougall, M. Effects of a Very Low-Fat, Vegan Diet in Subjects with Rheumatoid Arthritis. J. Altern. Complement. Med. 2002, 8, 71–75. [CrossRef]
- 12. Abkevich, V.I.; Gutin, A.M.; Shakhnovich, E.I. Specific Nucleus as the Transition State for Protein Folding: Evidence from the Lattice Model. *Biochemistry* **1994**, *33*, 10026–10036. [CrossRef] [PubMed]
- 13. Dobson, C.M. Protein Folding and Misfolding. Nature 2003, 426, 884–890. [CrossRef] [PubMed]
- 14. Faísca, P.F.N. The Nucleation Mechanism of Protein Folding: A Survey of Computer Simulation Studies. *J. Phys. Condens. Matter* 2009, *21*, 373102. [CrossRef] [PubMed]
- 15. Kmiecik, S.; Kolinski, A. Folding Pathway of the B1 Domain of Protein G Explored by Multiscale Modeling. *Biophys. J.* **2008**, *94*, 726–736. [CrossRef]
- 16. Muñoz, V. Conformational Dynamics and Ensembles in Protein Folding. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 395–412. [CrossRef]
- 17. Efimov, A.V. A Novel Super-Secondary Structure of β-Proteins A Triple-Strand Corner. FEBS Lett. 1992, 298, 261–265. [CrossRef]
- 18. Efimov, A.V. A Novel Super-Secondary Structure of Proteins and the Relation between the Structure and the Amino Acid Sequence. *FEBS Lett.* **1984**, *166*, 33–38. [CrossRef]
- 19. Efimov, A.V. Structure of Coiled Beta-Beta-Hairpins and Beta-Beta-Corners. FEBS Lett. 1991, 284, 288–292. [CrossRef]
- 20. Chothia, C.; Janin, J. Orthogonal Packing of Beta-Pleated Sheets in Proteins. Biochemistry 1982, 21, 3955–3965. [CrossRef]

- 21. Boshkova, E.A.; Efimov, A.V. Structures Closed into Cycles in Proteins Containing 3β-Corners. *Biochem. Biokhimiia* **2010**, *75*, 1258–1263. [CrossRef] [PubMed]
- 22. Hammarström, A.; Berndt, K.D.; Sillard, R.; Adermann, K.; Otting, G. Solution Structure of a Naturally-Occurring Zinc-Peptide Complex Demonstrates That the N-Terminal Zinc-Binding Module of the Lasp-1 LIM Domain Is an Independent Folding Unit. *Biochemistry* **1996**, *35*, 12723–12732. [CrossRef] [PubMed]
- Drew, E.D.; Janes, R.W. 2StrucCompare: A Webserver for Visualizing Small but Noteworthy Differences between Protein Tertiary Structures through Interrogation of the Secondary Structure Content. *Nucleic Acids Res.* 2019, 47, W477–W481. [CrossRef] [PubMed]
- 24. Heinig, M.; Frishman, D. STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. *Nucleic Acids Res.* 2004, 32 (Suppl. 2), W500–W502. [CrossRef]
- 25. Arnittali, M.; Rissanou, A.N.; Amprazi, M.; Kokkinidis, M.; Harmandaris, V. Structure and Thermal Stability of WtRop and RM6 Proteins through All-Atom Molecular Dynamics Simulations and Experiments. *Int. J. Mol. Sci.* **2021**, *22*, 5931. [CrossRef]
- Pérez-Alvarado, G.C.; Miles, C.; Michelsen, J.W.; Louis, H.A.; Winge, D.R.; Beckerle, M.C.; Summers, M.F. Structure of the Carboxy-Terminal LIM Domain from the Cysteine Rich Protein CRP. *Nat. Struct. Biol.* 1994, 1, 388–398. [CrossRef]
- Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* 2015, 1–2, 19–25. [CrossRef]
- 28. Schmid, N.; Eichenberger, A.P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A.E.; van Gunsteren, W.F. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843. [CrossRef]
- 29. Nandel, F.S.; Garla, R. Conformational Behavior of Stereo Regular Substituted Polyglycolides Is Side Chain Dependent. *J. Biophys. Chem.* **2011**, *2*, 285–299. [CrossRef]
- 30. Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. J. Chem. Phys. 2007, 126, 014101. [CrossRef]
- Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, A.; Haak, J.R. Molecular Dynamics with Coupling to an External Bath. J. Chem. Phys. 1984, 81, 3684–3690. [CrossRef]
- Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: A Linear Constraint Solver for Molecular Simulations. J. Comput. Chem. 1997, 18, 1463–1472. [CrossRef]
- 33. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. [CrossRef]
- 34. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W.F.; Mark, A.E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999**, *38*, 236–240. [CrossRef]





Article Aurisin A Complexed with 2,6-Di-O-methyl-β-cyclodextrin Enhances Aqueous Solubility, Thermal Stability, and Antiproliferative Activity against Lung Cancer Cells

Thanapon Charoenwongpaiboon¹, Amy Oo², Sutita Nasoontorn³, Thanyada Rungrotmongkol^{2,4}, Somdej Kanokmedhakul⁵ and Panupong Mahalapbutr^{3,*}

- ¹ Department of Chemistry, Faculty of Science, Silpakorn University, Nakhon Pathom 73000, Thailand
- ² Center of Excellence in Structural and Computational Biology, Department of Biochemistry, Chulalongkorn University, Bangkok 10330, Thailand
- ³ Department of Biochemistry, Center for Translational Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand
- ⁴ Ph.D. Program in Bioinformatics and Computational Biology, Graduate School, Chulalongkorn University, Bangkok 10330, Thailand
- ⁵ Natural Products Research Unit, Department of Chemistry and Center for Innovation in Chemistry, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand
- * Correspondence: panupma@kku.ac.th; Tel.: +66-43-363265

Abstract: Aurisin A (AA), an aristolane dimer sesquiterpene isolated from the luminescent mushroom Neonothopanus nambi, exhibits various biological and pharmacological effects. However, its poor solubility limits its use for further medicinal applications. This study aimed to improve the water solubility of AA via complexation with β -cyclodextrin (β CD) and its derivatives (2,6-di-O-methyl- β CD (DM β CD) and 2-hydroxypropyl- β CD (HP β CD). A phase solubility analysis demonstrated that the solubility of AA linearly enhanced with increasing concentrations of β CDs (ranked in the order of AA/DM β CD > AA/HP β CD > AA/ β CD). Notably, β CDs, especially DM β CD, increased the thermal stability of the inclusion complexes. The thermodynamic study indicated that the complexation between AA and β CD(s) was a spontaneous endothermic reaction, and AA/DM β CD possesses the highest binding strength. The complex formation between AA and $DM\beta CD$ was confirmed by means of FT-IR, DSC, and SEM. Molecular dynamics simulations revealed that the stability and compactness of the AA/DM β CD complex were higher than those of the DM β CD alone. The encapsulation of AA led to increased intramolecular H-bond formations on the wider rim of DM β CD, enhancing the complex stability. The antiproliferative activity of AA against A549 and H1975 lung cancer cells was significantly improved by complexation with DMBCD. Altogether, the satisfactory water solubility, high thermal stability, and enhanced antitumor potential of the AA/DM β CD inclusion complex would be useful for its application as healthcare products or herbal medicines.

Keywords: Aurisin A; beta-cyclodextrins; inclusion complex; lung cancer

1. Introduction

Aurisin A (AA, Figure 1A) is an aristolane dimer sesquiterpene isolated from the luminescent mushroom *Neonothopanus nambi* Speg. (Marasmiaceae), which is normally found on logs or dead wood in broad-leaved forests in the northeast of Thailand [1]. This compound exhibits various biological and pharmacological activities, including antimycobacterial activity toward *Mycobacterium tuberculosis*, antimalarial property against *Plasmodium falciparum* [1], and anticancer potential toward lung cancer cells (NCI-H187 and A549) [1,2], breast cancer cells (BC1), epidermoid carcinoma cells (KB), cholangiocarcinoma cells (KKU-100, KKU-139, KKU-156, KKU-213, and KKU-214) [1], and cervical cancer cells (Hela, CaSki, and SiHa), with no cytotoxic effect on normal white blood cells [3]. AA exerts its anticancer effects by (i) inhibiting cancer cell growth and migration and (ii) inducing

Citation: Charoenwongpaiboon, T.; Oo, A.; Nasoontorn, S.; Rungrotmongkol, T.; Kanokmedhakul, S.; Mahalapbutr, P. Aurisin A Complexed with 2,6-Di-O-methyl-β-cyclodextrin Enhances Aqueous Solubility, Thermal Stability, and Antiproliferative Activity against Lung Cancer Cells. *Int. J. Mol. Sci.* 2022, 23, 9776. https://doi.org/ 10.3390/ijms23179776

Academic Editor: Wajid Zaman

Received: 21 July 2022 Accepted: 22 August 2022 Published: 29 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). cell cycle arrest and apoptosis through activating caspase-3/9 as well as decreasing the expression of cyclin D1, cyclin-dependent kinase 2/4 (Cdk-2/4), B-cell lymphoma 2 (Bcl-2), epidermal growth factor receptor (EGFR), phosphorylated p38 (pp38), and vascular endothelial growth factor (VEGF) [2,3]. Although AA possesses promising biological and pharmacological activities, its low water solubility limits its use for further applications as herbal medicines or healthcare products.



Figure 1. Chemical structures of (**A**) AA and (**B**) β CD and its derivatives (DM β CD and HP β CD), where the functional substitutions used in this study, are shown below.

Cyclodextrin (CD) is a cyclic oligosaccharide linked by α -1,4 glycosidic bonds. Natural CD consists of six, seven, and eight glucose units, namely alpha-cyclodextrin (α CD), beta-cyclodextrin (β CD), and gamma-cyclodextrin (γ CD), respectively [4]. The structural arrangement of CD turns out to be a truncated cone shape structure possessing a hydrophilic outer surface with a hydrophobic inner cavity. By hosting lipophilic guest molecules into the central cavity of CD, their physico-chemical properties are tremendously improved [5-7], making CD the most frequently used excipient in pharmaceutical applications [8]. Among the three natural CDs, β CD (Figure 1B) has been widely used because of its suitable cavity size, commercial availability, desirable drug loading capacity, biocompatibility, and low price [9,10]. However, due to the limited solubility of β CD, derivatives of β CD, such as 2-hydroxypropyl- β CD (HP β CD) and 2,6-di-O-methyl- β CD $(DM\beta CD)$, are developed to improve water solubility and reduce the limitations of the parent β CD [11,12]. Many lines of evidence have shown that the water solubility, chemical stability, and biological activity of poorly soluble compounds are significantly enhanced by complexation with β CD derivatives [13–17]. However, the information on the inclusion complexation between AA and β CD(s) has never been reported.

In the present study, we aimed to enhance the water solubility, stability, and anticancer activity of AA by inclusion complexation with β CD and its derivatives (DM β CD and HP β CD). The obtained inclusion complex was then confirmed experimentally and theoretically using physical and chemical characterization techniques as well as molecular modeling studies. In addition, the anticancer potential of the inclusion complex was evaluated. We hope that the improved physical and biological properties of AA/ β CD(s) inclusion complex could pave the way for the further development of AA as herbal medicines or healthcare products.

2. Results and Discussion

2.1. Phase Solubility, Thermodynamic Parameters, and UV-Vis Spectra Analyses

The phase solubility diagrams of AA in aqueous solutions of β CD, DM β CD, and HP β CD at 20, 30, 40, and 50 °C are shown in Figure 2. The obtained results revealed that the solubility of AA linearly increased with increasing concentrations of β CDs (ranked in the order of AA/DM β CD > AA/HP β CD > AA/ β CD). This linear relationship is a characteristic of A_L-type solubility, indicating a 1:1 host–guest complexation [18–20]. Next, the stability constant (K_c) was calculated from the phase solubility diagrams to estimate the binding strength of all the studied inclusion complexes. As shown in Table 1, the highest K_c value was found in AA/DM β CD (209–237 M⁻¹), followed by AA/HP β CD (88–148 M⁻¹) and AA/ β CD (42–80 M⁻¹), respectively. These findings are consistent with many lines of evidence demonstrating that β CD derivatives, especially DM β CD, could significantly improve the stability and solubility of several poorly soluble compounds [15,21–23]. Interestingly, the increased temperature remarkably enhanced the stability (the K_c index) of all the investigated complexes.



Figure 2. Phase solubility diagrams of AA in complex with β CD, DM β CD, and HP β CD at 20 °C, 30 °C, 40 °C, and 50 °C. Data are expressed as mean \pm SEM of three independent experiments.

Tomporature (°C)		K_c (M $^{-1}$)	
Temperature (°C)	AA/βCD	AA/DMβCD	ΑΑ/ΗΡβCD
20	42	209	88
30	60	219	112
40	73	232	135
50	80	237	148

Table 1. Stability constant (K_c) of the AA/ β CDs inclusion complexes at different temperatures.

To obtain the thermodynamic parameters (i.e., ΔH , ΔS , and ΔG) for the AA/ β CD(s) inclusion complexation process, the Van't Hoff plot based on Equation (3) was then employed (Figure S1). As depicted in Table 2, the ΔH values were positive for all systems, indicating that the inclusion complex formation was an endothermic process [14]. As expected, the inclusion complex formation between AA and β CD(s) was spontaneous, as evidenced by the negative sign of ΔG . The lowest ΔG value was detected in AA/DM β CD (-3.24 kcal/mol), followed by AA/HP β CD (-2.82 kcal/mol) and AA/ β CD (-2.42 kcal/mol), respectively, which is consistent well with the aforementioned K_c values (Table 1).
Thermodynamic Parameters (kcal/mol) ^a	AA/βCD	AA/DMβCD	ΑΑ/ΗΡβCD
ΔΗ	4.05	0.84	3.32
$T\Delta S$	6.47	4.08	6.14
ΔG	-2.42	-3.24	-2.82

Table 2. Thermodynamic values for the inclusion complexation between AA and β CD(s).

^a Data were derived from Van't Hoff plots using R = 1.985×10^{-3} kcal·mol⁻¹·K⁻¹ and T = 303 K.

According to the UV-Vis spectra analysis (Figure 3), we found that the maximum absorption of AA (325 nm) bathochromically shifted to 328–332 nm in all the studied complexes, indicating a possible interaction between the AA and β CD(s). Similar bathochromic shifts of ligands after complexation were also found in luteolin and *trans*-ferulic acid in complex with the β CD derivatives [24,25]. Notably, the highest absorbance was detected in the AA/DM β CD complex (0.91), followed by AA/HP β CD (0.53), AA/ β CD (0.43), and AA alone (0.27), which is consistent well with the phase solubility study mentioned above (Figure 2).



Figure 3. UV-Vis spectra of AA, inclusion complexes, and the free form of β CDs.

Taken together, only the AA/DM β CD complex possessing the highest stability and solubility was selected for further structural characterizations in comparison with the free form of AA and DM β CD.

2.2. Inclusion Complex Characterization 2.2.1. FT-IR

FT-IR spectroscopy was used to determine the inclusion complex formation between AA and DM β CD. The obtained FT-IR spectra of the AA, DM β CD, and AA/DM β CD complex are shown in Figure 4. The spectrogram of AA showed characteristic stretching vibration peaks at (i) 1669 and 1634 cm⁻¹ for C=O, (ii) 1561 cm⁻¹ for C=C, (iii) 2954 cm⁻¹ for C–H, (iv) 1272 and 1206 cm⁻¹ for C–O, and (v) 3554 cm⁻¹ for O–H [1]. The FT-IR spectrum of DM β CD demonstrated a large band at 3399 cm⁻¹ (O–H stretching), 2923 cm⁻¹ (C–H stretching), and 1156, 1085, and 1045 cm⁻¹ (C–O and C–H stretching) [26]. After the inclusion complexation, the FT-IR spectrum of AA/DM β CD was distinctly different from that of the pure AA and DM β CD. The characteristic stretching vibration peaks of AA

at 1272, 1206, 1561, 1634, and 1669 cm⁻¹ totally disappeared in the FT-IR spectrum of the inclusion complex, which is similar to other reported hydrophobic compounds in complex with the β CD analogs [14,25,27]. This might be due to a restriction of the AA's C=O and C=C stretching vibrations, as well as a modification of the hydrophobic microenvironment inside the DM β CD cavity [28]. Moreover, the changes in the shape and position of the absorption bands of AA and DM β CD were observed in the AA/DM β CD complex. The vibration peaks of the C–H and O–H stretching of AA (2954 and 3554 cm⁻¹) and DM β CD (2923 and 3399 cm⁻¹) were shifted to 2925 and 3402 cm⁻¹ in the solid complex. Similarly, the vibration peaks of the C–O and C–H stretching of DM β CD (1156, 1085, and 1045 cm⁻¹) were redshifted to 1155, 1084, and 1044 cm⁻¹ after the complex formation. Altogether, these FT-IR results indicated that AA was completely embedded in the hydrophobic cavity of DM β CD, which is supported by the differential scanning calorimetry (DSC), scanning electron microscope (SEM), and molecular dynamics (MD) simulation results, as discussed later.



Figure 4. FT-IR spectra of the AA, DMβCD, and AA/DMβCD complex.

2.2.2. Thermal Analysis

The thermal properties of the AA, DM β CD, and AA/DM β CD complex were characterized in a solid state using DSC analysis. As shown in Figure 5, the characteristic endothermic/exothermic peaks of the free compounds were as follows: (i) AA at 93.0, 213.9, 229.7, and 284.1 °C and (ii) DM β CD at 51.8 °C. The endothermic peaks found at 213.9 and 229.7 °C corresponded to the melting point of AA, as previously reported [1], whereas the broad endothermic peak of DM β CD detected at 51.8 °C indicates the release of water molecules from the DM β CD's hydrophobic inner cavity [29]. In the thermogram of the freeze-dried AA/DM β CD inclusion complex, the characteristic thermal peaks of AA and DM β CD totally disappeared, coinciding with the appearance of a new endothermic peak at 72.0 °C and an exothermic peak at 198.7 °C, similar to other reports [15,16,23,26,30]. These findings indicated that the freeze-drying method successfully yielded the new solid phase between AA and $DM\beta CD$.



Figure 5. DSC thermograms of the AA, DM_βCD, and AA/DM_βCD complex.

2.2.3. SEM

Many lines of evidence have shown that the inclusion complexation process significantly changes the surface textures of the resulting products [27,31–33]. To visualize the surface morphology of all the studied compounds, the SEM technique was employed. The SEM photographs of the AA, DM β CD, and AA/DM β CD complex are given in Figure 6. Both AA and DM β CD presented a rod-like structure [23,30]; however, the particle size and shape of the AA were bigger and more spherical than those of the DM β CD. Upon molecular complexation, the surface morphology of the obtained freeze-dried AA/DM β CD inclusion complex, appearing as a plate-like structure, was different from that of the pure forms. These findings confirmed the successful formation between AA and DM β CD. Taken together, all of the structural characterization results (Figures 4–6) revealed that the inclusion complex between AA and DM β CD was successfully formed.



Figure 6. SEM photographs of the AA, DMβCD, and AA/DMβCD complex.

2.3. Molecular Modeling Studies

To further verify the aforementioned experimental results and to investigate the dynamic behavior of the $AA/DM\beta CD$ inclusion complex at the atomic level, all-atom MD

simulations in an aqueous solution and free energy calculations based on the molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) were performed.

2.3.1. System Stability

The stability of DM β CD and its inclusion complex, AA/DM β CD, along the simulation time was determined using the calculated time evolution of root-mean-square displacement (RMSD), radius of gyration (Rg), and number of atomic contacts (# Atom contacts). As shown in Figure 7A, the RMSD values of DM β CD (~2–3 Å) were higher than those of the AA/DM β CD complex (~1 Å), suggesting that the AA/DM β CD is more stable than the uncomplexed DM β CD. Similarly, the Rg values of DM β CD (~6.6–7.0 Å) were larger than those of the AA/DM β CD complex (~6.4–6.6 Å), indicating higher compactness of the AA/DM β CD structure, as evidenced by the final MD snapshots (Figure 7B). Although the # Atom contacts (native + nonnative) was highly stable along the simulation times (~100–120) for the three independent simulations, the high RMSD fluctuations at the first 150 ns of the simulations were detected. Therefore, in this work, the MD trajectories from 200-300 ns were extracted for further structural and energetic analyses. From Figure 7B, we found that the AA molecule was completely embedded in the hydrophobic cavity of DM β CD, where the C=O groups of AA at C1 and C1' (Figure 1A) were located at the center of DM β CD's cavity, while those at C8 and C8' were positioned near the secondary and primary rims of $DM\beta CD$. This complete formation of the AA/DM βCD complex is consistent well with the results of the inclusion complex characterization, as mentioned above.



Figure 7. (**A**) Time evolution of RMSD, Rg, and # Atom contacts of DMβCD and AA/DMβCD for three independent simulations (MD1–3). (**B**) Final MD snapshot of DMβCD and AA/DMβCD.

2.3.2. DMBCD Conformation upon AA Binding

The conformational changes of DM β CD upon AA encapsulation were investigated by calculating (i) the distance of the oxygen atoms on the wider rim of DM β CD (O3_(n)– O2_(n+1), dO₃₋₂), corresponding to a possibility of an intramolecular hydrogen bond (H-bond) formation (dO₃₋₂ \leq 3.5 Å), and (ii) the distance of glycosidic oxygen atoms (O4_(n)–O4_(n+1), dO₄₋₄). Afterward, these two parameters were converted to the free energy value, F(*x*,*y*), using Equation (1):

$$F(x,y) = -k_{\rm B}T \log[P(x,y)] \tag{1}$$

where k_B is the Boltzmann constant, *T* is the temperature (303 K), and P(x,y) is the probability of dO₃₋₂ (*x*) and dO₄₋₄ (*y*). The obtained 2D free energy landscape is shown in Figure 8. When compared to the unbound form of DM β CD, the molecular encapsulation of AA in the DM β CD could enhance the formation of intramolecular H-bonds on the wider rim of DM β CD, as evidenced by the increased population of dO₃₋₂ at ~3.0–3.5 Å. The H-bondoperated conformational changes of DM β CD upon the AA encapsulation are similar to other reported mansonones, pinostrobin, luteolin, pinocembrin, and neral in complex with various β CD derivatives [23,34–37]. In addition, the populations of dO₃₋₂ at ~3.0–5.0 Å and dO₄₋₄ at ~3.0–4.0 Å of the free form of DM β CD totally disappeared in the AA/DM β CD complex due to the adaptation of the DM β CD structure upon the insertion of AA to the hydrophobic cavity.



Figure 8. 2D free energy (F(x,y)) landscape of the intramolecular hydrogen bond distances, $O3_{(n)}$ – $O2_{(n+1)}$, against the adjacent glycosidic oxygen distances, $O4_{(n)}$ – $O4_{(n+1)}$, of the DM β CD (**top**) and AA/DM β CD complex (**bottom**) for three independent simulations (MD1–3).

2.3.3. Water Accessibility toward the Inclusion Complex

The water distribution around a spherical radius r of the oxygen atoms of AA (Figure 9A) was visualized using radial distribution function (RDF, g(r)) calculation, and the obtained results are given in Figure 9B. In addition, the integration number (n(r)) values at the first minima, corresponding to the number of water molecules approaching the targeted oxygens, are depicted in Table 3. From the RDF plots of all the systems, no dominant peak was detected within ~3 Å of the O, O1, O2, and O2' of AA (Figure 9B), indicating that these oxygen atoms were deeply embedded in the hydrophobic inner cavity of

DM β CD (Figure 9A). This phenomenon is in good agreement with the previously reported flavonoids, demonstrating that their oxygen atom on the chromone ring, embedded at the center of the β CD cavity, displayed no sharp RDF peak at the first solvation shell [36,38]. In contrast, the other oxygen atoms (O1', O8, O8', O9, and O9') displayed the first sharp peak at ~2.5 Å, corresponding to the water distribution around these oxygens. The O8 and O8' of AA exhibited higher water accessibility than the other oxygens, suggesting that these oxygen atoms were positioned nearby either the secondary or primary rim and were feasibly accessible to the water molecules (Figure 9A).



Figure 9. (**A**) Final MD snapshot of the AA/DMβCD complex showing the oxygen atoms (O, O1, O1', O2, O2', O8, O8', O9, and O9') of AA and the surrounding water molecules (cyan dot) within 5 Å of AA. (**B**) RDF of water oxygen atoms around the oxygen atoms of AA in complex with DMβCD for three independent simulations (MD1–3).

Table 3. n(r) up to the first minimum for the oxygen atoms of AA inside the DM β CD hydrophobic cavity.

	0	01	O 1′	O2	O 2′	O 8	O 8′	O9	O9′
<i>n</i> (<i>r</i>) ^a	-	-	0.96 ± 0.02	-	-	5.77 ± 0.43	4.74 ± 0.18	4.15 ± 0.16	3.07 ± 0.11
^a Data d	erived	d from	Figure 9 are sho	wn as	mean	\pm SEM ($n = 3$).			

2.3.4. Binding Affinity of the Inclusion Complex

To estimate the binding affinity of the AA/DM β CD inclusion complex, the MM/PBSA calculation was performed using 100 snapshots taken from the last 100 ns MD simulations. As expected, due to the poor solubility of AA, the inclusion complexation in the gas phase was driven mainly by van der Waals (vdW) interactions ($\Delta E_{vdW} = -35.12 \pm 1.06$ kcal/mol) rather than electrostatic attraction ($\Delta E_{ele} = -17.05 \pm 0.19$ kcal/mol). Similarly, the summation of $\Delta G_{solv,non-polar} + \Delta E_{vdW}$ energies (-40.64 ± 1.12 kcal/mol) showed a negative value compared to that of $\Delta G_{solv,polar} + \Delta E_{ele}$ energies (12.64 ± 0.28 kcal/mol), indicating that the vdW forces play an important role in the complex formation between AA and DM β CD in an aqueous environment. This vdW-driven host–guest complexation process is consistent well with other lipophilic ligands in complex with β CDs [39–42]. Notably, the predicted $\Delta G_{bind,MM/PBSA}$ value (-3.87 ± 0.68 kcal/mol) was almost identical to the experimental ΔG (ΔG_{exp} , -3.24 kcal/mol) obtained from the Van't Hoff plot (Table 4), suggesting the successful calculation of the AA/DM β CD complex.

2.4. DM_βCD Enhances Cytotoxicity of AA against Lung Cancer Cells

The cytotoxic activity of AA and the AA/DM β CD inclusion complex against A549 and H1975 human lung cancer cells was evaluated using MTT assay. The obtained results are depicted in Figure 10. We found that both AA and AA/DM β CD decreased cell viability in a dose-dependent manner, in which the AA/DM β CD complex exhibited significantly lower cell viability than the uncomplexed AA at the concentration of 1, 3, and 10 μ M for both A549 and H1975 cell lines (Figure 10A,B). The half-maximal inhibitory concentration

(IC₅₀) values against the A549 and H1975 cells of the AA/DM β CD inclusion complex (17.14 \pm 2.34 and 15.67 \pm 1.33 μ M) were significantly lower than those of the AA alone (32.77 \pm 2.94 and 27.38 \pm 3.17 μ M) (Figure 10C). These findings are in good agreement with previous reports demonstrating that β CDs can distinctly enhance the anticancer activity of several hydrophobic compounds, such as camptothecin, luotonin A, resveratrol, mansonone G, curcumin, and scutellarein [23,43–46]. Thus, it was assumed that the enhanced antitumor effect of the AA/DM β CD inclusion complex was due to the improved water solubility and complex stability (Figure 2 and Table 1). In addition, DM β CD could infiltrate into the drug permeation barrier, called the unstirred water layer (UWL) [47], better than the uncomplexed AA, enhancing the flux of AA through the UWL [48].

Energy Component (kcal/mol) ^a	ΑΑ/DMβCD
$\Delta E_{\rm vdW}$	-35.12 ± 1.06
$\Delta E_{ m ele}$	-17.05 ± 0.19
$\Delta E_{\rm MM}$	-52.18 ± 0.92
$\Delta G_{ m solv,polar}$	29.69 ± 0.16
$\Delta G_{ m solv,non-polar}$	-5.51 ± 0.06
$\Delta G_{ m solv}$	24.18 ± 0.19
$\Delta G_{\rm solv,polar} + \Delta E_{\rm ele}$	12.64 ± 0.28
$\Delta G_{\rm solv,non-polar} + \Delta E_{\rm vdW}$	-40.64 ± 1.12
ΤΔS	-24.12 ± 0.16
$\Delta G_{\rm bind,MM/PBSA}$	-3.87 ± 0.68
ΔG_{exp}^{b}	-3.24

Table 4. $\Delta G_{\text{bind}, \text{MM}/\text{PBSA}}$ and its energy components (kcal/mol) of the AA/DM β CD complex.

^a Data are shown as mean \pm SEM (n = 3). ΔE_{MM} , molecular mechanics energy; ΔG_{solv} , solvation free energy comprising polar ($\Delta G_{\text{solv,polar}}$) and non-polar ($\Delta G_{\text{solv,non-polar}}$) terms; ΔS , entropy. ^b Data from Table 2.



Figure 10. Cell viability of (**A**) A549 and (**B**) H1975 human lung cancer cells after being treated with various concentrations of AA and AA/DM β CD for 48 h. The viable cells in the vehicle control (0.2% DMSO) were calculated as 100%. (**C**) IC₅₀ of the AA and AA/DM β CD complex against the A549 and H1975 cells. * *p* < 0.05, ** *p* < 0.01 vs. AA. Data are shown as mean ± SEM (*n* = 3).

3. Materials and Methods

3.1. Materials

AA was extracted from the culture liquid of the luminescent mushroom *Neonothopanus nambi* PW1 (Marasmiaceae), as previously described [1,3]. βCD and HPβCD were purchased from TCI (Nihonbashi-honcho, CK, Tokyo), whereas DMβCD was purchased from Sigma-Aldrich (St. Louis, MO, USA).

3.2. Phase Solubility Study

The phase solubility study was performed according to the methods described by Higushi and Connors [49]. An excess amount of AA was added to aqueous solutions containing increasing amounts of β CD(s) (0–10 mM). The mixtures were incubated in a shaking incubator at 20, 30, 40, and 50 °C and 250 rpm for 40 h. After that, the insoluble AA was separated from the suspension by centrifugation at 10,000 rpm for 5 min and then filtered by a 0.45-micron syringe filter [13,23,44,50–52]. Two volumes of ethanol were added into each inclusion complex solution before measuring the absorbance at 331 nm [3]. The apparent stability constant (K_c) was determined by Equation (2), where S₀ is the y-intercept.

$$K_c = \frac{\text{Slope}}{S_0(1 - \text{slope})} \tag{2}$$

The Van't Hoff equation (Equation (3)) was used to calculate the change in the enthalpy (ΔH) and entropy (ΔS) of the inclusion complexation, whereas the Gibbs free energy (ΔG) was determined by Equation (4).

$$\ln K_c = -\frac{\Delta H}{RT} + \frac{\Delta H}{R} \tag{3}$$

$$\Delta G = \Delta H - T \Delta S \tag{4}$$

3.3. Inclusion Complex Preparation

An excess amount of AA was added to a 10 mM DM β CD solution and incubated in a shaking incubator at 30 °C at 250 rpm. After that, the suspension was centrifuged (12,000 rpm for 15 min) and filtered through the 0.45-micron syringe filter, and lyophilized. The obtained freeze-dried powders were kept in a desiccator for further analysis.

3.4. Inclusion Complex Characterization

3.4.1. Ultraviolet-Visible (UV-Vis) Spectroscopy

AA and its inclusion complexes were suspended in DI water at 30 °C for 48 h. After that, the suspension was filtered using the 0.45-micron syringe filter. The UV-Vis spectra of the solutions were recorded by Eppendorf BioSpectrometer[™] (Eppendorf, Hamburgm, Germany).

3.4.2. Fourier Transform Infrared (FT-IR) Spectroscopy

The FT-IR spectra of the AA, DM β CD, and AA/DM β CD complex were recorded by a Nicolet 6700 FT-IR spectrometer (ThemoFisher Scientific, Waltham, MA, USA) over a scanning range of 500–4000 cm⁻¹ via the attenuated total reflectance (ATR) mode.

3.4.3. Differential Scanning Calorimetry (DSC)

The thermal behavior of the AA, DM β CD, and AA/DM β CD complex was characterized using NETZSCH DSC 204F1 Phoenix (Selb, Germany). Each solid sample (~1–2 mg) was heated from 25 °C to 300 °C in aluminum pans at a rate of 10 °C/min.

3.4.4. Scanning Electron Microscope (SEM)

The surface morphology of the AA, DMβCD, and AA/DMβCD complex was analyzed using a Scanning Electron Microscope (JEOL JSM-IT500HR, Tokyo, Japan). Samples were

coated with a thin layer of gold in a vacuum before viewing under 300 times magnification. Observations were performed using an accelerating voltage of 10 kV.

3.5. Computational Details

3.5.1. System Preparation and Molecular Docking

The 3D structure of DM β CD was taken from a previous study [23], whereas that of AA was downloaded from the PubChem database (PubChem CID: 71491081) and then optimized by the Gaussian09 program (Wallingford, CT, USA) [53] using the HF/6-31G* level of theory. The protonation state of AA was checked at a pH of 7.0 using MarvinSketch software (Budapest, Hungary). The inclusion complex model between the optimized AA and the DM β CD was generated using the CDOCKER module implemented in Accelrys Discovery Studio 2.5 (Accelrys Software Inc., San Diego, CA, USA). Among the resulting 100 docking poses, the AA/DM β CD inclusion complex with the lowest CDOCKER interaction energy was selected for further studies.

3.5.2. Molecular Dynamics (MD) Simulations

The MD simulations with the isothermal-isobaric ensemble (NPT) of each system were performed with a time step of 2 fs using an AMBER16 software package [54]. According to the standard procedures [36,55,56], the electrostatic potential (ESP) charges of AA were calculated with the HF/6-31(d) level of theory using the antechamber module, whereas the restrained ESP (RESP) charges of AA were computed using the parmchk module in AMBER16. The SHAKE algorithm [57] was applied to constrain all chemical bonds involving hydrogen atoms, while the Particle Mesh Ewald [58] method was used to treat long-range electrostatic interactions. The cutoff value for non-bonded interactions was set to 12 Å. The general AMBER force field (GAFF) [59] and the Glycam-06 force field [60] were applied on AA and DM β CD, respectively. The TIP3P water molecules [61] were added to solvate the inclusion complex with a spacing distance of 15 Å. Subsequently, the water molecules were minimized using the steepest descent (1500 steps) and conjugated gradient (3000 steps), followed by the minimization of the whole system. Each studied system was heated up from 10 K to 298 K for 100 ps and then equilibrated for 1000 ps. After that, all-atom MD simulations were performed under a periodic boundary condition at 1 atm and 298 K until reaching 300 ns. The MD simulations were performed in three replicates (n = 3) for each model.

3.5.3. Structural and Energetic Analyses

The CPPTRAJ module of AMBER16 was used to calculate the structural information, including the RMSD, Rg, # Atom contacts, free energy landscape, and RDF. For the energetic analysis, the binding affinity between the host and guest was calculated by the MM/PBSA method [62] using 100 snapshots extracted from the last 100 ns MD simulations.

3.6. Cell Lines and Culture

A549 and H1975 human lung cancer cells were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA). Both cells were cultured in a Dulbecco's Modified Eagle's Medium (Gibco, NY, USA) supplemented with a 10% heat-inactivated fetal bovine serum (Gibco, NY, USA), 100 U/mL penicillin, and 100 μ g/mL streptomycin (Gibco, NY, USA) and were maintained at 37 °C in a humidified 5% CO₂ atmosphere.

3.7. Cell Viability Assay

The A549 and H1975 cells were seeded into 96-well plates at a density of 1000 cells/well. After overnight incubation, the cells were treated with logarithmic concentrations (1, 3, 10, 30, and 100 μ M) of AA and AA/DM β CD for 48 h. Note that the amount of MG in free form and in complex form was equivalent. An MTT reagent was then added to the wells and incubated for 3 h. Subsequently, the culture medium was withdrawn, and 100 μ L of a

DMSO solution was added to dissolve the formazan crystals. Finally, the absorbance was measured at 540 nm.

3.8. Statistical Analysis

Data are shown as mean \pm standard error of the mean (SEM) of three independent experiments. Differences between AA and AA/DM β CD were determined using the *t* test. A *p* value of <0.05 was considered statistically significant.

4. Conclusions

This study aimed to improve the water solubility and biological activity of AA by complexation with β CD and its derivatives (DM β CD and HP β CD). The phase solubility diagrams indicated 1:1 AA/ β CD(s) binding stoichiometry, and the highest K_c was detected in the AA/DM β CD complex. Notably, β CDs, especially DM β CD, increased the thermal stability of the complexes. The thermodynamic study indicated that the inclusion complexation between AA and β CD(s) was a spontaneous endothermic reaction. The complex formation of the AA/DM β CD was confirmed by UV-Vis, FT-IR, DSC, and SEM techniques. MD simulations and MM/PBSA-based free energy calculations affirmed the vdW-driven formation of the AA/DM β CD complex in an aqueous environment. The anticancer effect of AA on A549 and H1975 lung cancer cells was significantly improved by complexation with DM β CD. Taken together, the satisfactory water solubility, high thermal stability, and enhanced antitumor potential of the AA/DM β CD complex would be potentially useful for its application as herbal medicines or healthcare products.

Supplementary Materials: The supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23179776/s1.

Author Contributions: T.C.: methodology, formal analysis, investigation, validation, and visualization. A.O.: formal analysis, investigation, validation, and visualization. S.N.: investigation, validation, and visualization. T.R.: resources, validation, and visualization. S.K.: purification, validation, and visualization. P.M.: conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft, writing—review & editing, validation, visualization, project administration, and supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the "Research and Graduate studies" of Khon Kaen University (RP64-Mushroom extract-001). P.M. would like to thank the Thailand Toray Science Foundation for financial support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article or Supplementary Material.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Kanokmedhakul, S.; Lekphrom, R.; Kanokmedhakul, K.; Hahnvajanawong, C.; Bua-art, S.; Saksirirat, W.; Prabpai, S.; Kongsaeree, P. Cytotoxic sesquiterpenes from luminescent mushroom Neonothopanus nambi. *Tetrahedron* 2012, 68, 8261–8266. [CrossRef]
- 2. Boueroy, P.; Boonmars, T.; Kanokmedhakul, S.; Chareonsudjai, S.; Lekphrom, R.; Srichangwang, S. Promising Anticancer Effect of Aurisin A Against the Human Lung Cancer A549 Cell Line. *Asian Pac. J. Cancer Prev.* **2020**, *21*, 49–54. [CrossRef]
- Boueroy, P.; Kanokmedhakul, S.; Pientong, C.; Ekalaksananan, T.; Saksirirat, W.; Ratanasuwan, P.; Lekphrom, R.; Srichangwang, S. Anticancer effects of aurisin A extracts from Neonothopanus nambi on human papillomavirus-infected cervical cancer cells. *Agric. Nat. Resour.* 2021, 55, 618–627.
- 4. Szejtli, J. Introduction and general overview of cyclodextrin chemistry. Chem. Rev. 1998, 98, 1743–1754. [CrossRef]
- Soe, H.M.H.; Chamni, S.; Mahalapbutr, P.; Kongtaworn, N.; Rungrotmongkol, T.; Jansook, P. The investigation of binary and ternary sulfobutylether-β-cyclodextrin inclusion complexes with asiaticoside in solution and in solid state. *Carbohydr. Res.* 2020, 498, 108190. [CrossRef]

- 6. Jiang, L.; Yang, J.; Wang, Q.; Ren, L.; Zhou, J. Physicochemical properties of catechin/β-cyclodextrin inclusion complex obtained via co-precipitation. *CyTA-J. Food* **2019**, *17*, 544–551. [CrossRef]
- Calabrò, M.L.; Tommasini, S.; Donato, P.; Raneri, D.; Stancanelli, R.; Ficarra, P.; Ficarra, R.; Costa, C.; Catania, S.; Rustichelli, C.; et al. Effects of α- and β-cyclodextrin complexation on the physico-chemical properties and antioxidant activity of some 3-hydroxyflavones. *J. Pharm. Biomed. Anal.* 2004, *35*, 365–377. [CrossRef]
- 8. Conceição, J.; Adeoye, O.; Cabral-Marques, H.M.; Lobo, J.M.S. Cyclodextrins as excipients in tablet formulations. *Drug Discov. Today* **2018**, *23*, 1274–1284. [CrossRef]
- 9. Del Valle, E.M.M. Cyclodextrins and their uses: A review. Process Biochem. 2004, 39, 1033–1046. [CrossRef]
- 10. Tommasini, S.; Raneri, D.; Ficarra, R.; Calabrò, M.L.; Stancanelli, R.; Ficarra, P. Improvement in solubility and dissolution rate of flavonoids by complexation with β-cyclodextrin. *J. Pharm. Biomed.* **2004**, *35*, 379–387. [CrossRef]
- Gould, S.; Scott, R.C. 2-Hydroxypropyl-β-cyclodextrin (HP-β-CD): A toxicology review. *Food Chem. Toxicol.* 2005, 43, 1451–1459. [CrossRef]
- 12. Szejtli, J. Dimethyl-β-cyclodextrin as parenteral drug carrier. J. Incl. Phenom. 1983, 1, 135–150. [CrossRef]
- Kicuntod, J.; Sangpheak, K.; Mueller, M.; Wolschann, P.; Viernstein, H.; Yanaka, S.; Kato, K.; Chavasiri, W.; Pongsawasdi, P.; Kungwan, N.; et al. Theoretical and Experimental Studies on Inclusion Complexes of Pinostrobin and β-Cyclodextrins. *Sci. Pharm.* 2018, *86*, 5. [CrossRef]
- 14. Tang, P.; Li, S.; Wang, L.; Yang, H.; Yan, J.; Li, H. Inclusion complexes of chlorzoxazone with β- and hydroxypropyl-β-cyclodextrin: Characterization, dissolution, and cytotoxicity. *Carbohydr. Polym.* **2015**, *131*, 297–305. [CrossRef]
- 15. Wang, L.; Li, S.; Tang, P.; Yan, J.; Xu, K.; Li, H. Characterization and evaluation of synthetic riluzole with β-cyclodextrin and 2,6-di-O-methyl-β-cyclodextrin inclusion complexes. *Carbohydr. Polym.* **2015**, *129*, 9–16. [CrossRef]
- 16. Yang, L.-J.; Ma, S.-X.; Zhou, S.-Y.; Chen, W.; Yuan, M.-W.; Yin, Y.-Q.; Yang, X.-D. Preparation and characterization of inclusion complexes of naringenin with β-cyclodextrin or its derivative. *Carbohydr. Polym.* **2013**, *98*, 861–869. [CrossRef]
- 17. Oo, A.; Kerdpol, K.; Mahalapbutr, P.; Rungrotmongkol, T. Molecular encapsulation of emodin with various β-cyclodextrin derivatives: A computational study. *J. Mol. Liq.* **2022**, *347*, 118002. [CrossRef]
- Liu, M.; Dong, L.; Chen, A.; Zheng, Y.; Sun, D.; Wang, X.; Wang, B. Inclusion complexes of quercetin with three β-cyclodextrins derivatives at physiological pH: Spectroscopic study and antioxidant activity. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 2013, *115*, 854–860. [CrossRef]
- Santos, P.S.; Souza, L.K.M.; Araújo, T.S.L.; Medeiros, J.V.R.; Nunes, S.C.C.; Carvalho, R.A.; Pais, A.C.C.; Veiga, F.J.B.; Nunes, L.C.C.; Figueiras, A. Methyl-β-cyclodextrin Inclusion Complex with β-Caryophyllene: Preparation, Characterization, and Improvement of Pharmacological Activities. ACS Omega 2017, 2, 9080–9094. [CrossRef]
- 20. Gao, S.; Bie, C.; Ji, Q.; Ling, H.; Li, C.; Fu, Y.; Zhao, L.; Ye, F. Preparation and characterization of cyanazine–hydroxypropyl-betacyclodextrin inclusion complex. *RSC Adv.* 2019, *9*, 26109–26115. [CrossRef]
- 21. Kikuchi, M.; Uekama, K. Effect of dimethyl beta-cyclodextrin on oral or rectal absorption of 1-hexylcarbamoyl-5-fluorouracil (HCFU). *Yakugaku Zasshi J. Pharm. Soc. Jpn.* **1988**, *108*, 156–163. [CrossRef]
- Tang, P.; Wang, L.; Ma, X.; Xu, K.; Xiong, X.; Liao, X.; Li, H. Characterization and In Vitro Evaluation of the Complexes of Posaconazole with β- and 2,6-di-O-methyl-β-cyclodextrin. AAPS PharmSciTech 2017, 18, 104–114. [CrossRef]
- Mahalapbutr, P.; Wonganan, P.; Charoenwongpaiboon, T.; Prousoontorn, M.; Chavasiri, W.; Rungrotmongkol, T. Enhanced Solubility and Anticancer Potential of Mansonone G By β-Cyclodextrin-Based Host-Guest Complexation: A Computational and Experimental Study. *Biomolecules* 2019, *9*, 545. [CrossRef]
- 24. Liu, B.; Li, W.; Zhao, J.; Liu, Y.; Zhu, X.; Liang, G. Physicochemical characterisation of the supramolecular structure of luteolin/cyclodextrin inclusion complex. *Food Chem.* **2013**, *141*, 900–906. [CrossRef]
- Wang, J.; Cao, Y.; Sun, B.; Wang, C. Characterisation of inclusion complex of trans-ferulic acid and hydroxypropyl-β-cyclodextrin. *Food Chem.* 2011, 124, 1069–1075. [CrossRef]
- 26. Wang, D.-W.; Ouyang, C.-B.; Liu, Q.; Yuan, H.-L.; Liu, X.-H. Inclusion of quinestrol and 2,6-di-O-methyl-β-cyclodextrin: Preparation, characterization, and inclusion mode. *Carbohydr. Polym.* **2013**, *93*, 753–760. [CrossRef]
- 27. Rajendiran, N.; Siva, S. Inclusion complex of sulfadimethoxine with cyclodextrins: Preparation and characterization. *Carbohydr. Polym.* **2014**, *101*, 828–836. [CrossRef]
- Vukic, M.D.; Vukovic, N.L.; Popovic, S.L.; Todorovic, D.V.; Djurdjevic, P.M.; Matic, S.D.; Mitrovic, M.M.; Popovic, A.M.; Kacaniova, M.M.; Baskic, D.D. Effect of β-cyclodextrin encapsulation on cytotoxic activity of acetylshikonin against HCT-116 and MDA-MB-231 cancer cell lines. *Saudi Pharm. J. SPJ Off. Publ. Saudi Pharm. Soc.* 2020, *28*, 136–146. [CrossRef]
- 29. Gu, F.; Ning, J.; Fan, H.; Wu, C.; Wang, Y. Preparation and characterization of simvastatin/DMβCD complex and its pharmacokinetics in rats. *Acta Pharm.* **2018**, *68*, 145–157. [CrossRef]
- 30. Choi, S.G.; Lee, S.-E.; Kang, B.-S.; Ng, C.L.; Davaa, E.; Park, J.-S. Thermosensitive and Mucoadhesive Sol-Gel Composites of Paclitaxel/Dimethyl-β-Cyclodextrin for Buccal Delivery. *PLoS ONE* **2014**, *9*, e109090. [CrossRef]
- 31. Rakmai, J.; Cheirsilp, B.; Mejuto, J.C.; Simal-Gándara, J.; Torrado-Agrasar, A. Antioxidant and antimicrobial properties of encapsulated guava leaf oil in hydroxypropyl-beta-cyclodextrin. *Ind. Crops Prod.* **2018**, *111*, 219–225. [CrossRef]
- Yallapu, M.M.; Jaggi, M.; Chauhan, S.C. β-Cyclodextrin-curcumin self-assembly enhances curcumin delivery in prostate cancer cells. *Colloids Surf. B Biointerfaces* 2010, 79, 113–125. [CrossRef]

- 33. Zhou, Q.; Wei, X.; Dou, W.; Chou, G.; Wang, Z. Preparation and characterization of inclusion complexes formed between baicalein and cyclodextrins. *Carbohydr. Polym.* **2013**, *95*, 733–739. [CrossRef]
- 34. Mahalapbutr, P.; Nutho, B.; Wolschann, P.; Chavasiri, W.; Kungwan, N.; Rungrotmongkol, T. Molecular insights into inclusion complexes of mansonone E and H enantiomers with various β-cyclodextrins. *J. Mol. Graph. Model.* **2018**, *79*, 72–80. [CrossRef]
- 35. Kicuntod, J.; Khuntawee, W.; Wolschann, P.; Pongsawasdi, P.; Chavasiri, W.; Kungwan, N.; Rungrotmongkol, T. Inclusion complexation of pinostrobin with various cyclodextrin derivatives. *J. Mol. Graph. Model.* **2016**, *63*, 91–98. [CrossRef]
- Mahalapbutr, P.; Thitinanthavet, K.; Kedkham, T.; Nguyen, H.; Theu, L.t.h.; Dokmaisrijan, S.; Huynh, L.; Kungwan, N.; Rungrotmongkol, T. A theoretical study on the molecular encapsulation of luteolin and pinocembrin with various derivatized beta-cyclodextrins. *J. Mol. Struct.* 2019, 1180, 480–490. [CrossRef]
- Wongpituk, P.; Nutho, B.; Panman, W.; Kungwan, N.; Wolschann, P.; Rungrotmongkol, T.; Nunthaboot, N. Structural dynamics and binding free energy of neral-cyclodextrins inclusion complexes: Molecular dynamics simulation. *Mol. Simul.* 2017, 43, 1356–1363. [CrossRef]
- Choi, Y.; Lee, J.; Kum, W.C.; Hwang, S.; Jeong, K.; Jung, S. Binding geometry of inclusion complex as a determinant factor for aqueous solubility of the flavonoid/β-cyclodextrin complexes based on molecular dynamics simulations. *Bull. Korean Chem. Soc.* 2005, 26, 1203–1208. [CrossRef]
- Chen, W.; Chang, C.E.; Gilson, M.K. Calculation of cyclodextrin binding affinities: Energy, entropy, and implications for drug design. *Biophys. J.* 2004, *87*, 3035–3049. [CrossRef]
- 40. Liu, L.; Guo, Q.-X. The Driving Forces in the Inclusion Complexation of Cyclodextrins. J. Incl. Phenom. Macrocycl. Chem. 2002, 42, 1–14. [CrossRef]
- 41. Fermeglia, M.; Ferrone, M.; Lodi, A.; Pricl, S. Host–guest inclusion complexes between anticancer drugs and β-cyclodextrin: Computational studies. *Carbohydr. Polym.* **2003**, *53*, 15–44. [CrossRef]
- 42. Alvira, E. Theoretical Study of the β-Cyclodextrin Inclusion Complex Formation of Eugenol in Water. *Molecules* **2018**, 23, 928. [CrossRef]
- González-Ruiz, V.; Cores, Á.; Martín-Cámara, O.; Orellana, K.; Cervera-Carrascón, V.; Michalska, P.; Olives, A.I.; León, R.; Martín, M.A.; Menéndez, J.C. Enhanced Stability and Bioactivity of Natural Anticancer Topoisomerase I Inhibitors through Cyclodextrin Complexation. *Pharmaceutics* 2021, 13, 1609. [CrossRef]
- 44. Hao, X.; Sun, X.; Zhu, H.; Xie, L.; Wang, X.; Jiang, N.; Fu, P.; Sang, M. Hydroxypropyl-β-Cyclodextrin-Complexed Resveratrol Enhanced Antitumor Activity in a Cervical Cancer Model: In Vivo Analysis. *Front. Pharmacol.* **2021**, *12*, 573909. [CrossRef]
- 45. Zhang, L.; Man, S.; Qiu, H.; Liu, Z.; Zhang, M.; Ma, L.; Gao, W. Curcumin-cyclodextrin complexes enhanced the anti-cancer effects of curcumin. *Environ. Toxicol. Pharmacol.* **2016**, *48*, 31–38. [CrossRef]
- Wang, F.; Yang, B.; Zhao, Y.; Liao, X.; Gao, C.; Jiang, R.; Han, B.; Yang, J.; Liu, M.; Zhou, R. Host-guest inclusion system of scutellarein with 2-hydroxypropyl-beta-cyclodextrin: Preparation, characterization, and anticancer activity. *J. Biomater. Sci. Polym. Ed.* 2014, 25, 594–607. [CrossRef]
- 47. Loftsson, T. Drug permeation through biomembranes: Cyclodextrins and the unstirred water layer. Pharmazie 2012, 67, 363–370.
- 48. Brewster, M.E.; Noppe, M.; Peeters, J.; Loftsson, T. Effect of the unstirred water layer on permeability enhancement by hydrophilic cyclodextrins. *Int. J. Pharm.* 2007, 342, 250–253. [CrossRef]
- 49. Higuchi, T. A phase solubility technique. Adv. Anal. Chem. Instrum. 1965, 4, 117–211.
- 50. Chierentin, L.; Garnero, C.; Chattah, A.K.; Delvadia, P.; Karnes, T.; Longhi, M.R.; Salgado, H.R. Influence of β-cyclodextrin on the Properties of Norfloxacin Form A. *AAPS PharmSciTech* **2015**, *16*, 683–691. [CrossRef]
- Lopalco, A.; Manni, A.; Keeley, A.; Haider, S.; Li, W.; Lopedota, A.; Altomare, C.D.; Denora, N.; Tuleu, C. In Vivo Investigation of (2-Hydroxypropyl)-β-cyclodextrin-Based Formulation of Spironolactone in Aqueous Solution for Paediatric Use. *Pharmaceutics* 2022, 14, 780. [CrossRef] [PubMed]
- 52. Piras, A.M.; Fabiano, A.; Chiellini, F.; Zambito, Y. Methyl-β-cyclodextrin quaternary ammonium chitosan conjugate: Nanoparticles vs. macromolecular soluble complex. *Int. J. Nanomed.* **2018**, *13*, 2531–2541. [CrossRef]
- 53. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 09*; Gaussian Inc.: Wallingford, CT, USA, 2009.
- 54. Case, D.A.; Betz, R.M.; Cerutti, D.S.; Cheatham, T., III; Darden, T.A.; Duke, R.E.; Giese, T.J.; Gohlke, H.; Goetz, A.W.; Homeyer, N.; et al. *Amber* 2016; University of California: San Francisco, CA, USA, 2016.
- 55. Kerdpol, K.; Kicuntod, J.; Wolschann, P.; Mori, S.; Rungnim, C.; Kunaseth, M.; Okumura, H.; Kungwan, N.; Rungrotmongkol, T. Cavity Closure of 2-Hydroxypropyl-β-Cyclodextrin: Replica Exchange Molecular Dynamics Simulations. *Polymers* 2019, *11*, 145. [CrossRef]
- 56. Mahalapbutr, P.; Sangkhawasi, M.; Kammarabutr, J.; Chamni, S.; Rungrotmongkol, T. Rosmarinic Acid as a Potent Influenza Neuraminidase Inhibitor: In Vitro and In Silico Study. *Curr. Top. Med. Chem.* **2020**, *20*, 2046–2055. [CrossRef] [PubMed]
- 57. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [CrossRef]
- 58. Luty, B.A.; van Gunsteren, W.F. Calculating Electrostatic Interactions Using the Particle–Particle Particle–Mesh Method with Nonperiodic Long-Range Interactions. *Am. J. Phys. Chem.* **1996**, *100*, 2581–2587. [CrossRef]
- 59. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [CrossRef]

- 60. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; González-Outeiriño, J.; Daniels, C.R.; Foley, B.L.; Woods, R.J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622–655. [CrossRef]
- 61. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. 83 Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]
- 62. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461. [CrossRef]





Article Exogenously Applied Sodium Nitroprusside Mitigates Lead Toxicity in Rice by Regulating Antioxidants and Metal Stress-Related Transcripts

Waqas Rahim^{1,†}, Murtaza Khan^{2,†}, Tiba Nazar Ibrahim Al Azzawi¹, Anjali Pande¹, Nusrat Jahan Methela¹, Sajid Ali², Muhammad Imran¹, Da-Sol Lee¹, Geun-Mo Lee¹, Bong-Gyu Mun¹, Yong-Sun Moon², In-Jung Lee¹ and Byung-Wook Yun^{1,*}

- ¹ Department of Applied Biosciences, Kyungpook National University, Daegu 41566, Korea
- ² Department of Horticulture and Life Science, Yeungnam University, Gyeongsan 38541, Korea

+ These authors contributed equally to this work.

Abstract: Sustainable agriculture is increasingly being put in danger by environmental contamination with dangerous heavy metals (HMs), especially lead (Pb). Plants have developed a sophisticated mechanism for nitric oxide (NO) production and signaling to regulate hazardous effects of abiotic factors, including HMs. In the current study, we investigated the role of exogenously applied sodium nitroprusside (SNP, a nitric oxide (NO) donor) in ameliorating the toxic effects of lead (Pb) on rice. For this purpose, plants were subjected to 1.2 mM Pb alone and in combination with 100 μ M SNP. We found that under 1.2 mM Pb stress conditions, the accumulation of oxidative stress markers, including hydrogen peroxide (H_2O_2) (37%), superoxide anion (O_2^-) (28%), malondialdehyde (MDA) (33%), and electrolyte leakage (EL) (34%), was significantly reduced via the application of 100 µM SNP. On the other hand, under the said stress of Pb, the activity of the reactive oxygen species (ROS) scavengers such as polyphenol oxidase (PPO) (60%), peroxidase (POD) (28%), catalase (CAT) (26%), superoxide dismutase (SOD) (42%), and ascorbate peroxidase (APX) (58%) was significantly increased via the application of 100 µM SNP. In addition, the application of 100 µM SNP rescued agronomic traits such as plant height (24%), number of tillers per plant (40%), and visible green pigments (44%) when the plants were exposed to 1.2 mM Pb stress. Furthermore, after exposure to 1.2 mM Pb stress, the expression of the heavy-metal stress-related genes OsPCS1 (44%), OsPCS2 (74%), OsMTP1 (83%), OsMTP5 (53%), OsMT-I-1a (31%), and OsMT-I-1b (24%) was significantly enhanced via the application of 100 µM SNP. Overall, our research evaluates that exogenously applied 100 mM SNP protects rice plants from the oxidative damage brought on by 1.2 mM Pb stress by lowering oxidative stress markers, enhancing the antioxidant system and the transcript accumulation of HMs stress-related genes.

Keywords: nitric oxide; antioxidants; metal-stress related transcripts; rice; Pb-stress

1. Introduction

Heavy metals (HMs), including lead (Pb), are released due to rapid industrialization, mining, economic growth, anthropogenic activities, and the excessive use of inorganic fertilizers, along with other agrochemicals [1,2]. Pollutants possessing heavy metals enter the soil via a wide range of routes and pose a threat to the sustainable agroecosystem, farming, and livelihood [3,4]. Pb remains stable in soil for a long period of time, does not easily dissociate, and hence accumulates in human and animal bodies through the consumption of contaminated plants [5], thus threatening their health. For example, it is estimated that a Pb concentration of more than $40\mu g dL^{-1}$ in an infant's blood can cause a blockage of hemoglobin synthesis, resulting in anemia [6]. Pb is ranked the second most toxic HM after arsenic (As) [7] and has a promising role in the impairment of plant growth and development by adversely affecting the plant's metabolism [8], seed

Citation: Rahim, W.; Khan, M.; Al Azzawi, T.N.I.; Pande, A.; Methela, N.J.; Ali, S.; Imran, M.; Lee, D.-S.; Lee, G.-M.; Mun, B.-G.; et al. Exogenously Applied Sodium Nitroprusside Mitigates Lead Toxicity in Rice by Regulating Antioxidants and Metal Stress-Related Transcripts. *Int. J. Mol. Sci.* 2022, *23*, 9729. https://doi.org/ 10.3390/ijms23179729

Academic Editor: Wajid Zaman

Received: 16 August 2022 Accepted: 24 August 2022 Published: 27 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

^{*} Correspondence: bwyun@knu.ac.kr

germination, seedling growth, cell division, the permeability of plasma membrane, and various ultrastructural modifications [9,10]. A high concentration of lead (1.2 mM) induced a significant reduction in the plant height, number of tillers, number of panicles per plant, and the number of spikelets per panicle in different rice cultivars [11]. However, different rice cultivars responded differently to the 1.2 mM Pb stress. However, some species showed the lowest drop in agronomic attributes to different stresses [12].

An elevated level of Pb stimulates ROS generation, which induces oxidative stress, damages the plasma membrane, and alters metabolism and physiological reactions [13]. To cope with oxidative stress, plants have developed a sophisticated antioxidant defense mechanism comprising the generation of SOD and PO [14–16], APX, PPO, and CAT [17,18]. Furthermore, when plants are exposed to HMs, they use their inherent complex mechanisms and strategies for metal uptake, storage, transportation, detoxification, elimination, and compartmentalization [19,20]. However, a variation exists among different species or varieties of a species in the uptake, translocation, and accumulation of Pb [21]. In addition, phytochelatins (PCs) are the prime inducers of responses to the vulnerability of various heavy metals in plants [22]. They are considered to bind metals via thiolate coordination, which is involved in HMs homeostasis and detoxification. Recently, OsPCS1 and OsPCS2 genes have been characterized in rice [23]. In addition, the cation diffusion facilitator (CDF) genes family transports either metal ions out of the cytosol into extracellular spaces or diffuses into the vacuoles [24] and are also called metal tolerance proteins (MTPs). Similarly, metallothionein (MTs) play an important role in the detoxification of heavy metals [25], maintaining the balance of intracellular metallic ions in plants [26], the scavenging of ROS [27], and the regulation of developmental processes [28]. Furthermore, NO was awarded "molecule of the year" in 1992 and is crucial in modulating various physiological and biochemical activities in plants [29]. It is highly diffusible and participates in a wide range of abiotic stress tolerance mechanisms in plants [30]. ROS metabolism also involves the participation of NO [31]. It is noteworthy that ROS/NO interaction can cause cytotoxicity, or it can be protective, depending on the relative concentrations of ROS and NO [32]. The alleviative effect of NO on abiotic stresses in plants has been documented [33].

Rice is a necessity for life, and it has significantly impacted millions of people's economics, diets, and cultures [34]. It is most often exposed to environmental hazards, including heavy metals such as lead, the main causes of which are the overuse of agrochemicals and repeated use of waste and sewage water during rice cultivation [35]. The European chemicals agency (ECHA) has assorted Pb in the group of chemicals of great perturbation for the environment [36]. Therefore, it is essential to identify the techniques that help improve the defense system of the rice against Pb stress. For example, NO applications decrease the uptake of Pb in *Arabidopsis thaliana* [37] and affect gene expression in *Zea mays* [38]. Therefore, in the current study, we investigated alleviating the toxic effects of lead on rice by applying sodium nitroprusside (SNP). The novelty of the present work is revealing the antioxidant machinery of NO in controlling Pb-induced oxidative damage in rice.

2. Results

2.1. SNP Improves Morphological Parameters of Rice under Pb Stress

Pb-stress adversely affects the growth attributes of different crops. Our results showed a significant improvement (16%) in the plant's shoot length under SNP treatment in Pbuntreated plants compared to Pb-untreated control plants, as shown in Figure 1A,B. The results for plants under Pb stress revealed a significant reduction in shoot length (29%) compared to the control (only water). However, SNP-treated plants under Pb stress significantly enhanced (24%) the overall shoot length compared to sole Pb-treated rice plants (Figure 1A,B). Furthermore, a considerable increase in the number of tillers (14%) in NO-treated plants was observed in Pb-untreated plants compared to their respective control plants (Figure 1A,C). Pb-treatment significantly reduced the number of tillers (29%) in Pb-treated plants compared to control Pb-untreated plants. However, SNP treatment in Pb-treated plants significantly improved the number of tillers (40%) compared to sole Pbtreated plants, as shown in Figure 1A,C. In addition, our results revealed that Pb-untreated plants supplied with SNP considerably enhanced (13%) green pigment content compared to the control Pb-untreated plants (Figure 1D). Pb stress showed a significant reduction (27%) in the visible green pigment compared to control Pb-untreated plants (Figure 1D). However, the treatment of SNP in plants under Pb stress revealed a significant increase (44%) in visible green pigment content compared to sole Pb-treated plants (Figure 1D).



Figure 1. Effects of exogenously applied SNP on rice with or without Pb-induced stress. (**A**) Phenotypic characteristics; (**B**) shoot length; (**C**) number of tillers per plant; (**D**) SPAD value for photosynthetic green pigment contents. Each data point indicates the mean \pm standard deviation (n = 3). Bars with different letters indicate significant differences, according to Duncan's multiple range test. The results are compared to the respective controls (Pb-untreated and Pb-treated plants).

2.2. SNP Enhances the Chlorophyll a, b, and Protein Contents

Plants under stress face the challenge of inhibition in photosynthesis and changes in chlorophyll contents. The results revealed that SNP applications considerably increased chlorophyll a and b levels (16 and 15%, respectively) in Pb-untreated plants compared to control Pb-untreated plants, as shown in Figure 2A,B. Plants under Pb stress showed a significant decrease (36 and 34%, respectively) in chlorophyll a and b contents compared to control Pb-untreated plants (Figure 2A,B). The data also showed a significant increase in chlorophyll a and b contents (38 and 36%, respectively) in SNP-treated plants under Pb stress compared to plants under sole Pb stress (Figure 2A,B). Furthermore, the obtained data showed a significant increase in protein content (15%) in SNP-supplied Pb-untreated plants compared to control Pb-untreated plants, as shown in Figure 2C. Plants under Pb-treatment showed a highly significant decrease (49%) in protein content compared to control Pb-untreated plants (Figure 2C). However, SNP applications in Pb-treated plants significantly increased the total protein content (35%) in Pb-treated plants compared to sole Pb-treated plants (Figure 2C).



Figure 2. Effect of exogenously applied SNP on rice with or without Pb-induced stress. (**A**) Chlorophyll a content; (**B**) chlorophyll b content; (**C**) protein content. Each data point indicates the mean \pm standard deviation (n = 3). Bars with different letters indicate significant differences, according to Duncan's multiple range test. The results are compared to the respective controls (Pb-untreated and Pb-treated plants).

2.3. Exogenously Applied SNP Mitigates Membrane Injury and Enhances Protection of Rice against Pb-Toxicity

In the current study, we compute the impact of SNP on ROS compounds, including H_2O_2 , O_2^- , MDA, and electrolyte leakage. Pb-treatment significantly enhanced (305%) MDA levels in plants compared to control Pb-untreated plants (Figure 3A). However, the SNP supply in Pb-treated plants significantly reduced (33%) MDA levels in Pb-treated plants compared to sole Pb-treated plants (Figure 3A). A significant increase in ROS (H_2O_2 and O_2^-) levels (311 and 81%, respectively) was observed in Pb-treated plants compared to control Pb-untreated plants (Figure 3B,C). However, SNP applications considerably mitigate H_2O_2 and O_2^- levels (37 and 28%, respectively) in plants under Pb stress compared to plants under sole Pb stress (Figure 3B,C). Furthermore, Pb treatments significantly enhanced (65%) ion leakage compared to control Pb-untreated plants (Figure 3D), which was significantly mitigated (34%) by SNP applications in Pb-treated plants, as shown in Figure 3D.



Figure 3. Effect of exogenously applied SNP on rice with or without Pb-induced stress. (**A**) MDA level; (**B**) H_2O_2 content; (**C**) superoxide anion level; (**D**) electrolyte leakage. Each data point indicates the mean \pm standard deviation (n = 3). Bars with different letters indicate significant differences, according to Duncan's multiple range test. The results are compared to the respective controls (Pb-untreated and Pb-treated plants).

2.4. SNP Regulates the Antioxidant Enzymes Machinery

Overall, the results revealed that NO treatments increased the level and activity of antioxidant enzymes under Pb stress. The activities of PPO, PO, CAT, SOD, and APX were enhanced (35, 18, 15, 17, and 44%, respectively) in SNP-supplemented Pb-untreated plants compared to control Pb-untreated plants, as shown in Figure 4A–E. Pb treatment considerably activates the production of PPO, PO, CAT, SOD, and APX (7, 24, 218, 52, and 68%, respectively) compared to control Pb-untreated plants (Figure 4A–E). However, the supply of SNP significantly enhanced the activity of the above antioxidants by 60, 28, 26, 42, and 58%, respectively, compared to sole Pb-supplied plants.



Figure 4. Effect of exogenously applied SNP on rice with or without Pb-induced stress. (**A**) Polyphenol oxidase; (**B**) peroxidase; (**C**) catalase; (**D**) superoxide dismutase; (**E**) ascorbate peroxidase. Each data point indicates the mean \pm standard deviation (n = 3). Bars with different letters indicate significant differences, according to Duncan's multiple range test. The results are compared to the respective controls (Pb-untreated and Pb-treated plants).

2.5. SNP Modulates the Expression of the Genes Related to Metal Stress

The results revealed that the relative expression of *OsPCS1* and *OsPCS2* is differentially and highly significantly enhanced (387 and 134%, respectively) in Pb-treated plants compared to the Pb-untreated control. However, SNP supplementation in Pb-treated plants shows significant improvements in the relative expression of *OsPCS1* (44%) and *OsPCS2* (74%) compared to sole Pb-supplied plants (Figure 5A,B).

Furthermore, our results showed that plants under Pb stress have significantly higher relative *OsMTP1* and *OsMTP5* expression (460 and 177%, respectively) when compared to control Pb-untreated plants. However, the SNP supply further improved the relative expression of *OsMTP1* and *OsMTP5* (83 and 53%, respectively) compared to sole Pb-treated plants (Figure 5C,D).

Pb treatments triggered a significant enhancement in relative expressions of *OsMT-I-1a* and *OsMT-I-1b* (151 and 1065%, respectively) compared to Pb-untreated control plants (Figure 5E,F). *However*, SNP supplies further improved the relative expression of *OsMT-I-1a* and *OsMT-I-1b* (31 and 24%, respectively) compared to plants solely treated with Pb.



Figure 5. Effect of exogenously applied SNP on the relative expression of rice phytochelatin, metal transporter, and metallothionein protein candidate genes with or without Pb-induced stress. (A) *OsPC1;* (B) *OsPC2;* (C) *OsMTP1;* (D) *OsMTP5;* (E) *OsM1-1a;* (F) *OsM1-1b.* Each data point indicates the mean \pm standard deviation (n = 3). Bars with different letters indicate significant differences, according to Duncan's multiple range test. The results are compared to the respective controls (Pb-untreated and Pb-treated plants).

3. Discussion

Nitric oxide acts as a "guardee" molecule to alleviate heavy-metal toxicities via stress perception, signaling, and the acclimatization of plants under heavy-metal stress [39]. In the current study, the observed inhibitory effects of Pb on key morphological characteristics of rice, i.e., plant height, the number of tillers, and visible green pigment content (Figure 1A–D), were significantly reduced by an exogenous application of SNP in the form of 100 μ M SNP solutions. Similarly, the toxic effects of Pb on wheat seedling growth were significantly reduced via the application of 100 μ M SNP [40].

Many findings reported changes in the physiological, biochemical, and molecular levels of plants, such as the accumulation and activity of chlorophyll a and b contents in plants under abiotic stress [41–46]. SNP significantly ameliorated Pb-induced negative impacts on chlorophyll a and b contents, as shown in Figure 2A,B. Similarly, under heavy-metal (Pb and Cd) stress, adding SNP increases chlorophyll and carotenoid concentrations in bamboo plants [47].

Plants activate a plethora of adaptive responses to withstand abiotic stresses, including a high accumulation of proteins [48]. The current study showed that the significant reduction in protein content due to Pb-induced damages is significantly mitigated by exogenous application of SNP, as shown in Figure 2C. This result is in accordance with the previous reported result [47].

Heavy metals cause damage at the cellular and molecular levels to plants, both directly and indirectly, through overproduction and the hyperaccumulation of ROS [49,50]. However, NO protects plants against oxidative damage by scavenging ROS [49]. In the present study, due to Pb stress, a significant enhancement in the production of ROS and electrolyte leakage was observed. However, it was significantly mitigated by exogenously applied SNPs (Figure 3A–D). These results are in accordance with the previous literature published [51,52]. Exogenous applications of NO in rice and perennial ryegrass under Cd

and Pb-induced stress, respectively, decreased the production of ROS and MDA, resulting in increased activities of antioxidant enzymes [53–55].

NO regulates the antioxidant enzyme machinery at the cellular level, affecting the cellular redox level [49,56]. Our results indicated that the application of NO exogenously in the form of SNP mitigates the adverse effects of Pb on rice plants by increasing the production of antioxidant enzymes such as PPO, POD, CAT, SOD (Figure 4A–D), and APX (Figure 4E), which is according to previous literature published for various heavy metals [57].

The increased tolerance to heavy metals is linked with phytochelatin synthesis, for example, cadmium stress [58,59]. Several published reports suggest that PC-deficient mutants increased heavy-metal sensitivity [60]. Therefore, the current study was designed to check the effect of the exogenous application of SNP on the expression of two candidate phytochelatins, i.e., *OsPCS1* and *OsPCS2*. Our results showed that the relative expression of both genes was significantly enhanced under lead stress compared to control plants (Figure 5A,B), suggesting the transcript's accumulation for metal binding and vacuolar compartmentalization. To elucidate if NO affects the expression of phytochelatin genes, the observed effects and data showed that the expression levels of transcripts were considerably enhanced in SNP-supplied Pb-treated plants compared to plants under sole Pb stress.

Plants under heavy-metal stress undergo a transcriptional regulation of CDF protein family members *OsMTP1* and *OsMTP5*. They play an important role in cation homeostasis, chelation, sequestration, or the expulsion of excess heavy metals [61]. The expression pattern of *OsMTP1* is enhanced during Cd-stress exposure; overexpression and gene silencing confirmed its role in the transportation of Cd [62]. The expression analyses of *OsMTP1 and OsMTP5* in the current study for Pb stress are tallied with the results in the cited literature. The current study revealed improved expressions of *OsMTP1 and OsMTP5* in plants treated with Pb compared to control Pb-untreated plants, as shown in Figure 5C,D. To elucidate if SNP affects the relative expression of *OsMTP1 and OsMTP5*, the results showed that the expression level was significantly enhanced in NO-treated rice under Pb stress compared to rice only under Pb stress (Figure 5C,D).

The overexpression of metallothionein genes in tobacco showed decreased and increased Arsenic accumulation in roots and shoots, respectively [63]. In the current study, we found that the relative expression of two candidate genes, i.e., *OsMT-I-1a* and *OsMT-I-1b*, is upregulated by Pb toxicity, as shown in Figure 5E,F, which is further improved by the application of SNP in Pb-treated plants. Although the mechanism of heavy-metal detoxification in plants by metallothionein is still elusive, our results are supported by the previous published literature [64].

4. Materials and Methods

4.1. Plant Material, Husbandry Preparation, and Growth Conditions

The experiment was performed in soil under greenhouse conditions at Kyungpook National University, Daegu, Republic of Korea. For the experiment, the Jinbu rice cultivar (*Oryza sativa* L. ssp. *Japonica*) was selected as genetic material. Seeds sterilization, germination, and sowing were conducted as previously described [65]. Lead (II) nitrate (Pb(NO₃)₂) measuring 1.2 mM was applied as described earlier [11]. The pots were divided into four treatments with three replicates each, as listed in Table 1. After three weeks of transplantation, the plants were supplied with 100 μ M SNP [40,54].

Treatments	Concentrations
Control	untreated
SNP-R	100 μM SNP
Pb	1.2 mM Pb (NO ₃) ₂
Pb + SNP-R	1.2 mM Pb (NO ₃) ₂ + 100 μM SNP

Table 1. Demonstrates different treatments and concentrations of the chemicals used in this study.

Pb(NO₃)₂: Lead (II) nitrate as a Pb source; SNP-R: sodium nitroprusside as a NO-donor applied through roots via sub-irrigation.

4.2. Measurement of Electrolyte Leakage and Visible Green Pigment Quantification

The electrolyte leakage assay was performed to estimate any ion leakage that would have resulted from Pb oxidative damage, as previously described [48]. Electrolyte leakage-1 (EL1) was measured using a portable conductivity meter (HURIBA Twin Cond B-173, Fukuoka, Japan). For electrolyte leakage-2 (EL2), the samples were autoclaved and cooled at room temperature. Electrolyte leakage (EL), expressed in percentage (%), was calculated using the following formula.

$EL\% = EL1/EL2 \times 100$

Visible green pigment content in leaves was measured in leaf samples using a SPAD meter (SPAD-502; Minolta Co. Ltd., Osaka, Japan), as previously described [11].

4.3. Quantification of Chlorophyll a and b Contents

To quantify and calculate chlorophyll a and b contents, a detailed method was followed as described previously [66,67].

4.4. Quantification of Lipid Peroxidation

The determination of lipid peroxidation was performed by calculating the amount of a byproduct of membrane bilayer oxidation, malondialdehyde (MDA), by using a published method [68].

4.5. Hydrogen Peroxide (H_2O_2) and Superoxide Anion (O_2^-) Content

 H_2O_2 content in rice leaf tissue was quantified and calculated using the method described [69] and expressed in units as μ mol g⁻¹ FW. Similarly, O_2^- contents in rice leaf tissue were quantified and calculated as previously described [70] and expressed in units as μ mol g⁻¹ FW.

4.6. Estimation of Antioxidant Activities

As previously mentioned, CAT, POD, PPO, and SOD activity were examined [48]. In brief, 400 mg of leaf samples was powdered using a chilled mortar and pestle. The crushed samples were homogenized with 0.1 M phosphate buffer (pH 6.8) and centrifuged at 4 $^{\circ}$ C for 15 min at 5000 rpm. The supernatant was used as the crude enzyme source for CAT, POD, and PPO activities.

The activity of PPO and POD was estimated [48], and the activity of CAT was measured [11] as previously described. Furthermore, SOD activities were analyzed by following the photoreduction of nitro blue tetrazolium (NBT) [66], and the APX's activity was determined [71], as mentioned earlier.

4.7. RNA Extraction and Quantitative Real-Time PCR

RNA was extracted using the standardized procedure followed by [67]. Additionally, complementary DNA (cDNA) and RT-PCR were carried out according to the previous literature [72]. The synthesized cDNA was used as a template for further assessments of transcript accumulation using qRT-PCR (EcoTM IlluminaTM San Diego, California, USA), as previously described [29]. The list of genes and corresponding primers with names and sequences is provided in Table 2.

Primer	Forward Sequence (5'-3')	Reverse Sequence (5'-3')
OsActin	GGA ACT GGT ATG GTC AAG GC	AGT CTC ATG GAT AAC CGC AG
OsPCS1	CGA AGA TTC CAT TTC CCA GA	TCG AGG ATA TCG GTG AAA GC
OsPCS2	TCC CTC TCC GTC GTC CTC	CCT CCG CCT TCA CCT TGT
OsMTP1	TCA AGA TGC TGC GCA ACA TCC	GAG CTC CTA CTC GCG CTC AAT G
OsMTP5	ACG CTC GTT GTC TGA TGG G	GTC ACT GCA AGC ATG ATG TCC AC
OsMT-I-1a	TGC GGA AG TAC CCT GA	TTC TCC GGC GCC ACA C
OsMT-I-1b	CTG TGG ATC AAG CTG TGG CT	GCT GCT GCT CTT CTC TTC CA

Table 2. List of primers used in this study.

4.8. Statistical Analysis

Independent experimental analyses were performed in triplicates using a completely randomized design (CRD). To assess the statistical significance between Pb-treated plants with control and SNP-supplied Pb-plants, Statistical Analysis Software SAS version 9.1 (SAS Institute Inc., Cary, NC, USA) was used, and all data were statistically evaluated with Duncan's multiple range test. The significance threshold was set at p < 0.05. GraphPad Prism software version 6.0 (San Diego, CA, USA) was used to present the results graphically.

5. Conclusions

By altering the activities of SOD, POD, and PPO, as well as the amount of CAT and gene expression, Pb stress induced the overproduction of ROS and disrupted the H_2O_2 and MDA scavenging system. Conclusively, the significant role of exogenous NO donors (SNP) in plants has a protective effect in alleviating lead stress in rice. Our results revealed that exogenously applied SNP improves the Pb stress tolerance in rice by activating the antioxidant system, lowering electrolyte leakage, reducing the production of H_2O_2 and MDA, and increasing the expression of heavy-metal stress-related genes. Based on the present results, it is suggested that the application of SNP will enhance the growth and productivity of rice under lead stress conditions.

Author Contributions: Conceptualization, methodology, and validation, B.-W.Y., I.-J.L. and M.K.; formal analysis, investigation, and data curation, W.R., M.K., T.N.I.A.A., M.I., N.J.M., D.-S.L. and G.-M.L.; writing—original draft preparation, W.R. and M.K.; writing—review and editing M.K., W.R., B.-W.Y., B.-G.M., A.P., Y.-S.M. and S.A.; equally contributed to this work and have the rights to claim first author titles in their CVs, W.R. and M.K.; funding acquisition, B.-W.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (Grant number 2020R111A3073247), Republic of Korea, and a project to train professional personnel in biological materials by the Ministry of Environment.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Hu, J.; Shi, G.; Xu, Q.; Wang, X.; Yuan, Q.; Du, K. Effects of Pb²⁺ on the active oxygen-scavenging enzyme activities and ultrastructure in *Potamogeton crispus* leaves. *Russ. J. Plant Physiol.* **2007**, *54*, 414–419. [CrossRef]
- Liaquat, F.; Munis, M.F.H.; Arif, S.; Haroon, U.; Shi, J.; Saqib, S.; Zaman, W.; Che, S.; Liu, Q. PacBio single-molecule long-read sequencing reveals genes tolerating manganese stress in *Schima superba* saplings. *Front. Genet.* 2021, 12, 635043. [CrossRef] [PubMed]
- 3. Anjum, S.A.; Ashraf, U.; Khan, I.; Tanveer, M.; Ali, M.; Hussain, I.; Wang, L.C. Chromium and aluminum phytotoxicity in maize: Morpho-physiological responses and metal uptake. *CLEAN Soil Air Water* **2016**, *44*, 1075–1084. [CrossRef]

- 4. Wang, X.; Zhang, C.; Qiu, B.; Ashraf, U.; Azad, R.; Wu, J.; Ali, S. Biotransfer of Cd along a soil-plant-mealybug-ladybird food chain: A comparison with host plants. *Chemosphere* **2017**, *168*, 699–706. [CrossRef]
- 5. Jiao, W.; Chen, W.; Chang, A.C.; Page, A.L. Environmental risks of trace elements associated with long-term phosphate fertilizers applications: A review. *Environ. Pollut.* **2012**, *168*, 44–53. [CrossRef]
- 6. Ogwuegbu, M.; Muhanga, W. Investigation of lead concentration in the blood of people in the copper belt province of Zambia. *J. Environ.* **2005**, *1*, 66–75.
- 7. Gaya, U.; Ikechukwu, S. Heavy metal contamination of selected spices obtained from Nigeria. J. Appl. Sci. Environ. Manag. 2016, 20, 681–688. [CrossRef]
- 8. Uzu, G.; Sobanska, S.; Aliouane, Y.; Pradere, P.; Dumat, C. Study of lead phytoavailability for atmospheric industrial micronic and sub-micronic particles in relation with lead speciation. *Environ. Pollut.* **2009**, *157*, 1178–1185. [CrossRef]
- Gupta, D.; Nicoloso, F.; Schetinger, M.; Rossato, L.; Pereira, L.; Castro, G.; Srivastava, S.; Tripathi, R. Antioxidant defense mechanism in hydroponically grown *Zea mays* seedlings under moderate lead stress. *J. Hazard. Mater.* 2009, 172, 479–484. [CrossRef]
- 10. Maestri, E.; Marmiroli, M.; Visioli, G.; Marmiroli, N. Metal tolerance and hyperaccumulation: Costs and trade-offs between traits and environment. *Environ. Exp. Bot.* **2010**, *68*, 1–13. [CrossRef]
- 11. Khan, M.; Al Azzawi, T.N.I.; Imran, M.; Hussain, A.; Mun, B.-G.; Pande, A.; Yun, B.-W. Effects of lead (Pb)-induced oxidative stress on morphological and physio-biochemical properties of rice. *Biocell* **2021**, *45*, 1413. [CrossRef]
- Faria, J.M.; Teixeira, D.M.; Pinto, A.P.; Brito, I.; Barrulas, P.; Carvalho, M. The protective biochemical properties of arbuscular mycorrhiza extraradical mycelium in acidic soils are maintained throughout the mediterranean summer conditions. *Agronomy* 2021, 11, 748. [CrossRef]
- Sunera, A.; Saqib, S.; Uddin, S.; Zaman, W.; Ullah, F.; Ayaz, A.; Asghar, M.; Rehman, S.; Munis, M. Characterization and phytostimulatory activity of bacteria isolated from tomato (*Lycopersicon esculentum* Mill.) rhizosphere. *Microb. Pathog.* 2020, 140, 103966. [CrossRef] [PubMed]
- 14. Rezaei-Chiyaneh, E.; Amirnia, R.; Machiani, M.A.; Javanmard, A.; Maggi, F.; Morshedloo, M.R. Intercropping fennel (*Foeniculum vulgare* L.) with common bean (*Phaseolus vulgaris* L.) as affected by PGPR inoculation: A strategy for improving yield, essential oil and fatty acid composition. *Sci. Hortic.* **2020**, *261*, 108951. [CrossRef]
- 15. Asghar, M.; Younas, M.; Arshad, B.; Zaman, W.; Ayaz, A.; Rasheed, S.; Shah, A.H.; Ullah, F.; Saqib, S. Bioactive potential of cultivated *Mentha arvensis* L. For preservation and production of health-oriented food. *J. Anim. Plant Sci.* 2022, *32*, 835–844.
- Azam, Z.; Ayaz, A.; Younas, M.; Qureshi, Z.; Arshad, B.; Zaman, W.; Ullah, F.; Nasar, M.Q.; Bahadur, S.; Irfan, M.M. Microbial synthesized cadmium oxide nanoparticles induce oxidative stress and protein leakage in bacterial cells. *Microb. Pathog.* 2020, 144, 104188. [CrossRef]
- 17. Cui, G.; Zhao, X.; Liu, S.; Sun, F.; Zhang, C.; Xi, Y. Beneficial effects of melatonin in overcoming drought stress in wheat seedlings. *Plant Physiol. Biochem.* **2017**, *118*, 138–149. [CrossRef]
- 18. Wang, L.; Liu, J.; Wang, W.; Sun, Y. Exogenous melatonin improves growth and photosynthetic capacity of cucumber under salinity-induced stress. *Photosynthetica* **2016**, *54*, 19–27. [CrossRef]
- Jiang, W.; Liu, D. Pb-induced cellular defense system in the root meristematic cells of *Allium sativum* L. *BMC Plant Biol.* 2010, 10, 40. [CrossRef]
- 20. Liaquat, F.; Munis, M.F.H.; Haroon, U.; Arif, S.; Saqib, S.; Zaman, W.; Khan, A.R.; Shi, J.; Che, S.; Liu, Q. Evaluation of metal tolerance of fungal strains isolated from contaminated mining soil of Nanjing, China. *Biology* **2020**, *9*, 469. [CrossRef]
- Cheng, W.-d.; Zhang, G.-p.; Yao, H.-g.; Wu, W.; Xu, M. Genotypic and environmental variation in cadmium, chromium, arsenic, nickel, and lead concentrations in rice grains. J. Zhejiang Univ. Sci. B 2006, 7, 565–571. [CrossRef] [PubMed]
- 22. Zenk, M.H. Heavy metal detoxification in higher plants-a review. *Gene* 1996, 179, 21–30. [CrossRef]
- 23. Yamazaki, S.; Ueda, Y.; Mukai, A.; Ochiai, K.; Matoh, T. Rice phytochelatin synthases OsPCS1 and OsPCS2 make different contributions to cadmium and arsenic tolerance. *Plant Direct* **2018**, *2*, e00034. [CrossRef] [PubMed]
- 24. Gustin, J.L.; Zanis, M.J.; Salt, D.E. Structure and evolution of the plant cation diffusion facilitator family of ion transporters. *BMC Evol. Biol.* **2011**, *11*, *76*. [CrossRef] [PubMed]
- Gautam, N.; Verma, P.K.; Verma, S.; Tripathi, R.D.; Trivedi, P.K.; Adhikari, B.; Chakrabarty, D. Genome-wide identification of rice class I metallothionein gene: Tissue expression patterns and induction in response to heavy metal stress. *Funct. Integr. Genom.* 2012, 12, 635–647. [CrossRef] [PubMed]
- Eun, H.-D.; Ali, S.; Jung, H.; Kim, K.; Kim, W.-C. Profiling of ACC synthase gene (ACS11) expression in *Arabidopsis* induced by abiotic stresses. *Appl. Biol. Chem.* 2019, 62, 42. [CrossRef]
- 27. Irvine, G.W.; Stillman, M.J. Cadmium binding mechanisms of isolated domains of human MT isoform 1a: Non-cooperative terminal sites and cooperative cluster sites. *J. Inorg. Biochem.* **2016**, *158*, 115–121. [CrossRef]
- 28. Mudalkar, S.; Golla, R.; Sengupta, D.; Ghatty, S.; Reddy, A.R. Molecular cloning and characterisation of metallothionein type 2a gene from *Jatropha curcas* L., a promising biofuel plant. *Mol. Biol. Rep.* **2014**, *41*, 113–124. [CrossRef]
- 29. Khan, M.; Al Azawi, T.N.I.; Pande, A.; Mun, B.-G.; Lee, D.-S.; Hussain, A.; Lee, B.-H.; Yun, B.-W. The Role of Nitric Oxide-Induced ATILL6 in Growth and Disease Resistance in *Arabidopsis thaliana*. *Front. Plant Sci.* **2021**, *12*, 685156. [CrossRef]
- 30. Khan, M.; Imran, Q.M.; Shahid, M.; Mun, B.-G.; Lee, S.-U.; Khan, M.A.; Hussain, A.; Lee, I.-J.; Yun, B.-W. Nitric oxide-induced AtAO3 differentially regulates plant defense and drought tolerance in *Arabidopsis thaliana*. *BMC Plant Biol.* **2019**, *19*, 602. [CrossRef]

- 31. Delledonne, M.; Zeier, J.; Marocco, A.; Lamb, C. Signal interactions between nitric oxide and reactive oxygen intermediates in the plant hypersensitive disease resistance response. *Proc. Natl. Acad. Sci.* **2001**, *98*, 13454–13459. [CrossRef] [PubMed]
- 32. Beligni, M.V.; Lamattina, L. Nitric oxide counteracts cytotoxic processes mediated by reactive oxygen species in plant tissues. *Planta* **1999**, *208*, 337–344. [CrossRef]
- Graziano, M.; Lamattina, L. Nitric oxide accumulation is required for molecular and physiological responses to iron deficiency in tomato roots. *Plant J.* 2007, 52, 949–960. [CrossRef] [PubMed]
- 34. Gnanamanickam, S.S. Rice and its importance to human life. In *Biological Control of rice Diseases*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–11.
- Fangmin, C.; Ningchun, Z.; Haiming, X.; Yi, L.; Wenfang, Z.; Zhiwei, Z.; Mingxue, C. Cadmium and lead contamination in japonica rice grains and its variation among the different locations in southeast China. *Sci. Total Environ.* 2006, 359, 156–166. [CrossRef] [PubMed]
- 36. Pourrut, B.; Shahid, M.; Dumat, C.; Winterton, P.; Pinelli, E. Lead Uptake, Toxicity, and Detoxification in Plants. In *Reviews of Environmental Contamination and Toxicology Volume 213*; Whitacre, D.M., Ed.; Springer: New York, NY, USA, 2011; pp. 113–136.
- Phang, C.; Leung, D.W.M.; Taylor, H.H.; Burritt, D.J. The protective effect of sodium nitroprusside (SNP) treatment on *Arabidopsis* thaliana seedlings exposed to toxic level of Pb is not linked to avoidance of Pb uptake. *Ecotoxicol. Environ. Saf.* 2011, 74, 1310–1315. [CrossRef] [PubMed]
- 38. Hermes, V.S.; Dall'Asta, P.; Amaral, F.P.; Anacleto, K.B.; Arisi, A.C.M. The regulation of transcription of genes related to oxidative stress and glutathione synthesis in *Zea mays* leaves by nitric oxide. *Biol. Plant.* **2013**, *57*, 620–626. [CrossRef]
- Terrón-Camero, L.C.; Peláez-Vico, M.Á.; Del-Val, C.; Sandalio, L.M.; Romero-Puertas, M.C. Role of nitric oxide in plant responses to heavy metal stress: Exogenous application versus endogenous production. J. Exp. Bot. 2019, 70, 4477–4488. [CrossRef] [PubMed]
- Kaur, G.; Singh, H.P.; Batish, D.R.; Mahajan, P.; Kohli, R.K.; Rishi, V.J.P.O. Exogenous nitric oxide (NO) interferes with lead (Pb)-induced toxicity by detoxifying reactive oxygen species in hydroponically grown wheat (*Triticum aestivum*) roots. *PLoS ONE* 2015, 10, e0138713. [CrossRef]
- 41. Chen, C.T.; Chen, L.-M.; Lin, C.C.; Kao, C.H. Regulation of proline accumulation in detached rice leaves exposed to excess copper. *Plant Sci.* 2001, *160*, 283–290. [CrossRef]
- Ayaz, A.; Huang, H.; Zheng, M.; Zaman, W.; Li, D.; Saqib, S.; Zhao, H.; Lü, S. Molecular cloning and functional analysis of GmLACS2-3 reveals its involvement in cutin and suberin biosynthesis along with abiotic stress tolerance. *Int. J. Mol. Sci.* 2021, 22, 9175. [CrossRef]
- 43. Naeem, M.; Shahzad, K.; Saqib, S.; Shahzad, A.; Younas, M.; Afridi, M.I. The *Solanum melongena* COP1LIKE manipulates fruit ripening and flowering time in tomato (*Solanum lycopersicum*). *Plant Growth Regul.* **2022**, *96*, 369–382. [CrossRef]
- Ali, S.; Khan, M.A.; Kim, W.-C.J.A.B.C. Pseudomonas veronii KJ mitigates flood stress-associated damage in Sesamum indicum L. Appl. Biol. Chem. 2018, 61, 575–585. [CrossRef]
- 45. Jung, H.; Ali, S.; Kim, J.Y.; Kim, W.-C. Transgenic Arabidopsis expressing acdS gene of *Pseudomonas veronii*-KJ alleviate the adverse effects of salt and water-logging stress. *Plant Breed. Biotechnol.* **2018**, *6*, 221–232. [CrossRef]
- 46. Huang, H.; Ayaz, A.; Zheng, M.; Yang, X.; Zaman, W.; Zhao, H.; Lü, S. *Arabidopsis KCS5* and *KCS6* Play Redundant Roles in Wax Synthesis. *Int. J. Mol. Sci.* 2022, 23, 4450. [CrossRef] [PubMed]
- 47. Emamverdian, A.; Ding, Y.; Barker, J.; Mokhberdoran, F.; Ramakrishnan, M.; Liu, G.; Li, Y. Nitric Oxide Ameliorates Plant Metal Toxicity by Increasing Antioxidant Capacity and Reducing Pb and Cd Translocation. *Antioxidants* **2021**, *10*, 1981. [CrossRef]
- 48. Khan, M.; Rolly, N.K.; Al Azzawi, T.N.I.; Imran, M.; Mun, B.-G.; Lee, I.-J.; Yun, B.-W. Lead (Pb)-Induced Oxidative Stress Alters the Morphological and Physio-Biochemical Properties of Rice (*Oryzasativa* L.). *Agronomy* **2021**, *11*, 409. [CrossRef]
- 49. Laspina, N.; Groppa, M.; Tomaro, M.; Benavides, M. Nitric oxide protects sunflower leaves against Cd-induced oxidative stress. *Plant Sci.* 2005, *169*, 323–330. [CrossRef]
- 50. Qian, H.; Chen, W.; Li, J.; Wang, J.; Zhou, Z.; Liu, W.; Fu, Z. The effect of exogenous nitric oxide on alleviating herbicide damage in *Chlorella vulgaris*. *Aquat. Toxicol.* **2009**, *92*, 250–257. [CrossRef]
- 51. Mostofa, M.G.; Seraj, Z.I.; Fujita, M. Exogenous sodium nitroprusside and glutathione alleviate copper toxicity by reducing copper uptake and oxidative damage in rice (*Oryza sativa* L.) seedlings. *Protoplasma* **2014**, *251*, 1373–1386. [CrossRef]
- 52. Mostofa, M.G.; Fujita, M. Salicylic acid alleviates copper toxicity in rice (*Oryza sativa* L.) seedlings by up-regulating antioxidative and glyoxalase systems. *Ecotoxicology* **2013**, *22*, 959–973. [CrossRef]
- 53. Delledonne, M. NO news is good news for plants. Curr. Opin. Plant Biol. 2005, 8, 390–396. [CrossRef] [PubMed]
- 54. Bai, X.; Dong, Y.; Wang, Q.; Xu, L.; Kong, J.; Liu, S. Effects of lead and nitric oxide on photosynthesis, antioxidative ability, and mineral element content of perennial ryegrass. *Biol. Plant.* **2015**, *59*, 163–170. [CrossRef]
- 55. Asghar, M.; Habib, S.; Zaman, W.; Hussain, S.; Ali, H.; Saqib, S. Synthesis and characterization of microbial mediated cadmium oxide nanoparticles. *Microsc. Res. Tech.* 2020, *83*, 1574–1584. [CrossRef] [PubMed]
- 56. Singh, H.P.; Batish, D.R.; Kaur, G.; Arora, K.; Kohli, R.K. Nitric oxide (as sodium nitroprusside) supplementation ameliorates Cd toxicity in hydroponically grown wheat roots. *Environ. Exp. Bot.* **2008**, *63*, 158–167. [CrossRef]
- 57. Kühnlenz, T.; Schmidt, H.; Uraguchi, S.; Clemens, S. *Arabidopsis thaliana* phytochelatin synthase 2 is constitutively active in vivo and can rescue the growth defect of the PCS1-deficient cad1-3 mutant on Cd-contaminated soil. *J. Exp. Bot.* **2014**, *65*, 4241–4253. [CrossRef]

- 58. Liang Zhu, Y.; Pilon-Smits, E.A.; Jouanin, L.; Terry, N. Overexpression of glutathione synthetase in Indian mustard enhances cadmium accumulation and tolerance. *Plant Physiol.* **1999**, *119*, 73–80. [CrossRef]
- Zhu, Y.L.; Pilon-Smits, E.A.; Tarun, A.S.; Weber, S.U.; Jouanin, L.; Terry, N. Cadmium tolerance and accumulation in Indian mustard is enhanced by overexpressing *γ*-glutamylcysteine synthetase. *Plant Physiol.* **1999**, *121*, 1169–1177. [CrossRef]
- 60. Andresen, E.; Mattusch, J.; Wellenreuther, G.; Thomas, G.; Arroyo Abad, U.; Küpper, H. Different strategies of cadmium detoxification in the submerged macrophyte *Ceratophyllum demersum* L. *Metallomics* **2013**, *5*, 1377–1386. [CrossRef]
- 61. Jan, R.; Khan, M.A.; Asaf, S.; Lee, I.-J.; Kim, K.M. Metal resistant endophytic bacteria reduces cadmium, nickel toxicity, and enhances expression of metal stress related genes with improved growth of *Oryza sativa*, via regulating its antioxidant machinery and endogenous hormones. *Plants* **2019**, *8*, 363. [CrossRef]
- Yuan, L.; Yang, S.; Liu, B.; Zhang, M.; Wu, K. Molecular characterization of a rice metal tolerance protein, OsMTP1. *Plant Cell Rep.* 2012, *31*, 67–79. [CrossRef]
- 63. Grispen, V.M.; Irtelli, B.; Hakvoort, H.W.; Vooijs, R.; Bliek, T.; Wilma, M.; Verkleij, J.A.; Schat, H. Expression of the *Arabidopsis* metallothionein 2b enhances arsenite sensitivity and root to shoot translocation in tobacco. *Environ. Exp. Bot.* **2009**, *66*, 69–73. [CrossRef]
- 64. Durner, J.; Klessig, D.F. Nitric oxide as a signal in plants. Curr. Opin. Plant Biol. 1999, 2, 369–374. [CrossRef] [PubMed]
- 65. Al Azzawi, T.N.I.; Khan, M.; Hussain, A.; Shahid, M.; Imran, Q.M.; Mun, B.-G.; Lee, S.-U.; Yun, B.-W.J.A. Evaluation of Iraqi rice cultivars for their tolerance to drought stress. *Agronomy* **2020**, *10*, 1782. [CrossRef]
- 66. Sirhindi, G.; Mir, M.A.; Abd-Allah, E.F.; Ahmad, P.; Gucel, S. Jasmonic acid modulates the physio-biochemical attributes, antioxidant enzyme activity, and gene expression in *Glycine max* under nickel toxicity. *Front Plant Sci* **2016**, *7*, 591. [CrossRef] [PubMed]
- 67. Imran, M.; Shazad, R.; Bilal, S.; Imran, Q.M.; Khan, M.; Kang, S.-M.; Khan, A.L.; Yun, B.-W.; Lee, I.-J. Exogenous Melatonin mediates the regulation of endogenous nitric oxide in *Glycine max* L. to reduce effects of drought stress. *Environ. Exp. Bot.* 2021, *188*, 104511. [CrossRef]
- 68. Khan, M.A.; Asaf, S.; Khan, A.L.; Jan, R.; Kang, S.-M.; Kim, K.-M.; Lee, I.-J. Extending thermotolerance to tomato seedlings by inoculation with SA1 isolate of Bacillus cereus and comparison with exogenous humic acid application. *PLoS ONE* **2020**, *15*, e0232228. [CrossRef]
- 69. Velikova, V.; Yordanov, I.; Edreva, A. Oxidative stress and some antioxidant systems in acid rain-treated bean plants: Protective role of exogenous polyamines. *Plant Sci.* 2000, *151*, 59–66. [CrossRef]
- 70. Misra, H.P.; Fridovich, I. The univalent reduction of oxygen by reduced flavins and quinones. J. Biol. Chem. **1972**, 247, 188–192. [CrossRef]
- Nakano, Y.; Asada, K. Hydrogen peroxide is scavenged by ascorbate-specific peroxidase in spinach chloroplasts. *Plant Cell Physiol.* 1981, 22, 867–880.
- Nabi, R.B.S.; Rolly, N.K.; Tayade, R.; Khan, M.; Shahid, M.; Yun, B.-W. Enhanced Resistance of atbzip62 against *Pseudomonas syringae* pv. tomato Suggests Negative Regulation of Plant Basal Defense and Systemic Acquired Resistance by AtbZIP62 Transcription Factor. *Int. J. Mol. Sci.* 2021, 22, 11541. [CrossRef]





Vaccine for Diabetes—Where Do We Stand?

Dinesh Kumar Chellappan ^{1,*}^(D), Richie R. Bhandare ^{2,3,*}, Afzal B. Shaik ⁴^(D), Krishna Prasad ⁵, Nurfatihah Azlyna Ahmad Suhaimi ⁶, Wei Sheng Yap ⁶, Arpita Das ⁷^(D), Pradipta Banerjee ⁸^(D), Nandini Ghosh ⁸, Tanner Guith ⁸, Amitava Das ⁸, Sarannya Balakrishnan ⁹, Mayuren Candasamy ¹^(D), Jayashree Mayuren ¹⁰, Kishneth Palaniveloo ¹¹^(D), Gaurav Gupta ^{12,13,14}, Sachin Kumar Singh ^{15,16}^(D) and Kamal Dua ^{16,17}^(D)

- ¹ Department of Life Sciences, School of Pharmacy, International Medical University, Kuala Lumpur 57000, Malaysia
- ² Department of Pharmaceutical Sciences, College of Pharmacy & Health Sciences, Ajman University, Al-Jruf, Ajman P.O. Box 346, United Arab Emirates
- ³ Center of Medical and Bio-Allied Health Sciences Research, Ajman University, Al-Jruf, Ajman P.O. Box 346, United Arab Emirates
- ⁴ St. Mary's College of Pharmacy, St. Mary's Group of Institutions Guntur, Chebrolu, Guntur 522212, India
- ⁵ Department of Clinical Sciences, College of Dentistry, Centre of Medical and Bio-Allied Health Science Research, Ajman University, Al-Jruf, Ajman P.O. Box 346, United Arab Emirates
- ⁶ School of Health Sciences, International Medical University, Kuala Lumpur 57000, Malaysia
- ⁷ Department of Biotechnology, Adamas University, Kolkata 700126, India
- ⁸ Department of Surgery, Indiana University School of Medicine, Indianapolis, IN 46202, USA
- ⁹ School of Pharmacy, International Medical University, Kuala Lumpur 57000, Malaysia
- ¹⁰ Department of Pharmaceutical Technology, School of Pharmacy, International Medical University, Kuala Lumpur 57000, Malaysia
- ¹¹ C302, Institute of Ocean and Earth Sciences, University of Malaya, Kuala Lumpur 50603, Malaysia
- ¹² School of Pharmacy, Suresh Gyan Vihar University, Jaipur 302017, India
- ¹³ Department of Pharmacology, Saveetha Dental College, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai 600077, India
- ¹⁴ Uttaranchal Institute of Pharmaceutical Sciences, Uttaranchal University, Dehradun 248007, India
- ¹⁵ School of Pharmaceutical Sciences, Lovely Professional University, Jalandhar-Delhi G.T Road, Phagwara 144411, India
 ¹⁶ Australian Research Contra in Complementary and Integrative Medicine, Faculty of Health
 - ² Australian Research Centre in Complementary and Integrative Medicine, Faculty of Health, University of Technology Sydney, Sydney, NSW 2007, Australia
- ¹⁷ Discipline of Pharmacy, Graduate School of Health, University of Technology Sydney, Sydney, NSW 2007, Australia
- * Correspondence: dinesh_kumar@imu.edu.my (D.K.C.); r.bhandareh@ajman.ac.ae (R.R.B.); Tel.: +60-12-636-1308 (D.K.C.); +971-6-705-6227 (R.R.B.)

Abstract: Diabetes is an endocrinological disorder with a rapidly increasing number of patients globally. Over the last few years, the alarming status of diabetes has become a pivotal factor pertaining to morbidity and mortality among the youth as well as middle-aged people. Current developments in our understanding related to autoimmune responses leading to diabetes have developed a cause for concern in the prospective usage of immunomodulatory agents to prevent diabetes. The mechanism of action of vaccines varies greatly, such as removing autoreactive T cells and inhibiting the interactions between immune cells. Currently, most developed diabetes vaccines have been tested in animal models, while only a few human trials have been completed with positive outcomes. In this review, we investigate the undergoing clinical trial studies for the development of a prototype diabetes vaccine.

Keywords: diabetes; vaccines; clinical trials; insulin; GLP

1. Introduction

Persisting as a major global health threat, diabetes mellitus (DM) affects individuals of all ages, ethnicities, and backgrounds, especially those associated with a prominent family history of diabetes and a multitude of environmental factors [1–4]. As reported by the

Citation: Chellappan, D.K.;

Bhandare, R.R.; Shaik, A.B.; Prasad, K.; Suhaimi, N.A.A.; Yap, W.S.; Das, A.; Banerjee, P.; Ghosh, N.; Guith, T.; et al. Vaccine for Diabetes—Where Do We Stand? *Int. J. Mol. Sci.* **2022**, *23*, 9470. https://doi.org/10.3390/ ijms23169470

Academic Editor: Wajid Zaman

Received: 31 July 2022 Accepted: 19 August 2022 Published: 22 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). World Health Organization (WHO), 422 million people globally suffer from diabetes and is rapidly progressing in intermediate and poverty-stricken nations [5]. Approximately 1.5 million deaths annually are caused by diabetes worldwide [5]. It has been reported that China contributed to the highest number of diabetics in 2021, with 149.1 million of its population between ages 20 and 79 being affected by this chronic disease. It is forecasted that China will have approximately 174 million diabetic patients by the year 2045. Meanwhile, a survey in 2014 by Kaveeshwar and Cornwall which reported an elevation approaching 8.3% in diabetic incidences further elucidates this observation [6–8]. Complications stemming from poorly managed DM represent a crucial cause of concern as a threat to mortality, indirectly impacting the economical status of a country [9]. The development of secondary complications worsens the mortality and morbidity caused by diabetes [10]. Initially, the classification for diabetes depended on its etiology and clinical course, before ultimately being categorized into Type 1 (T1DM) and Type 2 diabetes (T2DM), as the previous definition excluded many sufferers who exhibited atypical presentation and progression of the disease [11,12]. According to the American Diabetes Association, T1DM occurs due to defects in insulin production, whereas T2DM precipitates primarily from insulin resistance, followed by problematic reduction in insulin secretion, giving rise to hyperglycemia [13,14].

Orban et al. (2001) elucidated three phases whereby researchers may interrupt the underlying pathologic mechanisms behind T1DM, which are the autoimmunity development, autoantibody development, and clinical manifestation emergence with remaining residual β -cell function to be conserved [15]. The intention of halting these phases is to avert autoimmunity development in its initial stages as well as to inhibit clinical disease onset in high-risk individuals, since this phase is the root cause of the disease for the vast majority of patients [16]. However, as for the prevention of T2DM progression, the International Diabetes Federation (IDF), in 2006, proposed a method involving reduction in modifiable risk factors [17,18]. In terms of monitoring parameters, autoantibodies, such as insulin, insulinoma-associated protein 2 (IA-2), glutamic acid decarboxylase (GAD) or zinc transporter isoform 8 (ZT8), act as biomarkers to detect the preliminary onset of diabetes, as individuals who tested positive for more than 50% of these autoantibodies compared to single β -cell antigens are at a greater risk of developing T1DM [19,20]. On the other hand, β -cell destruction is mediated by different types of cytokines or by the direct activity of T-or B lymphocytes.

The pancreatic β -cell damage may be initiated by direct environmental toxins, a virus, or a primary immune attack against pancreatic β -cell antigens such as glutamic acid decarboxylase 65-kD antibody (GAD65). T-helper lymphocytes, such as CD4+, are activated by β -cell antigens and antigen-presenting cells, including the dendritic cells (DC) and macrophages. Interleukin (IL)-12 secreted by macrophages then stimulate the secretion of IL-2 and interferon (IFN)- γ by the CD4+ T-cells. IFN- γ then excites further resting macrophages to secrete other cytokines, such as the tumor necrosis factor (TNF- α), free radicals and IL-1 β , which are lethal for pancreatic β -cells. Additionally, activated T-helper cells produce cytokines which attract T- and B lymphocytes and trigger its multiplication in the islet of Langerhans, hence precipitating insulitis. With time, B lymphocytes would attack and harm the β cells by producing antibodies against secreted pancreatic β -cell antigens, whereas cytotoxic T-lymphocytes (CD8+) directly attack β cells which carry the target autoantigens [21–23].

As diabetes is a progressive disease, diabetic patients require effective, long-term treatment and the regular monitoring of treatment to achieve the suggested glycemic HbA1c levels. This management strategy may involve a combination of regimens of oral medicines, injectables, such as insulin or GLP-1 analogs, or both dosage forms. These combinational therapies or injectable therapies confer a high chance of inducing side effects, such as diarrhea or vomiting, with GLP-1 analogs and weight gain or hypoglycemia following insulin treatment. Although certain treatment regimens are unsuccessful at decreasing a patients' HbA1c to the desired level, the undesirable side effects of the medications itself causes patients to skip treatment, especially with higher doses, rendering the therapy ineffective. Presently, a patients' lack of adherence to their treatment plan remains a persistent clinical challenge, with over 50% of diabetic patients failing to strictly follow schedule of medication administration. In addition, although adherence to insulin treatment has improved in the past few years, due to the usage of pre-mixed formulas and smaller-sized needles, it remains sub-optimal at 63–65% [24,25].

Hence, in an effort to avert medication adherence problems for chronic diseases such as diabetes, there is a growing need for better prevention measures. Recent approaches other than intervening with environmental triggers to halt the onset of DM early has led to the discovery of effective vaccines. In this review article, we attempt to discuss advanced methods of diabetes prevention and the role of adjuvants in relation to vaccines. Some common practices for the prevention of diabetes at early stages is depicted in Figure 1. Ongoing debates and different opinions on various vaccine products made from proteins, antigens, and live pathogens were also examined. Besides that, we also reviewed other types of vaccines from different diseases which may be useful in paving the path for diabetes vaccine development.



Figure 1. Different modes of prevention of diabetes.

2. Vaccination

Vaccines elicit their responses in several ways: dampening the destructive Th1 immune response to a benign Th2 response, inciting antigen-specific T-reg cells, eradicating autoreactive T cells or arresting immune cell interaction [26]. A classification of different vaccines for treating diabetes is mentioned in Figure 2.

2.1. Early Diabetes Prevention

There are various approaches involved in preventing DM onset and progression, mainly by treating targeted individuals with a family history of diabetes or tenacious autoantibodies, intensive lifestyle interventions, the consumption of dietary fibers and the intake of vitamin D supplements [20,27–31]. However, the possibility of immunization being a method of prevention remains under-researched. At present, vaccines are employed as prophylactic measures in combating infectious diseases by using variants derived mainly from targeted live-attenuated pathogens. However, concerns pertaining to the safety profile of these vaccines has led to the investigation of more advanced bases known as adjuvants, which perform a key role in skewing immune responses and their fabrication [32–34].



Figure 2. A classification of different vaccines for treating diabetes.

2.2. Rationale behind Vaccine Adjuvant Action

A new era of vaccine development is presently emerging through novel combined therapy comprising adjuvants, which specifically activate and drive immune responses [35–38]. Traditionally, incorporating an adjuvant into a vaccine presents certain benefits, such as a reduced quantity of dose administered, leading to altered immune responses of greater quality, with minimal side effects [39–41]. A prime example of frequently used adjuvants includes alum adjuvants, which are readily available in the market today, as this compound assists in promoting humoral immunity in an individual [42,43].

3. Newly Designed Vaccine Products

Currently, several utilized vaccine products, such as autoantigens and non-autoantigenspecific therapies, are underway to be developed into vaccines for diabetes. A few of these have almost reached the final human testing stage. Hence, in this article, we aim to discuss various approaches incorporating prevention strategies with vaccines with respect to diabetes mellitus.

3.1. Protein-Based Approach in Vaccine Production

3.1.1. IL-1β-Targeted Epitope Peptide (1βEPP) as a New Vaccine Product for T2DM

Inflammation of the pancreatic islet in T2DM leading to β-cell apoptosis and disruption of insulin production is mainly caused by IL-1 β cytokine, a key mediator that induces insulin resistance within the peripheral tissue [44–48]. However, several studies have interestingly shown that IL-1 β is capable and has potential to be enhanced as a T2DM future therapy. To elaborate on this revelation, not only did a newly developed IL-1βtargeted epitope peptide vaccine adjuvant with polylactic acid microparticles (1 β EPP) stimulate the level of glucose tolerance and provide a hyperglycaemia shield when tested in diabetic KK-Aay mice model, but it also caused a reduction in the lipid profile and β-cell apoptosis action [49]. However, a few alterations were made to create a securely modified anti-IL-1 β to address issues derived from the phase I and II trial studies [50]. Another noteworthy research highlighting the use of a combination therapy (CT) of anti-IL-1 β and GAD65 DNA vaccine demonstrated the immense potential in reversing diabetes in its early stages [51]. Besides that, another vaccine, hlL1bQb was initially developed and tested in a preclinical simian phase before being tested in T2DM patients, where it was well documented that the hlL1bQb immunization caused harm to human subjects involved in the trial [52].

3.1.2. Dipeptidyl Peptidase-4 Inhibitor (DPP4) as Novel Vaccine Product for T2DM

Incretins are hormones which are secreted from enteroendocrine cells into the blood within minutes after food intake to regulate the amount of insulin to be secreted. Incretins essentially consist of two variants, one being the glucose-dependent insulinotropic peptide (GIP) and another being the glucagon-like peptide-1 (GLP-1). Although these hormones share numerous common actions in the pancreas, they exhibit very distinct actions outside of the pancreas. Since both types of incretins are quickly deactivated by a dipeptidyl peptidase-4 inhibitor (DPP4) enzyme, DPP4 inhibitors, on the other hand, raise the concentrations of these hormones, resulting in enhanced β -cell responsiveness to raised glucose concentrations as well as the suppression of glucagon secretion [53,54]. On the other hand, activation of glucagon-like-peptide (GLP) receptor agonists results in insulin secretion and caspase-mediated cell death inhibition in pancreatic β -cells through the action of DPP4. Earlier researchers recognized GLP and DPP4 as both efficacious and long-lasting agents for future approaches in treating T2DM [55–62], as elucidated by a study performed in 2014 which succeeded in synthesizing a DPP4 vaccine. When trials were performed in C57BL/6] mice model, results revealed a rise in GLP-1 level and an enhanced sensitivity of insulin without prompting adverse autoimmune responses. This occurrence is mainly observed in B- and T-cell epitopes, where a significant increase in anti-DPP4 antibody titre is detected along with the Th2 humoral response [63]. Recently, a study was conducted using D41-IP, a newly combinatorial peptide vaccine, synthesized using the B-cell epitope D41 within DPP4 and B-cell epitope of insulinoma antigen-2 (IA-2) in an effort to enhance therapeutic responses [64,65]. However, this multi-epitope study is perceived more as an alternative diabetic therapy, compared to the previous study, which focused on diabetes onset prevention [65].

3.1.3. CTB-InsB Vaccination Product to Treat T1DM

Bombyx mori, a classic host which secretes recombinant proteins in its fifth instar stage from the lumen of silk glands, has been used in silkworm biotechnology for ages [66–68]. This discovery by Dr. Maeda in 1985 originated from silkworm larvae and has drawn much attention due to its high level of recombinant protein expression [69]. Cholera toxin subunit B (CTB), which is known for its toxic characteristics, is usually used as strong adjuvant, which, together with its antigen coupling action, will eventually lead to a down-regulation response in the onset of T1DM [70]. On the other hand, the B chain of insulin (INSB) is briefly known as a 30-amino-acid chain with an immunogenic epitope. When used in combination, both CTB and INSB make up a consumable vaccine to induce immune tolerance in diabetic patients, with its strong influences in evading insulitis when tested in NOD mice. This specific tolerance increases Foxp3+ regulatory T-cell proportions in peripheral lymph tissues and suppresses the biological functions of spleen lymphocytes in mice. This important research proved the effectiveness of the CTB-InsB oral protein vaccine against diabetes development [71–75].

3.2. Specific Self-Antigens Approach in Vaccine Production

Upon administration and absorption of the vaccine-adjuvant into the T1DM patients' skin, toll-like-receptor (TLR) acts on depot-containing antigens through pattern recognition receptors (PRRs) activation, which mainly leads to antigen presenting cells (APC) maturation (primarily DC). Activated APCs on major histocompatibility complexes (MHC) surface then interact with antigen-specific T cells and secrete IL-10 cytokine to suppress Th1 by Th0 stimulation. Two pathways—the Th2 anti-inflammatory process and Treg cells induction that suppresses Th cells development—eventually stimulate insulin secretion as diagrammatically shown in Figure 3 [36,76–78].



Figure 3. T1DM and vaccine adjuvant's mechanisms of action.

3.2.1. IA-2 as New Vaccine Product for T1DM

The pharmacological action of the D41-IP vaccine in using IA-2 protein as an islet autoantigen has been tested in few different studies [79]. IA-2, a tyrosine phosphatase protein, is commonly known as a T1DM major islet antigen [80]. A study conducted in 2012 by Guan et al. claimed that the IA-2 vaccination is capable of delaying the onset and the late stages of autoimmune diabetes either on its own or when co-administered with plasmid IL-4/MCP-1, which proposed a promising future for T1DM patients [81,82]. Later, a newly designed novel peptide vaccine, IA-2-P2, was introduced to the public, where the overall idea of the vaccine was initiated based on the previous findings of Guan et al. A drastic drop in the blood glucose levels of normoglycemic mice were obtained when tested with the IA-2-P2 vaccine in comparison to P277 and control mice in the study. Hence, it was concluded that IA-2-P2 was a suitable ameliorate vaccine to combat T1DM. The P277 peptide, a human 60 kDa heat shock protein (hsp60) is a causative factor for the onset of diabetes of non-obese diabetic (NOD) mice which are genetically prone to developing spontaneous autoimmune diabetes [83]. However, Lu et al. reported that the fusion of the His-Hsp65-6IA2P2 protein vaccine through nasal inoculation is believed to serve well in regulating the T1DM response [84]. Apart from IA-2, the zinc transporter (ZnT8) has been also studied as a major autoantigen target in T1DM immunotherapy. However, no investigations have been conducted to test ZnT8 as a diabetes vaccine in trials [85–88].

3.2.2. Glutamic Acid Decarboxylase 65-kD (GAD65): A New Vaccine Product for T1DM

The GAD65 antibody is an isoform of GAD targeted by self-reactive T cells that exhibits susceptibility marker detection in T1DM more frequently as compared to the IA-2 autoantigen [89,90]. On the other hand, aluminum hydroxide is the most used adjuvant [91,92]. In 2011, the GAD-alum vaccine (Diamyd) comprising GAD and aluminum hydroxide, was introduced, in which the efficacy and safety of the vaccine was tested in phase II trials preceding four years of close pharmacovigilance [93–99]. However, no desirable effects were observed within T1DM subjects, although two to three drops of injected vaccine were used in each subject throughout the three trial stages of experiment, rendering it ineffective [100–102]. In fact, HbA1c and insulin were not altered by the GAD-alum treatment [100]. In addition to that, a review article by Cook et al. also highlighted the insufficiency of the GAD-alum vaccine when tested in clinical trials of larger sample sizes [103]. However, the combination of CTB-insulin and CTB-GAD with IL-10 as a newly proposed multi-component vaccine has proven to suppress β -cell autoreactivity in T1DM [104]. Other studies also revealed that GAD65 antibodies elicit activity against glial

fibrillary acidic protein (GFAP), a predictive biomarker that is expressed within peri-islet Schwann cells in the event of the onset of T1DM, making it a suitable molecule to be incorporated in the production of an immune tolerizing vaccine [105,106]. Hyperglycemia was suppressed, whereas C-peptide secretion was enhanced significantly in T1DM using this GFAP vaccine by acting upon T-cell entrance into pancreatic islets, which subsequently shifts T-cell differentiation from a cytotoxic Th1- to a Th2-biased humoral response in NOD mice [107]. Considering the complex mechanism of the action of self-antigens toward the immune system, synthetic materials were used instead in one study to co-deliver the immunomodulatory signals, where fabricated microparticle (MP) vaccines were recently established via in vivo and in vitro methods. Therefore, the first biomaterial-based vaccine product, hydrogel/microparticle, was introduced by using dual subcutaneous immunization, and subsequently, a revised version involving the delivery of three shots of the vaccine into NOD mice models was tested [108,109]. On the other hand, Phillips et al. successfully created the first antisense oligonucleotide-formulated microsphere vaccine capable of suppressing diabetes and boosting Foxp3+ T-reg cells without inducing any unfavorable responses [110,111]. Multiple benefits were observed from these autoantigenspecific interventions over those involving nonspecific immune suppression as an immune tolerance therapy, including the reduction in pathogenic peptide epitope response and any related side effects [112]. Even so, it is unable to be predicted if similar favorable outcomes may be obtained from the phase III clinical trial [113]. In a recent double-blinded, randomized, placebo-controlled Phase IIb clinical trial, the intra-lymphatic administration of GAD-alum with vitamin D supplementation appears to preserve C-peptide in patients with recent-onset T1D carrying HLA DR3-DQ2 [114].

3.2.3. Insulin as a Target in New Vaccine Product for T1DM

The BHT-3021 vaccine, which is made of proinsulin, a precursor insulin prohormone, was proven effective in enhancing insulin production within its early developmental stage [115]. Longer-term research for this vaccine was recently announced by NHS Choice in which larger group of 200 participants will be involved. In another study, the reaction between insulin-like growth factor 2 (IGF-2), an insulin dominant self-antigen, and an insulin autoantigen initiated a response resembling negative/tolerogenic self-vaccination, indicating a possible cure for T1DM, as reported by Chentoufi [116]. To further consolidate this revelation, other studies have proven that self-antigen vaccination is one of the most secure strategies against autoimmune diabetes [117].

3.3. Non-Antigen Specific Approach in Vaccine Production

Certain immunologically active microbes and their products have been reported to prevent autoimmune diabetes in different animal models [118]. These agents may confer a protective effect in humans by stimulating the immune system especially during childhood development [118]. As per "the hygiene hypothesis," the growing cases of autoimmune diseases may be caused by insufficient microbial exposure due to the improved hygienic conditions of the developed world [118]. Some of these microbial approaches are discussed below.

3.3.1. Live Pathogen Salmonella as Vector Vaccine

Live recombinant attenuated *Salmonella*-vectored vaccines exhibit great potential as resources to improve human health by achieving long-lasting mucosal, humoral and cellular immunity against a variety of non-*Salmonella* pathogens at a low cost. The use of recombinant DNA has been a major breakthrough in antigen mucosal delivery for years through the generation of a live attenuated *Salmonella* oral vaccine since it was initially tested in phase I clinical trials [119]. It is the T-cell autoreactive downregulation response in the *Salmonella* vaccine which conferred greater therapeutic effectiveness besides its simplicity and relative safety in comparison to antigen-specific approaches, thus making it a potential future therapy for T1DM despite an uncertain mechanism of action [120].

A current study utilized *Salmonella typhimurium* bacteria in combination with other small regulatory proteins called cytokines and a low dose of an immunosuppressive drug, Anti-CD3. Results revealed that the vaccine reinstates balance to the immune system and prevents the attack of insulin-producing cells.

3.3.2. Inactivated Microbial Vaccines

Apart from alum and CTB adjuvants, complete Freund's adjuvant (CFA) has also been studied as a vital T-regulatory cell inducer which successfully reverses autoimmune diabetes through multicomponent immunization. However, the injection of a combined anti-CXCL10 vaccine with CFA is not advisable for human use due to its high toxicity profile, despite its proven effectiveness, as elucidated in a study by Oikawa et al. (2010), whereby the anti-CXCL10 vaccine used successfully reversed T1DM [121]. In few years, further studies involving CFA were performed using CTB:GAD fusion protein, which successfully induced the protective effects in tested NOD mice [122]. In more recent times, a multivalent islet lysate-negative vaccine tested elicited a positive immunogenic response in the pathophysiology of diabetes, in which incomplete Freund's adjuvant (IFA) was used [123].

3.3.3. BCG as Vector Vaccine in Clinical Trial Studies

In 2001, tumor necrosis factor (TNF- α) activation following BCG vaccine administration to restore endogenous β -cell function resulted in the discovery of a potential T1DM reversal mechanism [124]. In addition to that, Faustman (2012) developed another BCG vaccine which was tested in a phase I randomized control trial, resulting in a proven ability of the BCG vaccine in triggering TNF to induce apoptosis in autoreactive T cells. Furthermore, an increase in the restoration rate of pancreatic β-cell function was observed in a rodent model response [125]. Likewise, a minimum of two doses of the BCG vaccine is recommended, particularly with the first dose being administered in neonates. However, further studies pertaining to this claim is required, as its mechanism of action in human subjects remains unclear [126,127]. Nevertheless, no definite linkage was found in the preservation of β -cell function in T1DM after prophylactic administration of the BCG as hypothesized by other randomized clinical studies between the years 1999 and 2005 [128]. In contrast, approval by the Food and Drug Administration (FDA) led to the commencement of a phase II clinical trial study regarding the activity of the BCG vaccine related to T1DM reversal. This study design involves 130 participants, of which approximately over 100 candidates have received a minimum of one dose of the BCG vaccine, and progress or response is actively being monitored for five years from the date of administration.

4. Potential of Other Disease-Vaccines in Treating Diabetes

Interestingly, previous research elucidated evidence for enteroviruses (EV) being a causative factor for the onset of type I diabetes. Hence, avid and ongoing effort in conducting research and development to develop and synthesize an EV vaccine is currently in progress with the aid of technological advances. Yet, various issues are still under discussion, such as diabetogenic EV serotypes, safety concerns pertaining to it, and relevance as well as accuracy of the accumulated literature reviews, before this novel vaccine can be tested in clinical trials [129,130]. However, a noteworthy accomplishment was achieved in a recent preclinical study involving the invention of the first multivalent formalin inactivated CVB1 vaccine, where the vaccine proved to be effective with no adverse effects [131]. Apart from the usage of vaccines indicated for viral infections, the tuberculosis DNA vaccine known as DNA-HSP65 exemplifies high possibility in becoming the latest immunotherapeutic agent for the management of diabetes [132]. A list of new vaccine products in the pipeline that could be employed for diabetes can be found in Table 1.

Int. J. Mol. Sci. 2022, 23, 9470

TAUTE	I. FIST OF VACUUE PLOUDERS 101 MADELES.			
Type of Vaccine	Vaccine Name	Indication	Mechanism of Action	References
Adjuvant-based vaccines	Alum adjuvant-based vaccine	T1DM and T2DM	Activation of immune response; Promoting humoral immunity	[42,43]
Protein-based vaccines	IL-1 β -targeted epitope peptide (1 β EPP)	T2DM	Alters the level of glucose tolerance and provides a hyperglycaemia shield	[49]
	Dipeptidyl peptidase-4 inhibitor (DPP4) based vaccine	T2DM	Inhibition of dipeptidyl peptidase-4 inhibitor (DPP4) enzyme	[53,54]
	CTB-InsB vaccination product	T1DM	Down-regulation response in the onset of T1DM; Induction of immune tolerance	[71–75]
Specific self-antigen-based approach in vaccine production	IA-2 as a vaccine product	T1DM	Islet autoantigen mechanism; delaying the onset and the late stages of autoimmune diabetes	[81,82]
	Glutamic Acid Decarboxylase 65-kD (GAD65)-Alum vaccine	T1DM	Suppression of β -cell autoreactivity	[104]
	Insulin autoantigen-based vaccine	T1DM	Enhancing insulin production	[115]
Non-Antigen-based vaccines	Live pathogen Salmonella-based vector vaccine	T1DM	T-cell autoreactive downregulation response	[119]
	Inactivated microbial vaccine	T1DM	Positive immunogenic response induction	[121]
	BCG as a vector vaccine	T1DM	Restoration of endogenous β -cell function	[124]
Other Disease-Based Vaccines	Enteroviruses (EV)-based vaccine	T1DM	Delayed onset response	[129,130]

Table 1. List of vaccine products for diabetes.
Future Perspectives

Despite various prospects and the effectiveness of each approach discussed previously, limitations which arise should be thoroughly considered and evaluated from various perspectives, where modifications may be made for future immunotherapies and later implemented in phase III and IV trials. To date, the majority of the diabetes vaccine development has been studied using animal models, with relatively few human trials [26]. A major limitation of using animal models, such as NOD mice, is their profound sensitivity to diabetes protection. As a result, several successful animal studies failed in human trials, including the Diabetes Prevention Trial-1 [26]. Recently studies have suggested the administration of non-depleting anti-CD3 antibodies or a peptide from heat shock protein 60 to be beneficial against a recent onset of T1DM [26]. The revolutionary Diamyd vaccine for the prevention of diabetes, for example, could possibly be improved via the exploration of other antigen delivery pathways, such as self-antigen DNA vaccine administration, with consideration of divergent dose administration or replacement with a different type of autoantigen [133,134]. Besides that, understanding various other routes of administration could potentially be an area of further research. Being another pivotal part of vaccine development, feasible study designs not only contribute toward grasping a better understanding of the Alum-GAD system but also could pave the path to develop newer combination therapies studies for diabetes.

5. Conclusions

The pressing concern pertaining to the rising number of patients falling victim to diabetes has garnered the interest of numerous scientists and researchers worldwide in the search for the most effective management strategies, as demonstrated with the abundance of literature available to date. The urgent need for more effective prophylaxis in addition to conventional dietary modification advice to patients has prompted various attempts in developing vaccines to delay or prevent the onset of this chronic disease, as seen in recent approaches where protein-based, self-antigen and non-antigen specific interventions have exhibited promising potential for use as vaccines against diabetes in the near future. However, a deeper understanding is still required to ameliorate and invent potent therapies with minimal side effects regardless of the cause-related factors, especially for chronic diseases, such as T2DM.

Author Contributions: Conceptualization, D.K.C.; resources, A.B.S., K.P. (Krishna Prasad), N.A.A.S., W.S.Y., A.D. (Arpita Das), P.B., N.G., T.G., A.D. (Amitava Das), S.B., M.C., J.M., K.P. (Kishneth Palaniveloo), G.G., S.K.S. and K.D.; data curation, A.B.S., K.P. (Krishna Prasad), N.A.A.S., W.S.Y., A.D. (Arpita Das), P.B., N.G., T.G., A.D. (Amitava Das), S.B., M.C., J.M., K.P., G.G., S.K.S. and K.D.; writing—original draft preparation, A.B.S., K.P. (Krishna Prasad), N.A.A.S., W.S.Y., A.D. (Arpita Das), P.B., N.G., T.G., A.D. (Amitava Das), S.B., M.C., J.M., K.P., G.G., S.K.S. and K.D.; writing—original draft preparation, A.B.S., K.P. (Krishna Prasad), N.A.A.S., W.S.Y., A.D. (Arpita Das), P.B., N.G., T.G., A.D. (Amitava Das), S.B., M.C., J.M., K.P. (Krishna Prasad), G.G., S.K.S. and K.D.; writing—review and editing, D.K.C., R.R.B., A.B.S., K.P. (Krishna Prasad), N.A.A.S., W.S.Y., A.D. (Arpita Das), P.B., N.G., T.G., A.D. (Amitava Das), S.B., M.C., J.M., K.P. (Krishna Prasad), S.K.S. and K.D.; writing—review and editing, D.K.C., R.R.B., A.B.S., K.P. (Krishna Prasad), N.A.A.S., W.S.Y., A.D. (Arpita Das), P.B., N.G., T.G., A.D. (Amitava Das), S.B., M.C., J.M., K.P. (Kishneth Palaniveloo), G.G., S.K.S. and K.D.; supervision, D.K.C.; project administration, D.K.C.; funding acquisition, R.R.B. All authors have read and agreed to the published version of the manuscript.

Funding: The Article Processing Charges (APC) of this review is funded by the Deanship of Graduate Studies and Research, Ajman University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: RRB is thankful for Deanship of Graduate Studies and Research, Ajman University for providing funding towards the Article Processing Charges (APC).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

16EPP	IL-16-targeted epitope peptide vaccine adjuvant with polylactic acid microparticles
ADA	American Diabetes Association
Alum	Aluminum
APC	Antigen presenting cells
BCG	Bacillus Calmette–Guerin
CFA	Complete Freund's adjuvant
CTB	Cholera toxin B subunit
DC	Dendritic cell
DM	Diabetes mellitus
DNA	Deoxyribonucleic acid
DPP4	Dipeptidyl-peptidase-4-inhibitor
EV	Enteroviruses
FDA	Food and Drug Administration
GAD65	Glutamic acid decarboxylase 65-kD antibody
GFAP	Glial fibrillary acidic protein
GLP	Glucagon-like-peptide
IA-2	Insulinoma-associated protein 2
IDF	International Diabetes Federation
IFA	Incomplete Freund's adjuvant
INSB	B chain of insulin
IL-1β	Interleukin-1β cytokine
IL-10	Interleukin-10
MHC	Major histocompatibility complexes
MP	Microparticle
NOD	Non-obese diabetic
PRR	Pattern recognition receptors
T1DM	Type 1 diabetes mellitus
T2DM	Type 2 diabetes mellitus
Th0	Naive helper T cell
Th1	Cytotoxic helper T cell
Th2	Humoral helper T cell
TLR	Toll-like receptor
TNF	Tumor necrosis factor
Treg cells	Regulatory T cells
WHO	World Health Organization
ZnT8	Zinc transporter isoform 8

References

- 1. Nair, M. Diabetes mellitus, part 1: Physiology and complications. Br. J. Nurs. 2007, 16, 184–188. [CrossRef] [PubMed]
- 2. Rewers, M.; Gottlieb, P. Immunotherapy for the Prevention and Treatment of Type 1 Diabetes: Human trials and a look into the future. *Diabetes Care* **2009**, *32*, 1769–1782. [CrossRef]
- 3. Alwan, A. Raising the priority accorded to diabetes in global health and development: A promising response. *Int. J. Diabetes Mellit.* **2010**, *3*, 139–140. [CrossRef]
- 4. Slim, I.B.H.S. Cardiovascular risk in type 1 diabetes mellitus. Indian J. Endocrinol. Metab. 2013, 17 (Suppl. 1), S7–S13. [CrossRef]
- 5. Diabetes. Available online: http://www.who.int/health-topics/diabetes#tab=tab_1 (accessed on 1 June 2022).
- 6. Hanson, M.A.; Gluckman, P.D.; Ma, R.C.; Matzen, P.; Biesma, R.G. Early life opportunities for prevention of diabetes in low and middle income countries. *BMC Public Health* **2012**, *12*, 1–9. [CrossRef] [PubMed]
- 7. Kaveeshwar, S.A.; Cornwall, J. The current state of diabetes mellitus in India. Australas Med. J. 2014, 7, 45–48. [CrossRef]
- 8. Maffi, P.; Secchi, A. The Burden of Diabetes: Emerging Data. Dev. Ophthalmol. 2017, 60, 1–5.
- Deshpande, A.D.; Harris-Hayes, M.; Schootman, M. Epidemiology of diabetes and diabetes-related complications. *Phys. Ther.* 2008, *88*, 1254–1264. [CrossRef]
- 10. Bagchi, D.; Das, A.; Roy, S. Wound Healing, Tissue Repair, and Regeneration in Diabetes; Academic Press: London, UK, 2020; p. 609.
- 11. Rother, K.I. Diabetes treatment-bridging the divide. N. Engl. J. Med. 2007, 356, 1499-1501. [CrossRef]
- 12. Piero, M.N.; Nzaro, G.M.; Njagi, J.M. Diabetes mellitus-a devastating metabolic disorder. Asian J. Biomed. Pharm. Sci. 2015, 5, 1.

- 13. Staeva, T.P.; Chatenoud, L.; Insel, R.; Atkinson, M.A. Recent lessons learned from prevention and recent-onset type 1 diabetes immunotherapy trials. *Diabetes* **2013**, *62*, 9–17. [CrossRef] [PubMed]
- 14. DeFronzo, R.A.; Ferrannini, E.; Groop, L.; Henry, R.R.; Herman, W.H.; Holst, J.J.; Hu, F.B.; Kahn, C.R.; Raz, I.; Shulman, G.I.; et al. Type 2 diabetes mellitus. *Nat. Rev. Dis. Prim.* **2015**, *1*, 1–22. [CrossRef] [PubMed]
- 15. Orban, T.; Kent, S.C.; Malik, P.; Milner, J.D.; Schuster, K.; Jackson, R.A.; Hafler, D.A. Heterophile antibodies indicate progression of autoimmunity in human type 1 diabetes mellitus before clinical onset. *Autoimmunity* **2001**, *34*, 247–264. [CrossRef] [PubMed]
- 16. Orban, T.; Kis, J.T. Prevention of type 1 diabetes mellitus using a novel vaccine. *Ther. Adv. Endocrinol. Metab.* **2011**, *2*, 9–16. [CrossRef] [PubMed]
- Alberti, K.G.M.M.; Zimmet, P.; Shaw, J. International Diabetes Federation: A consensus on Type 2 diabetes prevention. *Diabet. Med.* 2007, 24, 451–463. [CrossRef]
- Li, G.; Zhang, P.; Wang, J.; Gregg, E.W.; Yang, W.; Gong, Q.; Li, H.; Li, H.; Jiang, Y.; An, Y.; et al. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: A 20-year follow-up study. *Lancet* 2008, 371, 1783–1789. [CrossRef]
- 19. Van Belle, T.L.; Coppieters, K.T.; Von Herrath, M.G. Type 1 diabetes: Etiology, immunology, and therapeutic strategies. *Physiol. Rev.* **2011**, *91*, 79–118. [CrossRef]
- 20. Lernmark, Å.; Larsson, H.E. Immune therapy in type 1 diabetes mellitus. Nat. Rev. Endocrinol. 2013, 9, 92–103. [CrossRef]
- Gelber, C.; Paborsky, L.; Singer, S.; McAteer, D.; Tisch, R.; Jolicoeur, C.; Buelow, R.; McDevitt, H.; Fathman, C.G. Isolation of nonobese diabetic mouse T-cells that recognize novel autoantigens involved in the early events of diabetes. *Diabetes* 1994, 43, 33–39. [CrossRef]
- Maclaren, N.; Lan, M.; Coutant, R.; Schatz, D.; Silverstein, J.; Muir, A.; Clare-Salzer, M.; She, J.X.; Malone, J.; Crockett, S.; et al. Only multiple autoantibodies to islet cells (ICA), insulin, GAD65, IA-2 and IA-2β predict immune-mediated (type 1) diabetes in relatives. J. Autoimmun. 1999, 12, 279–287. [CrossRef]
- Sabbah, E.; Savola, K.; Kulmala, P.; Veijola, R.; Vahasalo, P.; Karjalainen, J.; Akerblom, H.K.; Knip, M. Childhood Diabetes in Finland Study Group. Diabetes-associated autoantibodies in relation to clinical characteristics and natural course in children with newly diagnosed type 1 diabetes. J. Clin. Endocrinol. Metab. 2011, 84, 1534–1539.
- 24. Montgomery, S.M.; Ehlin, A.G.C.; Ekbom, A.; Wakefield, A.J. Pertussis infection in childhood and subsequent type 1 diabetes mellitus. *Diabet. Med.* **2002**, *19*, 986–993. [CrossRef] [PubMed]
- Blom, L.; Nystrom, L.; Dahlquist, G. The Swedish childhood diabetes study. Vaccinations and infections as risk determinants for diabetes in childhood. *Diabetologia* 1991, 34, 176–181. [CrossRef] [PubMed]
- 26. Petrovsky, N.; Silva, D.; Schatz, D.A. Vaccine therapies for the prevention of type 1 diabetes mellitus. *Paediatr Drugs* **2003**, *5*, 575–582. [CrossRef]
- 27. Fonseca, V.A. Early identification and treatment of insulin resistance: Impact on subsequent prediabetes and type 2 diabetes. *Clin. Cornerstone* **2007** *8*, S7–S18. [CrossRef]
- Weickert, M.O.; Pfeiffer, A.F. Metabolic effects of dietary fiber consumption and prevention of diabetes. J. Nutr. 2008, 138, 439–442. [CrossRef]
- 29. Zipitis, C.S.; Akobeng, A.K. Vitamin D supplementation in early childhood and risk of type 1 diabetes: A systematic review and meta-analysis. *Arch. Dis. Child* **2008**, *93*, 512–517. [CrossRef]
- Narayan, K.M.V.; Williamson, D.F. Prevention of Type 2 Diabetes: Risk Status, Clinic, and Community. J. Gen. Intern. Med. 2010, 25, 154–157. [CrossRef]
- 31. Saaristo, T.; Moilanen, L.; Korpi-Hyovalti, E.; Vanhala, M.; Saltevo, J.; Niskanen, L.; Jokelainen, J.; Peltonen, M.; Oksa, H.; Tuomilehto, J.; et al. Lifestyle intervention for prevention of type 2 diabetes in primary health care: One-year follow-up of the Finnish National Diabetes Prevention Program (FIN-D2D). *Diabetes Care* 2010, *33*, 2146–2151. [CrossRef]
- 32. Nicholls, E.F.; Madera, L.; Hancock, R.E. Immunomodulators as adjuvants for vaccines and antimicrobial therapy. *Ann. N. Y. Acad. Sci.* **2010**, *1213*, 46–61. [CrossRef]
- 33. Combadiere, B.; Mahe, B. Particle-based vaccines for transcutaneous vaccination. *Comp. Immunol. Microbiol. Infect. Dis.* 2008, 31, 293–315. [CrossRef] [PubMed]
- 34. De Gregorio, E.; Caproni, E.; Ulmer, J.B. Vaccine adjuvants: Mode of action. Front. Immunol. 2013, 4, 214. [CrossRef] [PubMed]
- 35. Foged, C. Subunit vaccines of the future: The need for safe, customized and optimized particulate delivery systems. *Ther. Deliv.* **2011**, *2*, 1057–1077. [CrossRef]
- 36. Schijns, V.E.; Lavelle, E.C. Trends in vaccine adjuvants. Expert Rev. Vaccines 2011, 10, 539–550. [CrossRef]
- 37. Wu, T.Y.H.; Singh, M.; Miller, A.T.; De Gregorio, E.; Doro, F.; D'Oro, U.; Skibinski, D.A.; Mbow, M.L.; Bufali, S.; Herman, A.E.; et al. Rational design of small molecules as vaccine adjuvants. *Sci. Transl. Med.* **2014**, *6*, 263ra160. [CrossRef]
- 38. O'Hagan, D.T.; Fox, C.B. New generation adjuvants-from empiricism to rational design. *Vaccine* 2015, 33, B14–B20. [CrossRef]
- 39. Aguilar, J.C.; Rodriguez, E.G. Vaccine adjuvants revisited. *Vaccine* 2007, 25, 3752–3762. [CrossRef]
- 40. Coffman, R.L.; Sher, A.; Seder, R.A. Vaccine adjuvants: Putting innate immunity to work. Immunity 2010, 33, 492–503. [CrossRef]
- 41. Reed, S.G.; Orr, M.T.; Fox, C.B. Key roles of adjuvants in modern vaccines. Nat. Med. 2013, 19, 1597–1608. [CrossRef]
- 42. Lambrecht, B.N.; Kool, M.; Willart, M.A.; Hammad, H. Mechanism of action of clinically approved adjuvants. *Curr. Opin. Immunol.* **2009**, *21*, 23–29. [CrossRef]

- 43. Oleszycka, E.;Lavelle, E.C. Immunomodulatory properties of the vaccine adjuvant alum. *Curr. Opin. Immunol.* **2014**, *28*, 1–5. [CrossRef] [PubMed]
- 44. Larsen, C.M.; Faulenbach, M.; Vaag, A.; Vølund, A.; Ehses, J.A.; Seifert, B.; Mandrup-Poulsen, T.; Donath, M.Y. Interleukin-1–receptor antagonist in type 2 diabetes mellitus. *N. Engl. J. Med.* **2007**, *356*, 1517–1526. [CrossRef] [PubMed]
- 45. Ren, K.; Torres, R. Role of interleukin-1beta during pain and inflammation. *Brain Res. Rev.* 2009, *60*, 57–64. [CrossRef] [PubMed]
 46. Lopez-Castejon, G.; Brough, D. Understanding the mechanism of IL-1β secretion. *Cytokine Growth Factor Rev.* 2011, *22*, 189–195.
- [CrossRef]47. Banerjee, M. Saxena, M. Interleukin-1 (IL-1) family of cytokines: Role in type 2 diabetes. *Clin. Chim. Acta* 2012, 413, 1163–1170.
- [CrossRef] [PubMed]
 48. Handa, M.; Vanegas, S.; Maddux, B.A.; Mendoza, N.; Zhu, S.; Goldfine, I.D.; Mirza, A.M. XOMA 052, an anti-IL-1β monoclonal
- antibody, prevents IL-1 β -mediated insulin resistance in 3T3-L1 adipocytes. *Obesity* **2013**, *21*, 306–309. [CrossRef] [PubMed]
- Zha, J.; Chi, X.W.; Yu, X.L.; Liu, X.M.; Liu, D.Q.; Zhu, J.; Ji, H.; Liu, R.T. Interleukin-1β-targeted vaccine improves glucose control and β-cell function in a diabetic KK-Ay mouse model. *PLoS ONE* 2016 11, e0154298. [CrossRef]
- 50. Spohn, G.; Schori, C.; Keller, I.; Sladko, K.; Sina, C.; Guler, R.; Schwarz, K.; Johansen, P.; Jennings, G.T.; Bachmann, M.F. Preclinical efficacy and safety of an anti-IL-1β vaccine for the treatment of type 2 diabetes. *Mol.-Ther. Clin. Dev.* **2014**, *1*, 14048. [CrossRef]
- 51. Pagni, P.P.; Bresson, D.; Rodriguez-Calvo, T.; Bel Hani, A.; Manenkova, Y.; Amirian, N.; Blaszczak, A.; Faton, S.; Sachithanantham, S.; von Herrath, M.G. Combination therapy with an anti–IL-1β antibody and GAD65 DNA vaccine can reverse recent-onset diabetes in the RIP-GP mouse model. *Diabetes* 2014, *63*, 2015–2025. [CrossRef]
- 52. Cavelti-Weder, C.; Timper, K.; Seelig, E.; Keller, C.; Osranek, M.; Lassing, U.; Spohn, G.; Maurer, P.; Muller, P.; Jennings, G.T.; et al. Development of an interleukin-1β vaccine in patients with type 2 diabetes. *Mol. Ther.* **2016**, *24*, 1003–1012. [CrossRef]
- 53. Kim, W.; Egan, J.M. The Role of Incretins in Glucose Homeostasis and Diabetes Treatment. *Pharmacol. Rev.* **2008**, *60*, 470–512. [CrossRef] [PubMed]
- 54. Vella, A. Mechanism of action of DPP-4 inhibitors-new insights. J. Clin. Endocrinol. Metab. 2012, 97, 2626–2628. [CrossRef] [PubMed]
- 55. Langley, A.K.; Suffoletta, T.J.; Jennings, H.R. Dipeptidyl peptidase IV inhibitors and the incretin system in type 2 diabetes mellitus. *Pharmacotherapy* **2007**, *27*, 1163–1180. [CrossRef] [PubMed]
- 56. Madsbad, S.; Krarup, T.; Deacon, C.F.; Holst, J.J. Glucagon-like peptide receptor agonists and dipeptidyl peptidase-4 inhibitors in the treatment of diabetes: A review of clinical trials. *Curr. Opin. Clin. Nutr. Metab. Care* **2008**, *11*, 491–499. [CrossRef] [PubMed]
- 57. Pacini, F. Thyroid nodules and cancer. Preface. Best Pract. Res. Clin. Endocrinol. Metab. 2008, 22, vii. [CrossRef]
- Davidson, J.A. Advances in therapy for type 2 diabetes: GLP-1 receptor agonists and DPP-4 inhibitors. *Cleve Clin. J. Med.* 2009, 76, S28–S38.
- Dicker, D. DPP-4 inhibitors: Impact on glycemic control and cardiovascular risk factors. *Diabetes Care* 2011, 34, S276–S278. [CrossRef]
- 60. Donnelly, D. The structure and function of the glucagon-like peptide-1 receptor and its ligands. *Br. J. Pharmacol.* **2012**, *166*, 27–41. [CrossRef]
- 61. Li, Y.; Xiao, J.; Tian, H.; Pei, Y.; Lu, Y.; Han, X.; Liu, Y.; Zhong, W.; Sun, B.; Fang, F.; et al. The DPP-4 inhibitor MK0626 and exercise protect islet function in early pre-diabetic kkay mice. *Peptides* **2013**, *49*, 91–99. [CrossRef]
- 62. Pratley, R.E.; Nauck, M.; Bailey, T.; Montanya, E.; Cuddihy, R.; Filetti, S.; Thomsen, A.B.; Sondergaard, R.E.; Davies, M.; 1860-LIRA-DPP-4 Study Group. Liraglutide versus sitagliptin for patients with type 2 diabetes who did not have adequate glycaemic control with metformin: A 26-week, randomised, parallel-group, open-label trial. *The Lancet* **2010**,375, 1447–1456. [CrossRef]
- 63. Pang, Z.; Nakagami, H.; Osako, M.K.; Koriyama, H.; Nakagami, F.; Tomioka, H.; Shimamura, M.; Kurinami, H.; Takami, Y.; Morishita, R.; et al. Therapeutic vaccine against DPP4 improves glucose metabolism in mice. *Proc. Natl. Acad. Sci. USA* **2014**, 111, E1256–E1263. [CrossRef] [PubMed]
- 64. Bresson, D.; Fradkin, M.; Manenkova, Y.; Rottembourg, D.; Von Herrath, M. Genetic-induced variations in the GAD65 T-cell repertoire governs efficacy of anti-CD3/GAD65 combination therapy in new-onset type 1 diabetes. *Mol. Ther.* **2010**, *18*, 307–316. [CrossRef] [PubMed]
- 65. Li, Z.; Fang, J.; Jiao, R.; Wei, X.; Ma, Y.; Liu, X.; Cheng, P.; Li, T. A novel multi-epitope vaccine based on Dipeptidyl Peptidase 4 prevents streptozotocin-induced diabetes by producing anti-DPP4 antibody and immunomodulatory effect in C57BL/6J mice. *Biomed. Pharmacother.* **2017**, *89*, 1467–1475. [CrossRef] [PubMed]
- 66. Mondal, M.; Trivedy, K.; Nirmal Kumar, S. The silk proteins, sericin and fibroin in silkworm, Bombyx Mori Linn.- A review. *Casp. J. Environ. Sci.* 2007, *5*, 63–76.
- 67. Kato, T.; Kajikawa, M.; Maenaka, K.; Park, E.Y. Silkworm expression system as a platform technology in life science. *Appl. Microbiol. Biotechnol.* **2010**, *85*, 459–470. [CrossRef]
- 68. Dong, Z.; Zhao, P.; Zhang, Y.; Song, Q.; Zhang, X.; Guo, P.; Wang, D.; Xia, Q. Analysis of proteome dynamics inside the silk gland lumen of Bombyx mori. *Sci. Rep.* **2016**, *6*, 21158. [CrossRef]
- 69. Park, E.Y.; Abe, T.; Kato, T. Improved expression of fusion protein using a cysteine-protease-and chitinase-deficient Bombyx mori (silkworm) multiple nucleopolyhedrovirus bacmid in silkworm larvae. *Biotechnol. Appl. Biochem.* **2008**, *49*, 135–140. [CrossRef]
- 70. Baldauf, K.J.; Royal, J.M.; Hamorsky, K.T.; Matoba, N. Cholera toxin B: One subunit with many pharmaceutical applications. *Toxins* **2015**, *7*, 974–996. [CrossRef]

- 71. Gong, Z.; Jin, Y.; Zhang, Y. Suppression of diabetes in non-obese diabetic (NOD) mice by oral administration of a cholera toxin B subunit–insulin B chain fusion protein vaccine produced in silkworm. *Vaccine* **2007**, *25*, 1444–1451. [CrossRef]
- 72. Langridge, W.; Denes, B.; Fodor, I. Cholera toxin B subunit modulation of mucosal vaccines for infectious and autoimmune diseases. *Curr. Opin. Investig Drugs* **2010**, *11*, 919–928.
- 73. Meng, Q.; Wang, W.; Shi, X.; Jin, Y.; Zhang, Y. Protection against autoimmune diabetes by silkworm-produced GFP-tagged CTB-insulin fusion protein. *Clin. Dev. Immunol.* **2011**, 2011, 831704. [CrossRef] [PubMed]
- 74. Mbongue, J.C.; Nicholas, D.A.; Zhang, K.; Kim, N.S.; Hamilton, B.N.; Larios, M.; Zhang, G.; Umezawa, K.; Firek, A.F.; Langridge, W.H. Induction of indoleamine 2, 3-dioxygenase in human dendritic cells by a cholera toxin B subunit—Proinsulin vaccine. *PLoS ONE* 2015, 10, e0118562. [CrossRef]
- 75. Stratmann, T. Cholera Toxin Subunit B as Adjuvant–An Accelerator in Protective Immunity and a Break in Autoimmunity. *Vaccines* **2015**, *3*, 579–596. [CrossRef] [PubMed]
- 76. Awate, S.; Babiuk, L.A.; Mutwiri, G. Mechanisms of Action of Adjuvants. Front. Immunol. 2013, 4, 114. [CrossRef] [PubMed]
- Nicholas, D.; Odumosu, O.; Langridge, W.H. Autoantigen based vaccines for type 1 diabetes. *Discov Med.* 2011, *11*, 293–301.
 [PubMed]
- 78. Schuch, R.A.; Oliveira, T.L.; Collares, T.F.; Monte, L.G.; Inda, G.R.; Dellagostin, O.A.; Vendruscolo, C.T.; Moreira, A.D.S.; Hartwig, D.D. The use of xanthan gum as vaccine adjuvant: An evaluation of immunostimulatory potential in balb/c mice and cytotoxicity in vitro. *BioMed Res. Int.* 2017, 2017, 3925024. [CrossRef] [PubMed]
- 79. Arvan, P.; Pietropaolo, M.; Ostrov, D.; Rhodes, C.J. Islet autoantigens: Structure, function, localization, and regulation. *Cold Spring Harb. Perspect. Med.* **2012**, *2*, a007658. [CrossRef]
- Wenzlau, J.M.; Juhl, K.; Yu, L.; Moua, O.; Sarkar, S.A.; Gottlieb, P.; Rewers, M.; Eisenbarth, G.S.; Jensen, J.; Davidson, H.W.; et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc. Natl. Acad. Sci. USA* 2007, 104, 17040–17045. [CrossRef]
- 81. Guan, Y.; Zhang, M.; Li, Y.; Cao, W.; Ji, M.; Liu, Y. Vaccination with IA-2 autoantigen can prevent late prediabetic nonobese diabetic mice from developing diabetes mellitus. *Diabetes Res. Clin. Pract.* **2012**, *95*, 93–97. [CrossRef]
- Liu, X.; Zhang, S.; Li, X.; Zheng, P.; Hu, F.; Zhou, Z. Vaccination with a co-expression DNA plasmid containing GAD65 fragment gene and IL-10 gene induces regulatory CD4+ T cells that prevent experimental autoimmune diabetes. *Diabetes/Metabolism Res. Rev.* 2016, 32, 522–533. [CrossRef]
- 83. Shen, L.; Lu, S.; Huang, D.; Li, G.; Liu, K.; Cao, R.; Zong, L.; Jin, L.; Wu, J. A rationally designed peptide IA-2-P2 against type 1 diabetes in streptozotocin-induced diabetic mice. *Diabetes Vasc. Dis. Res.* 2017, 14, 184–190. [CrossRef] [PubMed]
- 84. Lu, S.; Li, G.; Liu, K.; Yang, X.; Cao, R.; Zong, L.; Long, J.; Jin, L.; Wu, J. Fusion protein His-Hsp65-6IA2P2 prevents type 1 diabetes through nasal immunization in NOD Mice. *Int. Immunopharmacol.* **2016**, *35*, 235–242. [CrossRef] [PubMed]
- 85. Kawasaki, E. ZnT8 and type 1 diabetes. Endocr J. 2012, 59, 531–537. [CrossRef] [PubMed]
- Yi, B.; Huang, G.; Zhou, Z.G. Current and Future Clinical Applications of Zinc Transporter-8 in Type 1 Diabetes Mellitus. *Chin. Med. J.* 2015, 128, 2387–2394. [CrossRef] [PubMed]
- Wu, Q.; Wang, X.; Gu, Y.; Zhang, X.; Qin, Y.; Chen, H.; Xu, X.; Yang, T.; Zhang, M. Screening and identification of human ZnT8-specific single-chain variable fragment (scFv) from type 1 diabetes phage display library. *Sci. China Life Sci.* 2016, *59*, 686–693. [CrossRef] [PubMed]
- 88. Xu, X.; Gu, Y.; Bian, L.; Shi, Y.; Cai, Y.; Chen, Y.; Chen, H.; Qian, L.; Wu, X.; Xu, K.; et al. Characterization of immune response to novel HLA-A2-restricted epitopes from zinc transporter 8 in type 1 diabetes. *Vaccine* **2016**, *34*, 854–862. [CrossRef]
- 89. Hjorth, M.; Axelsson, S.; Ryden, A.; Faresjo, M.; Ludvigsson, J.; Casas, R. GAD-alum treatment induces GAD65-specific CD4+ CD25highFOXP3+ cells in type 1 diabetic patients. *Clin. Immunol.* **2011**, *138*, 117–126. [CrossRef]
- Morales, A.E.; Thrailkill, K.M. GAD-alum immunotherapy in Type 1 diabetes mellitus. *Immunotherapy* 2011, *3*, 323–332. [CrossRef]
 Marrack, P.; McKee, A.S.; Munks, M.W. Towards an understanding of the adjuvant action of aluminium. *Nat. Rev. Immunol.* 2009,
- 9, 287–293. [CrossRef]
 92. Dinarello, C.A. Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood* 2011, 117, 3720–3732. [CrossRef]
- Agardh, C.D.; Lynch, K.F.; Palmer, M.; Link, K.; Lernmark, Å. GAD65 vaccination: 5 years of follow-up in a randomised dose-escalating study in adult-onset autoimmune diabetes. *Diabetologia* 2009, 52, 1363–1368. [CrossRef] [PubMed]
- 94. Ludvigsson, J. The role of immunomodulation therapy in autoimmune diabetes. J. Diabetes Sci. Technol. 2009 3, 320–330. [CrossRef]
- 95. Larsson, H.E.; Lernmark, A. Vaccination against type 1 diabetes. J. Intern. Med. 2011, 269, 626–635. [CrossRef] [PubMed]
- 96. Ludvigsson, J.; Hjorth, M.; Cheramy, M.; Axelsson, S.; Pihl, M.; Forsander, G.; Nilsson, N.O.; Samuelsson, B.O.; Wood, T.; Aman, J.; et al. Extended evaluation of the safety and efficacy of GAD treatment of children and adolescents with recent-onset type 1 diabetes: A randomised controlled trial. *Diabetologia* 2011, 54, 634–640. [CrossRef]
- 97. Harrison, L.C.; Wentworth, J.M.; Zhang, Y.; Bandala-Sanchez, E.; Bohmer, R.M.; Neale, A.M.; Stone, N.L.; Naselli, G.; Bosco, J.J.; Auyeung, P.; et al. Antigen-based vaccination and prevention of type 1 diabetes. *Curr. Diabetes Rep.* 2013, 13, 616–623. [CrossRef] [PubMed]
- 98. Beam, C.A.; MacCallum, C.; Herold, K.C.; Wherrett, D.K.; Palmer, J.; Ludvigsson, J. GAD vaccine reduces insulin loss in recently diagnosed type 1 diabetes: Findings from a Bayesian meta-analysis. *Diabetologia* **2017**, *60*, 43–49. [CrossRef] [PubMed]

- 99. Tavira, B.; Cheramy, M.; Axelsson, S.; Åkerman, L.; Ludvigsson, J.; Casas, R. Effect of simultaneous vaccination with H1N1 and GAD-alum on GAD65-induced immune response. *Diabetologia* **2017**, *60*, 1276–1283. [CrossRef]
- 100. Wherrett, D.K.; Bundy, B.; Becker, D.J.; DiMeglio, L.A.; Gitelman, S.E.; Goland, R.; Gottlieb, P.A.; Greenbaum, C.J.; Herold, K.C.; Marks, J.B.; et al. Antigen-based therapy with glutamic acid decarboxylase (GAD) vaccine in patients with recent-onset type 1 diabetes: A randomised double-blind trial. *Lancet* 2011, 378, 319–327. [CrossRef]
- 101. Ludvigsson, J.; Krisky, D.; Casas, R.; Battelino, T.; Castaño, L.; Greening, J.; Kordonouri, O.; Otonkoski, T.; Pozzilli, P.; Robert, J.J.; et al. GAD65 antigen therapy in recently diagnosed type 1 diabetes mellitus. *N. Engl. J. Med.* **2012**, *366*, 433–442. [CrossRef]
- 102. Ludvigsson, J.; Cheramy, M.; Axelsson, S.; Pihl, M.; Åkerman, L.; Casas, R.; Clinical GAD-Study Group in Sweden. GADtreatment of children and adolescents with recent-onset type 1 diabetes preserves residual insulin secretion after 30 months. *Diabetes/Metabolism Res. Rev.* 2014, 30, 405–414. [CrossRef]
- 103. Cook, D.P.; Gysemans, C.; Mathieu, C. Prospects of a type 1 diabetes vaccine. Expert Opin. Biol. Ther. 2017, 17, 403–406. [CrossRef]
- 104. Denes, B.; Fodor, I.; Langridge, W.H. Autoantigens plus interleukin-10 suppress diabetes autoimmunity. *Diabetes Technol. Ther.* **2010**, *12*, 649–661. [CrossRef]
- 105. Tsui, H.; Chan, Y.; Tang, L.; Winer, S.; Cheung, R.K.; Paltser, G.; Selvanantham, T.; Elford, A.R.; Ellis, J.R.; Becker, D.J.; et al. Targeting of pancreatic glia in type 1 diabetes. *Diabetes* **2008**, *57*, 918–928. [CrossRef] [PubMed]
- 106. Pang, Z.; Kushiyama, A.; Sun, J.; Kikuchi, T.; Yamazaki, H.; Iwamoto, Y.; Koriyama, H.; Yoshida, S.; Shimamura, M.; Higuchi, M.; et al. Glial fibrillary acidic protein (GFAP) is a novel biomarker for the prediction of autoimmune diabetes. *FASEB J.* 2017, 31, 4053–4063. [CrossRef]
- 107. Pang, Z.; Higuchi, M.; Koriyama, H.; Yoshida, S.; Kurinami, H.; Shimamura, M.; Takami, Y.; Rakugi, H.; Morishita, R.; Nakagami, H. Evaluating the potential of the GFAP-KLH immune-tolerizing vaccine for type 1 diabetes in mice. *FEBS Lett.* 2017, 591, 129–136. [CrossRef] [PubMed]
- 108. Lewis, J.S.; Dolgova, N.V.; Zhang, Y.; Xia, C.Q.; Wasserfall, C.H.; Atkinson, M.A.; Clare-Salzler, M.J.; Keselowsky, B.G. A combination dual-sized microparticle system modulates dendritic cells and prevents type 1 diabetes in prediabetic NOD mice. *Clin. Immunol.* 2015, 160, 90–102. [CrossRef] [PubMed]
- 109. Yoon, Y.M.; Lewis, J.S.; Carstens, M.R.; Campbell-Thompson, M.; Wasserfall, C.H.; Atkinson, M.A.; Keselowsky, B.G. A combination hydrogel microparticle-based vaccine prevents type 1 diabetes in non-obese diabetic mice. *Sci. Rep.* 2015, *5*, 1–13. [CrossRef] [PubMed]
- 110. Phillips, B.; Nylander, K.; Harnaha, J.; Machen, J.; Lakomy, R.; Styche, A.; Gillis, K.; Brown, L.; Lafreniere, D.; Gallo, M.; et al. A microsphere-based vaccine prevents and reverses new-onset autoimmune diabetes. *Diabetes* 2008, 57, 1544–1555. [CrossRef] [PubMed]
- 111. Keselowsky, B.G.; Xia, C.Q.; Clare-Salzler, M. Multifunctional dendritic cell-targeting polymeric microparticles: Engineering new vaccines for type 1 diabetes. *Hum. Vaccines* **2011**, *7*, 37–44. [CrossRef]
- 112. Miller, S.D.; Turley, D.M.; Podojil, J.R. Antigen-specific tolerance strategies for the prevention and treatment of autoimmune disease. *Nat. Rev. Immunol.* 2007, 7, 665–677. [CrossRef]
- 113. Roep, B.O.; Solvason, N.; Gottlieb, P.A.; Abreu, J.R.; Harrison, L.C.; Eisenbarth, G.S.; Yu, L.; Leviten, M.; Hagopian, W.A.; Buse, J.B.; et al. Plasmid-encoded proinsulin preserves C-peptide while specifically reducing proinsulin-specific CD8+ T cells in type 1 diabetes. *Sci. Transl. Med.* 2013, *5*, 191ra82. [CrossRef]
- 114. Ludvigsson, J.; Sumnik, Z.; Pelikanova, T.; Nattero Chavez, L.; Lundberg, E.; Rica, I.; Martínez-Brocca, M.A.; Ruiz de Adana, M.; Wahlberg, J.; Katsarou, A.; et al. Intralymphatic glutamic acid decarboxylase with vitamin D supplementation in recent-onset type 1 diabetes: A double-blind, randomized, placebo-controlled phase IIb trial. *Diabetes Care* 2021, 44, 1604–1612. [CrossRef] [PubMed]
- 115. Geenen, V.; Mottet, M.; Dardenne, O.; Kermani, H.; Martens, H.; Francois, J.M.; Galleni, M.; Hober, D.; Rahmouni, S.; Moutschen, M. Thymic self-antigens for the design of a negative/tolerogenic self-vaccination against type 1 diabetes. *Curr. Opin. Pharmacol.* 2010, 10, 461–472. [CrossRef] [PubMed]
- 116. Harrison, L.C. Vaccination against self to prevent autoimmune disease: The type 1 diabetes model. *Immunol. Cell Biol.* **2008**, *86*, 139–145. [CrossRef] [PubMed]
- 117. Lewis, G.K. Live-attenuated Salmonella as a prototype vaccine vector for passenger immunogens in humans: Are we there yet? *Expert Rev. Vaccines* 2007, *6*, 431–440. [CrossRef] [PubMed]
- 118. Petrovsky, N. Immunomodulation with microbial vaccines to prevent type 1 diabetes mellitus. *Nat. Rev. Endocrinol.* **2010**, *6*, 131–138. [CrossRef]
- 119. Husseiny, M.I.; Rawson, J.; Kaye, A.; Nair, I.; Todorov, I.; Hensel, M.; Kandeel, F.; Ferreri, K. An oral vaccine for type 1 diabetes based on live attenuated Salmonella. *Vaccine* **2014**, *32*, 2300–2307. [CrossRef]
- 120. Cheminay, C.; Hensel, M. Rational design of Salmonella recombinant vaccines. Int. J. Med. Microbiol. 2008, 298, 87–98. [CrossRef]
- 121. Oikawa, Y.; Shimada, A.; Yamada, Y.; Okubo, Y.; Katsuki, T.; Shigihara, T.; Miyazaki, J.I.; Narumi, S.; Itoh, H. CXC chemokine ligand 10 DNA vaccination plus Complete Freund's Adjuvant reverses hyperglycemia in non-obese diabetic mice. *Rev. Diabet. Stud. RDS* **2010**, *7*, 209.
- 122. Denes, B.; Fodor, I.; Langridge, W.H. Persistent suppression of type 1 diabetes by a multicomponent vaccine containing a cholera toxin B subunit-autoantigen fusion protein and complete Freund's adjuvant. *Clin. Dev. Immunol.* **2013**, 2013, 578786. [CrossRef]

- 123. Lin, M.S.; Tse, H.M.; Delmastro, M.M.; Bertera, S.; Wong, C.T.; Lakomy, R.; He, J.; Sklavos, M.M.; Coudriet, G.M.; Pietropaolo, M.; et al. A multivalent vaccine for type 1 diabetes skews T cell subsets to Th2 phenotype in NOD mice. *Immunol. Res.* 2011, 50, 213–220. [CrossRef] [PubMed]
- 124. Lilani, Z.; Ahmed, A.; Tazeem, R.; Naeem, E. BCG vaccine—An innovation for diabetes treatment in Pakistan? J. Ayub Med. Coll. Abbottabad 2017, 29, 1–2. [PubMed]
- 125. Faustman, D.L.; Wang, L.; Okubo, Y.; Burger, D.; Ban, L.; Man, G.; Zheng, H.; Schoenfeld, D.; Pompei, R.; Avruch, J.; et al. Proof-of-concept, randomized, controlled clinical trial of Bacillus-Calmette-Guerin for treatment of long-term type 1 diabetes. *PLoS ONE* **2012**, *7*, e41756. [CrossRef]
- 126. Karaci, M. The protective effect of the BCG vaccine on the development of type 1 diabetes in humans. In *The Value of BCG and TNF in Autoimmunity;* Academic Press: Cambridge, MA, USA, 2014; pp. 52–62.
- 127. Kowalewicz-Kulbat, M.; Locht, C. BCG and protection against inflammatory and auto-immune diseases. *Expert Rev. Vaccines* **2017**, *16*, 699–708. [CrossRef] [PubMed]
- 128. Bluestone, J.A.; Herold, K.; Eisenbarth, G. Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature* 2010, 464, 1293–1300. [CrossRef]
- 129. Nurminen, N.; Oikarinen, S.; Hyoty, H. Virus infections as potential targets of preventive treatments for type 1 diabetes. *Rev. Diabet. Stud. RDS* **2012**, *9*, 260. [CrossRef]
- 130. Hyoty, H.; Knip, M. Developing a vaccine for type 1 diabetes through targeting enteroviral infections. *Expert Rev. Vaccines* **2014**, *13*, 989–999. [CrossRef]
- 131. Larsson, P.G.; Lakshmikanth, T.; Laitinen, O.H.; Utorova, R.; Jacobson, S.; Oikarinen, M.; Domsgen, E.; Koivunen, M.R.; Chaux, P.; Devard, N.; et al. A preclinical study on the efficacy and safety of a new vaccine against Coxsackievirus B1 reveals no risk for accelerated diabetes development in mouse models. *Diabetologia* 2015, *58*, 346–354. [CrossRef]
- Rodrigues dos Santos, R., Jr.; Sartori, A.; Deperon Bonato, V.L.; Coelho Castelo, A.A.M.; Vilella, C.A.; Zollner, R.L.; Lopes Silva, C. Immune modulation induced by tuberculosis DNA vaccine protects non-obese diabetic mice from diabetes progression. *Clin. Exp. Immunol.* 2007, 149, 570–578. [CrossRef]
- 133. Stratton, K.R.; Howe, C.J.; Johnston, R.B. Adverse Effects of Vaccines: Evidence and Causality; National Academies Press (US) National Academy of Sciences Press: Washington, DC, USA, 2011.
- 134. Boettler, T.; Pagni, P.P.; Jaffe, R.; Cheng, Y.; Zerhouni, P.; von Herrath, M. The clinical and immunological significance of GAD-specific autoantibody and T-cell responses in type 1 diabetes. *J. Autoimmun.* **2013**, *44*, 40–48. [CrossRef]





Article **Toxicity of** *Bacillus thuringiensis* **Strains Derived from the Novel Crystal Protein Cry31Aa with High Nematicidal Activity against Rice Parasitic Nematode** *Aphelenchoides besseyi*

Zhao Liang ^{1,2,†}, Qurban Ali ^{1,2,†}, Yujie Wang ^{1,2}, Guangyuan Mu ³, Xuefei Kan ³, Yajun Ren ^{1,2}, Hakim Manghwar ⁴, Qin Gu ^{1,2}, Huijun Wu ^{1,2} and Xuewen Gao ^{1,2,*}

- ¹ The Sanya Institute of Nanjing Agricultural University, Sanya 572024, China; zhaol0619@163.com (Z.L.); 2020202068@stu.njau.edu.cn (Q.A.); yjiewang57@163.com (Y.W.); renyajun0719@163.com (Y.R.); guqin@njau.edu.cn (Q.G.); hjwu@njau.edu.cn (H.W.)
- ² Key Laboratory of Integrated Management of Crop Diseases and Pests, Ministry of Education, Department of Plant Pathology, College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China
- ³ Shenzhen Batian Ecotypic Engineering Co., Ltd., Shenzhen 518057, China; muguangyuan1@sina.com (G.M.); xavier2014@163.com (X.K.)
- ⁴ Lushan Botanical Garden, Chinese Academy of Sciences, Jiujiang 332000, China; hakim@lsbg.cn
- Correspondence: gaoxw@njau.edu.cn; Tel.: +86-025-8439-5268
- + These authors contributed equally to this study.

Abstract: The plant parasitic nematode, Aphelenchoides besseyi, is a serious pest causing severe damage to various crop plants and vegetables. The Bacillus thuringiensis (Bt) strains, GBAC46 and NMTD81, and the biological strain, FZB42, showed higher nematicidal activity against A. besseyi, by up to 88.80, 82.65, and 75.87%, respectively, in a 96-well plate experiment. We screened the whole genomes of the selected strains by protein-nucleic acid alignment. It was found that the Bt strain GBAC46 showed three novel crystal proteins, namely, Cry31Aa, Cry73Aa, and Cry40ORF, which likely provide for the safe control of nematodes. The Cry31Aa protein was composed of 802 amino acids with a molecular weight of 90.257 kDa and contained a conserved delta-endotoxin insecticidal domain. The Cry31Aa exhibited significant nematicidal activity against A. besseyi with a lethal concentration (LC50) value of 131.80 μ g/mL. Furthermore, the results of in vitro experiments (i.e., rhodamine and propidium iodide (PI) experiments) revealed that the Cry31Aa protein was taken up by A. besseyi, which caused damage to the nematode's intestinal cell membrane, indicating that the Cry31Aa produced a poreformation toxin. In pot experiments, the selected strains GBAC46, NMTD81, and FZB42 significantly reduced the lesions on leaves by up to 33.56%, 45.66, and 30.34% and also enhanced physiological growth parameters such as root length (65.10, 50.65, and 55.60%), shoot length (68.10, 55.60, and 59.45%), and plant fresh weight (60.71, 56.45, and 55.65%), respectively. The number of nematodes obtained from the plants treated with the selected strains (i.e., GBAC46, NMTD81, and FZB42) and A. besseyi was significantly reduced, with 0.56, 0.83., 1.11, and 5.04 seedling mL^{-1} nematodes were achieved, respectively. Moreover, the qRT-PCR analysis showed that the defense-related genes were upregulated, and the activity of hydrogen peroxide (H₂O₂) increased while malondialdehyde (MDA) decreased in rice leaves compared to the control. Therefore, it was concluded that the Bt strains GBAC46 and NMTD81 can promote rice growth, induce high expression of rice defenserelated genes, and activate systemic resistance in rice. More importantly, the application of the novel Cry31Aa protein has high potential for the efficient and safe prevention and green control of plant parasitic nematodes.

Keywords: Bacillus thuringiensis; Cry toxin; nematicidal activity; pore-formation

Citation: Liang, Z.; Ali, Q.; Wang, Y.; Mu, G.; Kan, X.; Ren, Y.; Manghwar, H.; Gu, Q.; Wu, H.; Gao, X. Toxicity of *Bacillus thuringiensis* Strains Derived from the Novel Crystal Protein Cry31Aa with High Nematicidal Activity against Rice Parasitic Nematode *Aphelenchoides besseyi. Int. J. Mol. Sci.* **2022**, 23, 8189. https:// doi.org/10.3390/ijms23158189

Academic Editor: Wajid Zaman

Received: 20 June 2022 Accepted: 19 July 2022 Published: 25 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

Plant pathogens, such as fungi, bacteria, viruses, and nematodes, cause different plant diseases, resulting in yield loss in crops around the globe [1,2]. Plant parasitic nematodes are currently serious pests, contributing to USD 157 billion in annual agricultural losses worldwide [3]. *Aphelenchoides besseyi* is one of the most damaging plant parasitic nematodes and can cause severe harm to a variety of significant crops and vegetables. The significant rice seed-borne pest, *A. besseyi*, spreads predominantly through infected seeds. This is the cause of rice white tip disease and is widely distributed throughout rice-growing areas [4]. In general, *A. besseyi* causes yield losses ranging from 10% to 30%, and in strongly infested areas, yield losses might reach 50% [5,6].

Rhizobacteria have been widely reported as biological control agents for plant parasitic nematodes. Numerous *Bacillus* species have been studied for their role in reducing nematodes and insect pests [7,8]. *Bacillus subtilis* and *Bacillus thuringiensis* (Bt) are two of the *Bacillus* species that have been investigated against various pathogens, including nematodes [9,10]. Bt strains can produce a number of parasporal crystal proteins (Cry or d-endotoxins), which are poisonous to a variety of insects and pests [11,12]. Bt strains can also produce various crystal protein toxins (Cry5, Cry6, Cry12, Cry13, Cry14, Cry21, and Cry55) during the growth process. These Cry toxins exhibit a substantial biological activity against various insects, nematodes, and other pathogenic pests [13,14].

Currently, the high efficiency and specificity of these toxic proteins enable Bt strains to be safely used in controlling various insect pests of crops and are considered to be green and safe microbial pesticides—making them one of the most successful microbial pesticides [15,16]. These are widely used in global agricultural production, effectively reducing the number of chemical pesticides [17]. In addition, Bt toxin proteins have also been used in the field of plant genetic engineering to control crop pests [18]. For example, transgenic rapeseed containing the Cry1Ac gene can control diamondback moth, hairy bug, and cotton bollworm [19], and transgenic pigeon bean containing the Cry2Aa gene is resistant to pod borers [20]. The Cry3A gene in transgenic spruce showed toxicity to spruce bark beetles, indicating that Bt and its toxin proteins add significant value to biological control [21].

In addition to insects, plant parasitic nematodes also affect the agricultural economy [22,23]. Many studies have demonstrated that Cry proteins produced by Bt strains have nematicidal activity [16,23]. Cry proteins with nematicidal activity mainly include Cry5, Cry6, Cry14, Cry21, and Cry55 [14,24]; among these, the 3D-Cry protein Cry5 and non-3D-Cry protein Cry6 have been reported to kill nematodes [25,26]. In order to find and identify novel Cry toxins with broad-spectrum activity against insects and nematodes, we compared Bt strains GBAC46 and NMTD81 based on genome-wide screening and found three unknown Cry proteins in the GBAC46 strain. Among the identified Cry proteins, Cry31Aa showed the highest nematicidal activity in killing nematodes in the intestinal cell membrane.

Furthermore, the beneficial properties of Bt strains are not limited to their biological control activity; however, a variety of endophytes possess the ability to indirectly or directly promote plant growth and development [27,28]. Bt is one of the most important phosphatedissolving bacteria [29], which can be insoluble in the soil through enzymatic activity. The phosphate was hydrolyzed into plant-absorbable phosphorus ions [30]. Bt produces siderophores through a nonribosomal synthetic peptide pathway to help plants absorb iron [31,32]. Additionally, it can synthesize plant hormones (such as indole acetic acid (IAA), deaminase, and volatile organic compounds (VOCs) to promote plant growth [33]. Thus, Bt strains can be used as a biofertilizer agent to promote the uptake and transport of plant mineral nutrients.

Therefore, in the current study, we used different *Bacillus thurgensis* (Bt) strains, GBAC46 and NMTD81, which were previously isolated in the Qinghai–Tibetan Plateau, China, and a well-known biological strain, *B. velezensis* FZB42, against the plant parasitic nematode *A. besseyi*. In this study, we identified a novel plasmid-encoded Cry protoxin

Cry31Aa produced by Bt GBAC46, which acts synergistically to kill free-living *A. besseyi* nematodes via intestinal damage. We also determined the lethal concentrations at LC_{50} (131.80 µg/mL) for the selected Cry proteins Cry31Aa, Cry73Aa, and Cry40OR and found that the only Cry31Aa had strong nematicidal toxicity against *A. besseyi*. The present work sheds new light on the mechanism of Bt strains and their Cry31Aa proteins that regulate the genes involved in the defense mechanism that *A. besseyi*-infested rice plants use to efficiently control white tip disease in rice. Thus, the toxins produced from the Bt strain could be employed for effective *A. besseyi* control, benefiting both the rice crop and the environment, offering a more sustainable management technique than the overuse of nonspecific chemical nematicides.

2. Results

2.1. Sequence Analysis of Cry31Aa and Familiar Cry Protein

Bt is a main pathogenic bacterium, the pathogenicity of which is largely dependent on the parasporal crystal protein (Cry) which it produces during biological control activity [13]. In order to find new Cry toxins, we used SeqHunter 2 software to screen the whole genome of the GBAC46 and NMTD81 strains, and the results predicted 3 unknown Cry protein genes in GBAC46, namely, Cry31Aa, Cry73Aa, and Cry40ORF, while in the NMTD81 strain, there were no new Cry proteins observed (Figure 1A). The Cry31Aa, Cry73Aa, and Cry40ORF proteins contained 802, 395, and 568 amino acids and were 90.257, 44.647, and 65.1902 kDa in size, respectively. The Cry31Aa and Cry73Aa, which were 110–338Aa and 24–227Aa, respectively, at the N-terminal, were in the delta-endotoxin insecticidal domain. Furthermore, the Cry73Aa C-terminal amino acids 524-659 were at the carbohydrate-binding module 6 (CBM6) functional domain, which determines the nematicidal specificity by binding with carbohydrates. In addition, the analysis of the amino acid sequences of three Cry proteins in the GBAC46 strain found that the Cry31Aa protein was 65.48%, similar to the Cry70Ba1 protein (Figure 1B). The Cry40ORF protein had more similarity to the Cry48Ab1 protein, but the similarity of the amino acid sequence was 14.26%. The percentage of amino acid sequence between Cry73Aa and MPP46Aa1 protein was the highest, i.e., 11.62%. These three Cry proteins of the GBAC46 strain are novel and unreported Cry proteins. Furthermore, the SWISS-MODEL was used to predict and analyze the three-dimensional structures of the three Cry proteins of the GBAC46 strain. The result showed that the 3D model from the crystal structure of three Cry proteins was rich in α -helix and β -sheet and predict that the respective proteins may be a perforating toxins (Figure 1C).

2.2. Expression, Immunoblot, and Purification of the Three Cry Proteins of GBAC46

The RT-PCR analysis was used for the detection of the three Cry proteins from Bt strain GBAC46. The RT-PCR results showed that the three Cry proteins of the GBAC46 strain were normally transcribed when reverse-transcribed cDNA was used as a template and strong bands were detected in the gel. When RNA was employed as a negative template, no bands were obtained, whereas the size was in agreement with the calculated size of the three proteins, Cry31Aa, Cry40ORF, and Cry73Aa, which was 100, 150, and 150 bp nucleotides, respectively (Figure 2A). In addition, the three Cry proteins of the GBAC46 strain were heterologously expressed in BL21 *E. coli* using the pOPT *E. coli* expression vector. The size of the three proteins, Cry31Aa, Cry31Aa, Cry73Aa, and Cry40ORF, was 104.39, 58.78, and 79.32 kDa, respectively. The results of the Western blot and SDS-PGAE showed that the three Cry protein target bands were correct as shown in Figure 2B.



Figure 1. Phylogenetic tree construction was based on the sequence alignment of full-length Cry31Aa and familiar Cry protein (**A**); the amino acid sequences alignment of the Cry31Aa, Cry73Aa, and Cry40ORF proteins (**B**); three-dimensional structure prediction of the three Cry proteins in the GBAC46 strain (**C**).



Figure 2. Detection of Cry protein genes and the identification of BL21 *E. coli* using the pOPT *E. coli* expression vector. (**A**) Detection of the transcription of *Cry31Aa* through RT-PCR. Total RNA was extracted from GBAC46 and BL21/pOPT. cDNA was used as a positive control, and RNA was used as a negative control. (**B**) SDS-PAGE expression analysis of Cry proteins (i.e., Cry31Aa, Cry40ORF, and Cry73Aa) after IPTG induction at 16 °C for 16 h. Levels of the three Cry proteins encoded by BL21/pOPT detected via Western blotting. The experiments were repeated independently three times.

2.3. Virulence of Bacillus Thuringiensis Strains to A. besseyi

The nematicidal activity of the selected Bt strains against *A. besseyi* in vitro was evaluated through a 96-well plate. The plate counting method was used to detect the concentrations of strains being treated with nematodes at different gradients. After 24 h of incubation at 20 °C, the GBAC46, NMTD81, and FZB42 strains showed strong changes in their ability to kill *A. besseyi*, up to 88.80, 82.65, and 75.87%, respectively, compared to the control as shown in Figure 3A. If the nematodes were unable to restore movement after being touched with a needle under a microscope, they were considered dead. In addition, various concentrations in CFU/mL of the selected strains were used. The results showed that with a high concentration in CFU/mL, the mortality rate of *A. besseyi* was 100% compared to the control, ddH₂O, as shown in Figure 3B.



Figure 3. Nematicidal activity of selected *Bacillus* strains against *A. besseyi* compared with the control: (**A**) microscopic figures under different objectives: $10 \times$, $40 \times$, and $100 \times$; (**B**) graphical details of the nematicidal activity of selected *Bacillus* strains. Different error bars show the different mean standard deviation of each treatment repeated three times with triplicates.

2.4. Morphological Observation and Pore-Formation of Bt Strains Via Staining

Bacillus is a typical rod-shaped bacterium. In the late stage of the stable culture period and the dying period, *Bacillus* forms spores in the cell to adapt to the external environment. We stained GBAC46 and NMTD81 strains cultured in SM medium for 48 h with biological staining reagents. The results showed that the vegetative morphology of the GBAC46 and NMTD81 strains were a typical short rod shape, while the spores were nearly elliptical (Figure 4A). Furthermore, a scanning electron microscope (SEM) was used to observe the spore and parasporal crystal protein of GBAC46 and NMTD81 after being cultured in an ICPM medium for 80 h. The outcomes showed that both selected strains formed more spores and Cry as shown in Figure 4B.



Figure 4. Detection of spore formation in GBAC46 and NMTD81 strains: (**A**) different colors represent different cells: blue—spores, red—living cells under the microscopic study; (**B**) observation of spore and crystal proteins in selected strains under a SEM. White arrow—spores; black arrow—circular crystal proteins; yellow arrow—bipyramidal crystal proteins; red arrow—irregular crystal proteins.

2.5. Toxicity of Cry Protein to A. besseyi

In this study, we tested the selected Cry proteins, Cry31Aa, Cry40ORF, and Cry73Aa, from GBAC46 to determine the nematicidal activity against *A. besseyi*. The results showed that the Cry31Aa protein had the highest nematicidal activity against *A. besseyi*, and its LC_{50} value was 131.80 µg/mL. In addition, we used rhodamine to label three Cry proteins and detect whether the protein was ingested by the nematodes, and PBS was used as the negative control. The results showed that red fluorescence could be observed in the gut of the nematodes after 4 h of treatment with rhodamine-labeled Cry31Aa protein, whereas no red fluorescence was observed in the negative control treated with PBS (Figure 5A). To verify that the Cry31Aa protein was a perforating toxin, we fed *A. besseyi* with the protein Cry31Aa and stained the nematodes with propidium iodide dye. The results showed that the Cry31Aa protein was treated, and the diffusion degree of propidium iodide was present in the intestine of *A. besseyi* (Figure 5B).





Figure 5. Observation of Cry proteins absorbed by *A. besseyi*: (**A**) the nematode *A. besseyi* after being treated with PBS as a control and rhodamine-labeled Cry31Aa were observed in bright and red fields; (**B**) the Cry31a protein produced the pore-formation toxin exposed to *A. besseyi* after staining with propidium iodide (PI), and fluorescence microscopy was used to monitor the signal of PI. The bright and red fields clearly indicate the changes in the intestinal cells. Different letters indicate significant differences at p < 0.005.

2.6. Plant Growth Promotion Traits of Bt Strains in a Pot Experiment

In the present study, the selected strains were used as a treatment, and control water was used as a CK. The rice seeds were soaked in a solution of *A. besseyi* for 36 h, and then seeds were dipped into selected bacterial solutions of $OD_{600} = 1.0$ and planted for 1 month in a greenhouse under controlled conditions. The result showed that the disease intensity was small, and the "white tip" lesions were white to yellow–brown in the control seedlings. Whereas, after being treated with the bacterial solution of selected strains, the disease symptoms on leaves were significantly reduced (Figure 6A). The length of the leaf lesions was calculated, and the disease index of different treatment groups was found to be significantly reduced up to 33.56% 45.66, and 30.34 by GBAC46, NMTD81, and FZB42, respectively.

Further, the growth of rice seedlings along with other physiological parameters were determined to be enhanced by *Bacillus* strains after treating with *A. besseyi*. It was observed that the selected strains, GBAC46, NMTD81, and FZB42, significantly increased root length by 65.10, 50.65, and 55.60%, shoot length by 68.10, 55.60, and 59.45%, and plant fresh weight by 60.71, 56.45, and 55.65%, respectively, compared to the control water treatment (Figure 6b). In addition, the whole plant of rice was cut into small pieces, the nematodes were isolated from the plants, and the concentration of nematodes isolated from the rice seedlings was calculated. The results showed that the number of nematodes

obtained from rice plants treated with the selected strains, GBAC46, NMTD81, and FZB42, and the nematode *A. besseyi* significantly decreased by 0.56, 0.83, 1.11, and 5.04 mL⁻¹ seedling⁻¹, respectively. In conclusion, these results indicated that the selected Bt strains could significantly overcome the rice white tip nematode disease.



Figure 6. The effect of various *Bacillus* strains against *A. besseyi*-infested rice plants in a greenhouse experiment. The photographic representation of the effect of the selected strains on the infested rice plants (**A**). The rice growth promotion parameters, such as root and shoot length, plant fresh weight, and the detection of several *A. besseyi* nematodes from rice seedlings, after being treated with the selected strains (**B**). The error bars indicate the mean standard deviation of each treatment when replicated three times. Significant differences are indicated by the letters above the columns. Fisher's LSD test was used to determine significant differences between treatments at $p \leq 0.05$.

2.7. Hydrogen Peroxide (H₂O₂) and Malondialdehyde (MDA) Analysis in Rice

To further examine the effects of GBAC46 and NMTD81 strains on the defense response of rice, we detected the contents of hydrogen peroxide (H_2O_2) and malondialdehyde (MDA), a product of lipid peroxidation in rice leaves after being treated with the selected strains for 24 h. The results showed that both the GBAC46 and NMTD81 strains could significantly induce the accumulation of H_2O_2 and reduced the MDA in rice leaves, revealing that the selected strains could induce the defense response of rice compared to the FZB42-positive and CK-negative controls (Figure 7).



Figure 7. Determination of H_2O_2 and MDA in the leaves of 30 day old seedlings following different treatments. The error bars indicate the mean standard deviation of each treatment when replicated three times. Significant differences are indicated by the letters above the columns. Fisher's LSD test was used to determine significant differences between treatments at $p \le 0.05$.

2.8. Relative Expression of Defense-Related Genes

The effects of the selected strains on the relative expression of rice plant defenserelated (*PR*) genes were studied. The results revealed that the GBAC46 strain significantly enhanced the relative expression levels of all defense-related genes, such as *PBZ1*, *PR1a*, *PR4*, and *LOX1*, whereas the *PR1a* genes were highly upregulated compared to other genes. Similar results were obtained after treatment with the NMTD81 strain, and the relative expression levels of *PR*-related genes, *PBZ1*, *PR1a*, and *PR4*, were significantly upregulated, indicating that the selected strains could induce high relative expression of rice defense response-related genes (Figure 8).



Figure 8. Relative expression analysis of defense genes *PR1a*, *PR4*, *LOX1*, *PBZ1*, and *PAL1* in rice plants infested with *A. besseyi* and treated with Bt strains. The error bars indicate the mean standard deviation of each treatment when replicated three times. Significant differences are indicated by the letters above the columns. Fisher's LSD test was used to determine significant differences between treatments at $p \leq 0.05$.

3. Discussion

Plant parasitic nematodes threaten agricultural production and human health [34]. *Bacillus thuringiensis* is a pathogenic bacterium with a pathogenicity that is largely dependent on the parasporal crystal protein which it produces against nematodes [13]. For the discovery and identification of Bt toxins, whole-genome sequencing has proven to be a helpful and efficient method [35,36]. Cry proteins, Cry21Ha, Cry1Ba, Vip1/Vip2, and -exotoxin, were identified as possible nematicidal factors by genome analysis of Bt 4A4, suggesting that this strain possesses many toxins with potential activity against nematodes method [35]. The discovery of new highly toxic Cry toxins will help in developing new pesticides to effectively control diseases related to different insects and nematodes in agriculture.

In this study, we used SeqHunter 2 software to screen the whole genome of the GBAC46 strain, where three unknown Cry protein genes were found, namely, Cry31Aa, Cry73Aa, and Cry40ORF (Figure 1a). The analysis of the amino acid sequences of the three Cry proteins in the GBAC46 strain revealed that the Cry31Aa protein was 65.48% similar to the Cry70Ba1 protein [37]. Recent studies reported that the Cry70Ba1 and Cry21Ha1 proteins had good insecticidal activity against the Lepidoptera beet armyworm and *C. elegans* [38]. The Cry40ORF protein had more similarity to the Cry48Ab1 protein, but the similarity of the amino acid sequence was 14.26%, within delta-endotoxin Cry1Ac-Domain V domain [39]. Furthermore, the SWISS-MODEL was used to predict and analyze the three-dimensional structures of the three Cry proteins of the GBAC46 strain. The results show that the 3D model from the crystal structure of the three Cry proteins was rich in α -helix and β -sheet and predicted that the respective proteins may be a perforating toxin, in agreement with a previous report [40].

Various Bt strains can effectively control different plant parasitic and animal nematodes [11,41]. Bt has been discovered to have a key function in the control of plant nematodes; it not only prevents eggs from hatching but also has nematicidal activity against juveniles (J2) [42]. Various scientists have reported the role of several Bt species in killing nematodes such as *Meloidogyne incognita, Rotylenchulus reniformis, Heterodera gylcines,* and *Caenorhabditis elegans* [11,36]. In this study, we explored the nematicidal activity of the Bt strains against *A. besseyi* in a 96-well plate in vitro. The selected strains GBAC46, NMTD81, and FZB42 revealed significant differences in their ability to kill *A. besseyi* after 24 h of incubation at 20 °C compared to the control. These findings are consistent with prior research that found *P. pacificus* to be more disease resistant than *C. elegans* [43,44]. Another study also found that the secondary metabolite of the FZB42 strain, plantazolicin, kills *C. elegans* and has insecticidal activity against *D. elegans* [45].

Many studies have shown that Cry proteins are highly toxic to nematodes [46,47]. Bt Cry toxins are almost entirely encoded by plasmids and can be used as nematicidal candidates in Bt DB27 [36]. Interestingly, previous crystals revealed a substantial relationship with spores, a trait known as spore–crystal interaction (SCA) [48]. To explore further, we tested whether Bt GBAC46 produced Cry toxins. After culturing for 80 h in an ICPM medium, the results demonstrated that the GBAC46 strain produced more spores, and the parasporal crystal proteins were detected using SEM. These results reveal strong similarities to the SCA of a rare filamentous Bt strain [49].

Cry5, Cry6, Cry14, Cry21, and Cry55 are among the cry proteins that have reported nematicidal activity [24]. One of the prerequisites for a Cry protein to induce its toxic activity is that the toxin protein can be ingested into the intestine by nematodes. It was found that plant parasitic nematodes can only ingest food through narrow needles, limiting the size of their food [50]. It has also been shown that the ability of nematodes to take up proteins is not only determined by the protein molecular weight but may also be affected by the overall size, shape, and electrostatic charge of the protein [51]. Here, we investigated the Cry31Aa proteins from GBAC46 strains. The Cry31Aa proteins were found to show a strong nematicidal activity against *A. besseyi*. In addition, (6)-rhodamine was employed to label three cells, since carboxytetramethylrhodamine is a biological stain with strong fluorescence

when dissolved in DMSO solvent and is commonly used for biological dyes [52]. After 4 h of treatment with the rhodamine-labeled Cry31Aa protein, red fluorescence was visible in the guts of the nematodes, which agrees with the results in [53]. Propidium iodide (PI) is a red fluorescent dye that can enter nematode cells through the holes formed by the Cry protein, staining the nucleus red while cells are intact; it is frequently used to determine whether a Cry protein is a perforating toxin protein [9]. The red fluorescence could be seen in the nuclei of the nematode intestinal cells if the Cry protein was a perforating toxin.

The biocontrol potential of the Bt selected isolates were evaluated for their effective control of the plant parasitic nematode *A. besseyi* under greenhouse conditions. The results showed that the disease severity was small and white tip lesions were white to yellow–brown in the control seedling. Whereas after being treated with a bacterial solution of the selected strains with an $OD_{600} = 1.0$, the disease symptoms on leaves significantly reduced. A similar result was found against root-knot nematode caused by *M. incognita* [54]. In addition, the growth of rice seedlings and the other physiological parameters were also determined. The results revealed that the root length and plant fresh weight significantly increased compared to the control water treatment, which is in agreement with [53].

Furthermore, *Bacillus* spp. and its products have been shown to stimulate the plant immune response and induce systemic resistance (ISR) and upregulate the pathogenicity-related genes in plants [55,56]. The rapid buildup of H_2O_2 is considered to be a significant phytopathogen defense signal in cellular plants [57]. The accumulation of malondialde-hyde (MDA) levels are important indicators in plants to combat various environmental stresses [10,58] whereas the decreased level of MDA indicated less membrane damage in plants inoculated with *Bacillus* strains under stress conditions [59]. In the current study, it was also found that the H_2O_2 levels were upregulated and the MDA levels were downregulated in a very efficient manner in rice plants treated with selected strains.

Bt can also directly promote plant growth by synthesizing phosphate lysing enzymes [60], siderophores [61], and plant hormones [62,63]. When analyzing the growthpromoting effects of GBAC46 and NMTD81 strains in rice, it was found that the selected strains significantly promoted rice physiological parameters such as root and shoot length [10,28]. Moreover, qRT-PCR experiments showed that the GBAC46 and NMTD81 strains were able to enhance the relative expression of rice plant defense-related (PR) genes. The experimental results showed that the selected strains significantly enhanced the relative expression of the defense related genes in rice plants; our results agree with those in [10,64].

4. Materials and Methods

4.1. Plasmids, Bacterial Strains, and Cultural Conditions

In the current study, we used *Bacillus thuringiensis* GBAC46 and NMTD81 strains, which were previously isolated at the Laboratory of Biocontrol and Bacterial Molecular Biology, Nanjing Agriculture University, Nanjing, China, along with the biological control strain *Bacillus velezensis* FZB42, which was used as a positive control, and double-distilled water (ddH₂O), which was used as a negative control (Supplementary Materials Table S1) [65]. All selected strains were incubated in 1/2 Luria-Bertani (LB) liquid medium (0.5% tryptone, 0.25% yeast extract, and 0.5% NaCl, pH 7.0) or agar plates at 30 °C. The *Escherichia coli* (*E. coli*) strains DH5 α and BL21 (DE3) were cultivated at 37 °C in LB agar plates or broth. The required antibiotic, ampicillin (Amp), was added at a final concertation of 100 µg/mL with shaking for 24 h at 200 rpm, according to [66], with some modifications.

4.2. Nematode Inoculum

The original culture of the plant parasitic nematode, *Aphelenchoides besseyi*, was provided by Xuan Wang (Laboratory of Plant Nematology, Nanjing Agricultural University, Jiangsu, China). *A. besseyi* was further cultured on *Botrytis cinerea* grown on a potato dextrose agar (PDA) medium for three weeks at 20 °C [67]. The mixture of various nematode stages was collected with a sufficient amount of ddH₂O in the Petri dishes. The harvested

A. besseyi was stored at 5 °C in the refrigerator until it was used in the experiment. In addition, in vivo studies were conducted using eggs obtained from a 25 μ m sieve after being rinsed twice with ddH₂O. The Baermann funnel method was used to gather second-stage juveniles (J2) after incubating the eggs for 5 days at 20 °C [68]. These juveniles (J2) were used to test the effect of Bt strains and their Cry31Aa toxin on *A. besseyi*.

4.3. Bioinformatics Analysis

The three *cry* genes were found by protein-nucleic acid alignment based on the whole genome of the Bt strains using SeqHunter 2.0 software which was created by Professor Daolong Dou Lab Nanjing Agricultural University China at 2010 based on Microsoft Visual Basic 6.0. A phylogenetic tree was constructed based on the multiple sequence alignment of the full-length of the three Cry proteins from the GBAC46 strain with known Cry proteins from BPPRC database (available online: https://camtech-bpp.ifas.ufl.edu/, accessed on 25 May 2021) using MEGA X 10.0.2. The alignment of the sequences of the Cry proteins was performed using DNAMAN. The crystal structures of the three Cry proteins were predicted based on protein homology using the SWISS-MODEL (available online: https://swissmodel.expasy.org/, accessed on 25 May 2021).

4.4. Expression of Cry31Aa Protein and Purification

The full-length PCR product of the Cry31Aa gene amplified from the GBAC46 strain's chromosomal DNA using primer Cry31Aa-F and Cry31Aa-R was cloned into the expression plasmid pOPTHis between *BamH I* and *Hind III* sites with N-terminal 6hismaltose-binding protein tag and C-terminal flag tag (for primers see Supplementary Materials Table S3). Then, the recombinant vector was transformed into BL21 (DE3), and the Cry31Aa protein was purified as described previously in [69], with some modifications. Briefly, the expression plasmid-carrying BL21 (DE3) cells were grown at 37 °C with 200 rpm shaking in LB medium containing 100 µg/mL of Amp until an OD₆₀₀ of 0.5–0.7 was achieved. Isopropyl thio- β -D-galactoside, at a final concentration of 1 mM, was added to flasks holding the cultures. The cells were extracted by centrifugation at 10,000× g for 20 min after 16 h of incubation at 16 °C with shaking at 200 rpm.

Using a high-pressure homogenizer, the bacterial pellets were resuspended in buffer A (50 mM Hepes, 300 mM NaCl, 5% glycerol, and 30 mM imidazole buffer at pH 8.0), and the cell debris was removed by centrifugation at $10,000 \times g$ for 20 min. The recombinant proteins were purified in a single chromatographic stage using a 5 mL HisTrapHP column (GE Healthcare, Milwaukee, WI, USA) at 5 mL/min on an AKTA avant 150 machine (GE Healthcare) [70]. The bacterial cell lysate was loaded onto the column, and non-adherent proteins were removed by washing the column with 20 volumes of wash buffer A. A 30–250 mM imidazole gradient was used to elute the proteins in wash buffer B (50 mM Hepes, 300 mM NaCl, 5% glycerol, and 250 mM imidazole buffer at pH 8.0). The purified enzymes were maintained at -80 °C after salt removal using Millipore Amicon Ultra. The proteins were detected and identified using SDS-PAGE. The protein concentration was determined using the Bradford Protein Assay Kit (P0006, Beyotime Institute of Biotechnology, Shanghai, China). For the Western blot, we followed the manufacturer's instructions and used method anti-Flag (F1804, Sigma-Aldrich, St. Louis, MO, USA).

4.5. Total RNA Extraction and cDNA Synthesis for RT-PCR

Total RNA was isolated from the GBAC46 strain after cultivating in liquid LB medium when the OD $_{= 600}$ reached up to 1.0, according to the Bacterial RNA extraction kit's procedure (OMEGA Bio-tek, Inc., Norcross, GA, USA). After being treated with DNase-I (Takara Shuzo, Takara, Japan), total RNA was reverse transcribed into cDNA using the HiScript II Q RT SuperMix Kit (Vazyme, Nanjing, China) according to the manufacturer's instructions and determined using the NanoDrop 1000 Spectrophotometer. In a ABI 7500 Fast Real-Time PCR Detection System, reverse transcriptase RT-PCR was performed using SYBR qPCR Master Mix (Vazyme, Nanjing, China).

4.6. Nematode Bioassays

The selected strains GBAC46, NMTD81, and FZB42 were cultured overnight in liquid LB medium, then washed with ddH₂O 2–3 times, and divided into various concentrations of colony-forming units (CFU/mL), 6.41×10^7 , 3.98×10^6 , 8.59×10^5 , and 7.21×10^4 , which were applied against *A. besseyi*. For nematode bioassays, the *A. besseyi* was treated with different concentrations of CFU/mL as mentioned above, and ddH₂O was applied as a negative control in a 96-well plate at 20 °C. The live and dead J2 were counted to calculate the mortality rate (%) using the microscopic examination.

4.7. Determination of Sporulation Formation

The production of spores was studied using sporulation medium (SM). Bt cells (1%) from an overnight culture were transferred to 20 mL SM media and cultured for 40 h at 37 °C with 200 rpm shaking. The spore development was observed in one microliter of the culture. Malachite green and safranine O were employed in the spore staining kit (HB8300-2, Haibo Biotechnology Co., Ltd., Suzhou, China). Malachite green was used to color the spores blue, and safranine O was used to counterstain the living cells red [69]. The experimental setups were carried out in accordance with the manufacturer's instructions. Using an Olympus BX43 microscope, spores and living cells dyed with various colors were seen using cell Sens Standard Software (Tokyo, Japan).

4.8. Phenotypic Observation by Scanning Electron Microscopy

The selected Bt strains were cultivated for 72 h at 30 °C with a 200 rpm shaking time on a liquid LB medium. Centrifugation was used to recover the spore–crystal mixture, and the spore–crystal mixtures (1 mL of liquid culture) were washed three times with 1 M NaCl and ice-cold ddH₂O and then rinsed 2–3 times with ddH₂O. The parasporal crystal protein was observed using scanning electron microscopy (SEM). Following the company protocols, the samples were fixed in 2.5% glutaraldehyde and 1% osmic acid, coated with gold particles, and then observed using a Hitachi S-3000N SEM at a voltage of 5 kV (Hitachi, Tokyo, Japan).

4.9. Endocytosis Assays with Rhodamine-Labeled Cry31Aa

The endocytosis assays were carried out as previously described in [52], with some modifications. The 5(6)-carboxytetramethylrhodamine is a biological stain with strong fluorescence when dissolved in DMSO solvent and is often used as a biological dye [52]. A of 1 mg/mL of rhodamine was used to label the three Cry proteins to detect whether the protein was ingested by nematodes, and PBS was used as the negative control. After 2 h, the nematode samples were washed three times with sterilized water before being observed with an Olympus fluorescent microscope. An objective lens $\times 40$ was used to capture the images. The rhodamine filter was used to detect fluorescence at the same time.

4.10. Propidium Iodide Uptake Assays

Propidium iodide (Sigma) uptake tests were used for pore-formation experiments, according to the methodology reported in [71], with some modifications. Nematodes were observed using a fluorescent microscope after being incubated with Cry31Aa for 4 h and 20 mM propidium iodide for 3 h. The wavelengths of excitation and emission were 555 and 580 nm, respectively. ImageJ was used to process the images.

4.11. In Vivo Experiments with Rice Plants

The rice cultivar (Daohuahua no. 4) was grown in a greenhouse in controlled condition. Seeds were surface sterilized by 70% ethanol for 30 s (s), 5% (w/v) sodium hypochlorite solution for 20 min, and finally washed with ddH₂O four times. The rice seeds were soaked in ddH₂O for 36 h to accelerate germination. After 36 h, the rice seeds were dipped into *A. besseyi* nematode solution (nematode concentration of approximately 500 juveniles/mL) and ddH₂O was used as a control (CK) and placed at room temperature for 36 h. The soil

was autoclaved at 140 °C for 40 min. A similar sized pot was filled with sterilized soil and used for sowing rice seeds. The 15 mL of overnight culture of the selected strains (i.e., GBAC466 and NMTD81) and the positive control, FZB42, with an $OD_{600nm} = 1.0$ was applied in rice seedlings grown in sterilized soil. Then, the rice seeds were cultured for one month in a greenhouse under control conditions.

4.12. Estimation of MDA Level (Lipid Peroxidation) in Rice Plants

Lipid peroxidation was examined in terms of the malondialdehyde (MDA) level studied in plant leaves under stress conditions according to [72], with some changes. Briefly, the fresh leaves (0.1 g) of rice plants treated with the selected strains as well as the control plants were taken and homogenized in a 0.1% (w/v) TCA 500 µL solution. The mixture was then centrifuged at 4 °C and 13,000 rpm. The supernatant of each treatment was then mixed with 1.5 mL of 0.5% TBA solution and heated to 95 °C in water for 25 min. The mixture was placed on ice for 5 min to stop any further reaction. The absorbance of the reaction was analyzed at 532 and 600 nm using a microplate reader (Spectrum max plus; Molecular Devices, Sunnyvale, CA, USA).

4.13. Determination of Hydrogen Peroxide

Rice leaves extracts were tested for hydrogen peroxides (H_2O_2) concentration. Fresh rice plant leaves (0.1 g) were homogenized at 4uC in a 1:9 (w/v) phosphate buffer (50 mM, pH 6.0). A hydrogen peroxide assay kit (Beyotime, Shanghai, China) was used to determine the amount of H_2O_2 in the leaf sample. The test tubes containing 50 mL test solutions were left at room temperature for 30 min before being measured using a spectrometer at 560 nm. The absorbance measurements were calibrated using a standard curve with known H_2O_2 concentrations [72].

4.14. Defense-Related Genes' Expression in Rice Plants

The expression profile of the defense genes (i.e., *PR1a*, *PR4*, *LOX1*, *PBZ1*, and *PAL1*) was carried out through RT q-PCR in rice plants grown after being treated with selected strains in a greenhouse experiment. For this, the selected gene sequences were taken from NCB1, followed by designing primers through the PrimerQuest tool; the primers are listed in the Supplementary Materials Table S2. The housekeeping gene elongation factor 1-alpha (*ef1*) was used in the present study. Briefly, RNA was extracted from fresh rice plant leaves inoculated with selected strains and ddH₂O was used as the control grown under infested and non-infested *A. besseyi* in greenhouse conditions after 4 days' post-inoculation (dpi) through the TRizole method. The Vazyme HiScript II Q RT SuperMix Kit (Vazyme, Nanjing, China) was used for cDNA synthesis. RT-qPCR was performed to analyze the expression profile of selected genes in rice plants through a ABI 7500 Fast Real-Time PCR Detection System (Thermo Fisher Scientific, San Jose, CA, USA). The PCR machine was programmed using the following steps: initial denaturation at 95 °C for 30 s, including 40 cycles of 95 °C for 5 s, and 34 s at 60 °C. Finally, relative quantification was performed according to the comparative C method of $2^{-\Delta\Delta CT}$ as described in [73].

4.15. Statistical Analysis

The data were analyzed using one-way analysis of variance (ANOVA) in SPSS version 26 software and reported as the mean \pm standard deviation from three biological replicates. For group comparisons, Fisher's LSD test was utilized. Outcomes were considered significant at p < 0.05.

5. Conclusions

In conclusion, the selected Bt strains from the Tibet region, China, showed high nematicidal activity against *A. besseyi*. Three proteins (i.e., Cry31Aa, Cry73Aa, and Cry40ORF) were also characterized and functionally assessed for food safety. The novel Cry protein, Cry31Aa, possessed highly efficient and selective nematicidal activity. Furthermore, the current study adds new insights into the mechanisms by which nematicidal Bt regulates defense-related genes and improves plant growth promotion parameters in rice as well as contributes to the control of plant parasitic nematode diseases.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23158189/s1.

Author Contributions: Z.L., Q.A. and X.G., planned and designed the research; Z.L. and Q.A., Y.W., performed the research and contributed to the methodology, writing, and editing; G.M., Y.R., X.K. and H.M., assisted in the analysis and compiled the data and results; Q.G., H.W. and X.G., critically revised and improved the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Guidance Foundation of the Sanya Institute of Nanjing Agricultural University (NAUSY-MS18), Fundamental Research Funds for the Central Universities (KYZZ2022001), and the Key Project of NSFC Regional Innovation and Development Joint Fund (U20A2039).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ab Rahman, S.F.S.; Singh, E.; Pieterse, C.M.J.; Schenk, P.M. Emerging microbial biocontrol strategies for plant pathogens. *Plant Sci.* 2018, 267, 102–111. [CrossRef] [PubMed]
- Ali, Q.; Zheng, H.; Rao, M.J.; Ali, M.; Hussain, A.; Saleem, M.H.; Nehela, Y.; Sohail, M.A.; Ahmed, A.M.; Kubar, K.A.; et al. Advances, limitations, and prospects of biosensing technology for detecting phytopathogenic bacteria. *Chemosphere* 2022, 296, 133773. [CrossRef] [PubMed]
- Jin, H.; Cui, H.; Yang, X.; Xu, L.; Li, X.; Liu, R.; Yan, Z.; Li, X.; Zheng, W.; Zhao, Y.; et al. Nematicidal activity against *Aphelenchoides* besseyi and Ditylenchus destructor of three biflavonoids, isolated from roots of Stellera chamaejasme. J. Asia Pac. Entomol. 2018, 21, 1473–1478. [CrossRef]
- 4. Cheng, C.; Gao, X.; Feng, B.; Sheen, J.; Ssheen, L.; He, P. Differential temperature operation of plant immune responses. *Nat. Commun.* **2013**, *4*, 2530. [CrossRef]
- 5. Zhao, B.; Lin, X.; Poland, J.; Trick, H.; Leach, J.; Hulbert, S. A maize resistance gene functions against bacterial streak disease in rice. *Proc. Natl. Acad. Sci. USA* 2005, 102, 15383–15388. [CrossRef]
- 6. Wang, F.; Li, D.; Wang, Z.; Dong, A.; Liu, L.; Wang, B.; Chen, Q.; Liu, X. Transcriptomic analysis of the rice white tip nematode, Aphelenchoides besseyi (Nematoda: Aphelenchoididae). *PLoS ONE* **2014**, *9*, e91591. [CrossRef]
- Gu, Q.; Qiao, J.; Wang, R.; Lu, J.; Wang, Z.; Li, P.; Zhang, L.; Ali, Q.; Khan, A.R.; Gao, X.; et al. The Role of Pyoluteorin from *Pseudomonas protegens* Pf-5 in Suppressing the Growth and Pathogenicity of *Pantoea ananatis* on Maize. *Int. J. Mol. Sci.* 2022, 23, 6431. [CrossRef]
- Yu, C.; Liu, X.; Zhang, X.; Zhang, M.; Gu, Y.; Ali, Q.; Mohamed, M.; Xu, J.; Shi, J.; Gao, X. Mycosubtilin Produced by *Bacillus subtilis* ATCC6633 Inhibits Growth and Mycotoxin Biosynthesis of Fusarium graminearum and Fusarium verticillioides. *Toxins* 2021, 13, 791. [CrossRef]
- 9. Jasmer, D.P.; Rosa, B.A.; Tyagi, R.; Mitreva, M. Rapid determination of nematode cell and organ susceptibility to toxic treatments. *Int. J. Parasitol. Drugs Drug Resist.* 2020, 14, 167–182. [CrossRef]
- Ayaz, M.; Ali, Q.; Farzand, A.; Khan, A.R.; Ling, H.; Gao, X. Nematicidal volatiles from bacillus atrophaeus gbsc56 promote growth and stimulate induced systemic resistance in tomato against meloidogyne incognita. *Int. J. Mol. Sci.* 2021, 22, 5049. [CrossRef]
- Yu, Z.; Xiong, J.; Zhou, Q.; Luo, H.; Hu, S.; Xia, L.; Sun, M.; Li, L.; Yu, Z. The diverse nematicidal properties and biocontrol efficacy of *Bacillus thuringiensis* Cry6A against the root-knot nematode Meloidogyne hapla. J. Invertebr. Pathol. 2015, 125, 73–80. [CrossRef]
- 12. Peng, D.; Wan, D.; Cheng, C.; Ye, X.; Sun, M. Nematode-specific cadherin CDH-8 acts as a receptor for Cry5B toxin in Caenorhabditis elegans. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 3663–3673. [CrossRef]
- 13. Schnepf, E.; Crickmore, N.; Van Rie, J.; Lereclus, D.; Baum, J.; Feitelson, J.; Zeigler, D.R.; Dean, D. *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 775–806. [CrossRef]
- 14. Wei, J.-Z.; Hale, K.; Carta, L.; Platzer, E.; Wong, C.; Fang, S.-C.; Aroian, R. V *Bacillus thuringiensis* crystal proteins that target nematodes. *Proc. Natl. Acad. Sci. USA* 2003, 100, 2760–2765. [CrossRef]

- 15. Bravo, A.; Likitvivatanavong, S.; Gill, S.S.; Soberón, M. *Bacillus thuringiensis*: A story of a successful bioinsecticide. *Insect Biochem. Mol. Biol.* **2011**, *41*, 423–431. [CrossRef]
- 16. Wan, L.; Lin, J.; Du, H.; Zhang, Y.; Bravo, A.; Soberón, M.; Sun, M.; Peng, D. *Bacillus thuringiensis* targets the host intestinal epithelial junctions for successful infection of Caenorhabditis elegans. *Environ. Microbiol.* **2019**, *21*, 1086–1098. [CrossRef]
- 17. Rosas-García, N.M. Biopesticide production from *Bacillus thuringiensis*: An environmentally friendly alternative. *Recent Pat. Biotechnol.* **2009**, *3*, 28–36. [CrossRef]
- Kumar, S.; Chandra, A.; Pandey, K.C. Bacillus thuringiensis (Bt) transgenic crop: An environment friendly insect-pest management strategy. J. Environ. Biol. 2008, 29, 641–653.
- 19. Stewart, C.N., Jr.; Adang, M.J.; All, J.N.; Raymer, P.L.; Ramachandran, S.; Parrott, W.A. Insect control and dosage effects in transgenic canola containing a synthetic *Bacillus thuringiensis* cryIAc gene. *Plant Physiol.* **1996**, *112*, 115–120. [CrossRef]
- Singh, S.; Kumar, N.R.; Maniraj, R.; Lakshmikanth, R.; Rao, K.Y.S.; Muralimohan, N.; Arulprakash, T.; Karthik, K.; Shashibhushan, N.B.; Vinutha, T. Expression of Cry2Aa, a *Bacillus thuringiensis* insecticidal protein in transgenic pigeon pea confers resistance to gram pod borer, Helicoverpa armigera. *Sci. Rep.* 2018, *8*, 8820. [CrossRef]
- Bříza, J.; Pavingerová, D.; Vlasák, J.; Niedermeierová, H. Norway spruce (*Picea abies*) genetic transformation with modified Cry3A gene of *Bacillus thuringiensis*. Acta Biochim. Pol. 2013, 60, 395–400. [CrossRef]
- Jones, J.; Haegeman, A.; Danchin, E.; Gaur, H.; Helder, J.; Jones, M.; Kikuchi, T.; Manzanilla-Lopez, R.; Palomares-Rius, J.; Wesemael, W. Top 10 plant-parasitic nematodes in molecular plant pathology. In Proceedings of the 6th International Congress of Nematology, Cape Town, South Africa, 4–9 May 2014.
- Ali, Q.; Ahmar, S.; Sohail, M.A.; Kamran, M.; Ali, M.; Saleem, M.H.; Rizwan, M.; Ahmed, A.M.; Mora-Poblete, F.; do Amaral Júnior, A.T.; et al. Research advances and applications of biosensing technology for the diagnosis of pathogens in sustainable agriculture. *Environ. Sci. Pollut. Res.* 2021, 28, 9002–9019. [CrossRef]
- 24. Ali, M.; Ali, Q.; Sohail, M.A.; Ashraf, M.F.; Saleem, M.H.; Hussain, S.; Zhou, L. Diversity and Taxonomic Distribution of Endophytic Bacterial Community in the Rice Plant and Its Prospective. *Int. J. Mol. Sci.* **2021**, *22*, 10165. [CrossRef]
- 25. Hu, Y.; Georghiou, S.B.; Kelleher, A.J.; Aroian, R.V. *Bacillus thuringiensis* Cry5B protein is highly efficacious as a single-dose therapy against an intestinal roundworm infection in mice. *PLoS Negl. Trop. Dis.* **2010**, *4*, e614. [CrossRef]
- 26. Feng, H.; Zhou, D.; Daly, P.; Wang, X.; Wei, L. Characterization and functional importance of two glycoside hydrolase family 16 genes from the rice white tip nematode Aphelenchoides besseyi. *Animals* **2021**, *11*, 374. [CrossRef]
- 27. Shi, J.; Peng, D.; Zhang, F.; Ruan, L.; Sun, M. The Caenorhabditis elegans CUB-like-domain containing protein RBT-1 functions as a receptor for *Bacillus thuringiensis* Cry6Aa toxin. *PLoS Pathog.* **2020**, *16*, e1008501. [CrossRef]
- Armada, E.; Probanza, A.; Roldán, A.; Azcón, R. Native plant growth promoting bacteria *Bacillus thuringiensis* and mixed or individual mycorrhizal species improved drought tolerance and oxidative metabolism in Lavandula dentata plants. *J. Plant Physiol.* 2016, 192, 1–12. [CrossRef]
- 29. Ali, Q.; Ayaz, M.; Yu, C.; Wang, Y.; Gu, Q.; Wu, H.; Gao, X. Cadmium tolerant microbial strains possess different mechanisms for cadmium biosorption and immobilization in rice seedlings. *Chemosphere* **2022**, *303*, 135206. [CrossRef]
- Ben Abdallah, D.; Frikha-Gargouri, O.; Tounsi, S. Rizhospheric competence, plant growth promotion and biocontrol efficacy of Bacillus amyloliquefaciens subsp. plantarum strain 32a. *Biol. Control* 2018, 124, 61–67. [CrossRef]
- Cherif-Silini, H.; Silini, A.; Yahiaoui, B.; Ouzari, I.; Boudabous, A. Phylogenetic and plant-growth-promoting characteristics of Bacillus isolated from the wheat rhizosphere. *Ann. Microbiol.* 2016, *66*, 1087–1097. [CrossRef]
- 32. Raddadi, N.; Cherif, A.; Ouzari, H.; Marzorati, M.; Brusetti, L.; Boudabous, A.; Daffonchio, D. *Bacillus thuringiensis* beyond insect biocontrol: Plant growth promotion and biosafety of polyvalent strains. *Ann. Microbiol.* **2007**, *57*, 481–494. [CrossRef]
- Liaquat, F.; Munis, M.F.H.; Arif, S.; Haroon, U.; Shi, J.; Saqib, S.; Zaman, W.; Che, S.; Liu, Q. PacBio Single-Molecule Long-Read Sequencing Reveals Genes Tolerating Manganese Stress in Schima superba Saplings. *Front. Genet.* 2021, 12, 635043. [CrossRef] [PubMed]
- 34. Azizoglu, U. *Bacillus thuringiensis* as a biofertilizer and biostimulator: A mini-review of the little-known plant growth-promoting properties of Bt. *Curr. Microbiol.* **2019**, *76*, 1379–1385. [CrossRef] [PubMed]
- 35. Iatsenko, I.; Boichenko, I.; Sommer, R.J. *Bacillus thuringiensis* DB27 produces two novel protoxins, Cry21Fa1 and Cry21Ha1, which act synergistically against nematodes. *Appl. Environ. Microbiol.* **2014**, *80*, 3266–3275. [CrossRef] [PubMed]
- Ye, W.; Zhu, L.; Liu, Y.; Crickmore, N.; Peng, D.; Ruan, L.; Sun, M. Mining new crystal protein genes from *Bacillus thuringiensis* on the basis of mixed plasmid-enriched genome sequencing and a computational pipeline. *Appl. Environ. Microbiol.* 2012, 78, 4795–4801. [CrossRef] [PubMed]
- 37. Crickmore, N.; Berry, C.; Panneerselvam, S.; Mishra, R.; Connor, T.R.; Bonning, B.C. A structure-based nomenclature for *Bacillus thuringiensis* and other bacteria-derived pesticidal proteins. *J. Invertebr. Pathol.* **2021**, *186*, 107438. [CrossRef]
- Peng, Q.; Yu, Q.; Song, F. Expression of cry genes in *Bacillus thuringiensis* biotechnology. *Appl. Microbiol. Biotechnol.* 2019, 103, 1617–1626. [CrossRef]
- Vilas-Bôas, G.T.; Peruca, A.P.S.; Arantes, O.M.N. Biology and taxonomy of *Bacillus cereus*, *Bacillus anthracis*, and *Bacillus thuringiensis*. Can. J. Microbiol. 2007, 53, 673–687. [CrossRef]
- 40. Vilchez, S. Making 3D-cry toxin mutants: Much more than a tool of understanding toxins mechanism of action. *Toxins* **2020**, *12*, 600. [CrossRef]

- 41. Xu, L.-N.; Wang, Y.-Q.; Wang, Z.-Y.; Hu, B.-J.; Ling, Y.-H.; He, K.-L. Transcriptome differences between Cry1Ab resistant and susceptible strains of Asian corn borer. *BMC Genom.* **2015**, *16*, 173. [CrossRef]
- Huang, T.; Lin, Q.; Qian, X.; Zheng, Y.; Yao, J.; Wu, H.; Li, M.; Jin, X.; Pan, X.; Zhang, L. Nematicidal activity of Cry1Ea11 from Bacillus thuringiensis BRC-XQ12 against the pine wood nematode (Bursaphelenchus xylophilus). Phytopathology 2018, 108, 44–51. [CrossRef]
- 43. Rai, V.; Acharya, S.; Dey, N. Implications of Nanobiosensors in Agriculture. J. Biomater. Nanobiotechnol. 2012, 3, 315–324. [CrossRef]
- 44. Sinha, A.; Rae, R.; Iatsenko, I.; Sommer, R.J. System wide analysis of the evolution of innate immunity in the nematode model species Caenorhabditis elegans and Pristionchus pacificus. *PLoS ONE* **2012**, *7*, e44255. [CrossRef]
- 45. Liu, Z.; Budiharjo, A.; Wang, P.; Shi, H.; Fang, J.; Borriss, R.; Zhang, K.; Huang, X. The highly modified microcin peptide plantazolicin is associated with nematicidal activity of Bacillus amyloliquefaciens FZB42. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 10081–10090. [CrossRef]
- 46. Guo, S.; Liu, M.; Peng, D.; Ji, S.; Wang, P.; Yu, Z.; Sun, M. New strategy for isolating novel nematicidal crystal protein genes from *Bacillus thuringiensis* strain YBT-1518. *Appl. Environ. Microbiol.* **2008**, *74*, 6997–7001. [CrossRef]
- Yu, Z.; Luo, H.; Xiong, J.; Zhou, Q.; Xia, L.; Sun, M.; Li, L.; Yu, Z. B acillus thuringiensis C ry6 A exhibits nematicidal activity to C aenorhabditis elegans bre mutants and synergistic activity with C ry5 B to C. elegans. *Lett. Appl. Microbiol.* 2014, 58, 511–519. [CrossRef]
- 48. Ji, F.; Zhu, Y.; Ju, S.; Zhang, R.; Yu, Z.; Sun, M. Promoters of crystal protein genes do not control crystal formation inside exosporium of *Bacillus thuringiensis* ssp. finitimus strain YBT-020. *FEMS Microbiol. Lett.* **2009**, 300, 11–17. [CrossRef]
- Ammons, D.R.; Reyna, A.; Granados, J.C.; Ventura-Suárez, A.; Rojas-Avelizapa, L.I.; Short, J.D.; Rampersad, J.N. A novel Bacillus thuringiensis Cry-like protein from a rare filamentous strain promotes crystal localization within the exosporium. Appl. Environ. Microbiol. 2013, 79, 5774–5776. [CrossRef]
- 50. Belair, G. Nematodes, these roundworms that harm plants ... by their roots. *Phytoprotection* 2005, 86, 65–69.
- 51. Eves-van den Akker, S.; Lilley, C.J.; Ault, J.R.; Ashcroft, A.E.; Jones, J.T.; Urwin, P.E. The feeding tube of cyst nematodes: Characterisation of protein exclusion. *PLoS ONE* **2014**, *9*, e87289.
- 52. Kahn, T.W.; Duck, N.B.; McCarville, M.T.; Schouten, L.C.; Schweri, K.; Zaitseva, J.; Daum, J. A *Bacillus thuringiensis* Cry protein controls soybean cyst nematode in transgenic soybean plants. *Nat. Commun.* **2021**, *12*, 3380. [CrossRef]
- 53. Ramalakshmi, A.; Sharmila, R.; Iniyakumar, M.; Gomathi, V. Nematicidal activity of native *Bacillus thuringiensis* against the root knot nematode, Meloidogyne incognita (Kofoid and White). *Egypt. J. Biol. Pest Control* **2020**, *30*, 90. [CrossRef]
- 54. Chandler, S.; Van Hese, N.; Coutte, F.; Jacques, P.; Höfte, M.; De Vleesschauwer, D. Role of cyclic lipopeptides produced by *Bacillus subtilis* in mounting induced immunity in rice (*Oryza sativa* L.). *Physiol. Mol. Plant Pathol.* **2015**, *91*, 20–30. [CrossRef]
- 55. Huang, H.; Ayaz, A.; Zheng, M.; Yang, X.; Zaman, W.; Zhao, H.; Lü, S. Arabidopsis KCS5 and KCS6 Play Redundant Roles in Wax Synthesis. *Int. J. Mol. Sci.* 2022, 23, 4450. [CrossRef]
- 56. Mengiste, T. Plant immunity to necrotrophs. Annu. Rev. Phytopathol. 2012, 50, 267–294. [CrossRef]
- 57. Manghwar, H.; Hussain, A. Mechanism of tobacco osmotin gene in plant responses to biotic and abiotic stress tolerance: A brief history. *Biocell* **2022**, *46*, 623–632. [CrossRef]
- Ayaz, A.; Huang, H.; Zheng, M.; Zaman, W.; Li, D.; Saqib, S.; Zhao, H.; Lü, S. Molecular cloning and functional analysis of GmLACS2-3 reveals its involvement in cutin and suberin biosynthesis along with abiotic stress tolerance. *Int. J. Mol. Sci.* 2021, 22, 9175. [CrossRef]
- 59. Weber, H.; Chételat, A.; Reymond, P.; Farmer, E.E. Selective and powerful stress gene expression in Arabidopsis in response to malondialdehyde. *Plant J.* 2004, *37*, 877–888. [CrossRef]
- Salehi Jouzani, G.; Seifinejad, A.; Saeedizadeh, A.; Nazarian, A.; Yousefloo, M.; Soheilivand, S.; Mousivand, M.; Jahangiri, R.; Yazdani, M.; Amiri, R.M.; et al. Molecular detection of nematicidal crystalliferous *Bacillus thuringiensis* strains of Iran and evaluation of their toxicity on free-living and plant-parasitic nematodes. *Can. J. Microbiol.* 2008, 54, 812–822. [CrossRef]
- 61. Wilson, M.K.; Abergel, R.J.; Raymond, K.N.; Arceneaux, J.E.L.; Byers, B.R. Siderophores of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*. *Biochem. Biophys. Res. Commun.* **2006**, *348*, 320–325. [CrossRef]
- Manghwar, H.; Hussain, A.; Ali, Q.; Saleem, M.H.; Abualreesh, M.H.; Alatawi, A.; Ali, S.; Munis, M.F.H. Disease Severity, Resistance Analysis, and Expression Profiling of Pathogenesis-Related Protein Genes after the Inoculation of Fusarium equiseti in Wheat. *Agronomy* 2021, *11*, 2124. [CrossRef]
- 63. Kong, M.; Sheng, T.; Liang, J.; Ali, Q.; Gu, Q.; Wu, H.; Chen, J.; Liu, J.; Gao, X. Melatonin and Its Homologs Induce Immune Responses via Receptors trP47363-trP13076 in *Nicotiana benthamiana*. *Front. Plant Sci.* **2021**, *12*, 691835. [CrossRef] [PubMed]
- Tahir, H.A.S.; Gu, Q.; Wu, H.; Raza, W.; Safdar, A.; Huang, Z.; Rajer, F.U.; Gao, X. Effect of volatile compounds produced by Ralstonia solanacearum on plant growth promoting and systemic resistance inducing potential of Bacillus volatiles. *BMC Plant Biol.* 2017, 17, 133. [CrossRef] [PubMed]
- 65. Wu, H.; Gu, Q.; Xie, Y.; Lou, Z.; Xue, P.; Fang, L.; Yu, C.; Jia, D.; Huang, G.; Zhu, B.; et al. Cold-adapted Bacilli isolated from the Qinghai—Tibetan Plateau are able to promote plant growth in extreme environments. *Environ. Microbiol.* **2019**, *21*, 3505–3526. [CrossRef]
- Fernandez-Luna, M.T.; Kumar, P.; Hall, D.G.; Mitchell, A.D.; Blackburn, M.B.; Bonning, B.C. Toxicity of *Bacillus thuringiensis*derived pesticidal proteins cry1Ab and cry1Ba against asian citrus psyllid, *Diaphorina citri* (Hemiptera). *Toxins* 2019, *11*, 173. [CrossRef]

- 67. Feng, H.; Shao, Y.; Wei, L.H.; Gao, C.Y.; Zhou, Y.J. The white-tip nematode, Aphelenchoides besseyi, exhibits an auxin-orientated behaviour affecting its migration and propagation. *Nematology* **2014**, *16*, 837–845. [CrossRef]
- 68. Viglierchio, D.R.; Schmitt, R. V On the methodology of nematode extraction from field samples: Baermann funnel modifications. *J. Nematol.* **1983**, *15*, 438–444.
- Liang, Z.; Qiao, J.Q.; Li, P.P.; Zhang, L.L.; Qiao, Z.X.; Lin, L.; Yu, C.J.; Yang, Y.; Zubair, M.; Gu, Q.; et al. A novel Rap-Phr system in Bacillus velezensis NAU-B3 regulates surfactin production and sporulation via interaction with ComA. *Appl. Microbiol. Biotechnol.* 2020, 104, 10059–10074. [CrossRef]
- 70. Wu, H.; Yang, Y.; Wang, S.; Qiao, J.; Xia, Y.; Wang, Y.; Wang, W.; Gao, S.; Liu, J.; Xue, P. Cloning, expression and characterization of a new aspartate aminotransferase from *Bacillus subtilis* B3. *FEBS J.* **2011**, *278*, 1345–1357. [CrossRef]
- 71. Kourtis, N.; Nikoletopoulou, V.; Tavernarakis, N. Small heat-shock proteins protect from heat-stroke-associated neurodegeneration. *Nature* **2012**, *490*, 213–218. [CrossRef]
- 72. Chen, H.; Gong, Y.; Han, R. Cadmium telluride quantum dots (CdTe-QDs) and enhanced ultraviolet-B (UV-B) radiation trigger antioxidant enzyme metabolism and programmed cell death in wheat seedlings. *PLoS ONE* **2014**, *9*, e110400. [CrossRef]
- 73. Kong, M.; Liang, J.; Ali, Q.; Wen, W.; Wu, H.; Gao, X.; Gu, Q. 5-Methoxyindole, a Chemical Homolog of Melatonin, Adversely Affects the Phytopathogenic Fungus *Fusarium graminearum*. *Int. J. Mol. Sci.* **2021**, 22, 10991. [CrossRef]



Remiern



Degradation Mechanism of Autophagy-Related Proteins and Research Progress

Yanhui Zhou^{1,2}, Hakim Manghwar¹, Weiming Hu^{1,2,*} and Fen Liu^{1,2,*}

- ¹ Lushan Botanical Garden, Chinese Academy of Sciences, Jiujiang 332900, China; zhouyanhui994@163.com (Y.Z.); hakim@lsbg.cn (H.M.)
- ² School of Life Sciences, Nanchang University, Nanchang 330031, China

Correspondence: huwm@lsbg.cn (W.H.); liuf@lsbg.cn (F.L.)

Abstract: In all eukaryotes, autophagy is the main pathway for nutrient recycling, which encapsulates parts of the cytoplasm and organelles in double-membrane vesicles, and then fuses with lysosomes/vacuoles to degrade them. Autophagy is a highly dynamic and relatively complex process influenced by multiple factors. Under normal growth conditions, it is maintained at basal levels. However, when plants are subjected to biotic and abiotic stresses, such as pathogens, drought, waterlogging, nutrient deficiencies, etc., autophagy is activated to help cells to survive under stress conditions. At present, the regulation of autophagy is mainly reflected in hormones, second messengers, post-transcriptional regulation, and protein post-translational modification. In recent years, the degradation mechanism of autophagy-related proteins has attracted much attention. In this review, we have summarized how autophagy-related proteins are degraded in yeast, animals, and plants, which will help us to have a more comprehensive and systematic understanding of the regulation mechanisms of autophagy. Moreover, research progress on the degradation of autophagy-related proteins in plants has been discussed.

Keywords: autophagy-related protein; degradation; ubiquitin; proteasome; autophagy

1. Introduction

Autophagy widely exists in eukaryotic cells, which is a relatively conservative process in evolution. It is responsible for the transport of certain cytoplasmic proteins and subcellular organelles into lysosomes/vacuoles for degradation, thereby contributing to the recycling of intracellular nutrients [1,2]. In yeast and mammals, autophagy consists of three types: microautophagy, macroautophagy, and chaperone-mediated autophagy (CMA). However, CMA is absent in plants and has been replaced by mega-autophagy [3]. Among them, the so-called autophagy usually refers to macroautophagy, which contains both selective and non-selective intracellular degradation processes [4] (Figure 1). Selective autophagy is distinguished from bulk autophagy by the use of selective autophagy receptors (SARs) [5]. Macroautophagy begins with a series of autophagy-related (ATG) proteins forming phagophore structures at the assembly site of the phagophore, and then recruiting substrate molecules, and forming autophagosome with double-layer membrane structure through the expansion and closure of vesicles. Subsequently, the outer membrane of the autophagosome fuses with the lysosomal or tonoplast membrane to release the contents into the lysosomal lumen or vacuole in the form of an autophagic body with only a single membrane, and is degraded into small molecular substances for recycling under the role of acid hydrolase [6–8]. Microautophagy is a process that isolates and uptakes cell components by direct enclosure with the lysosomal/vacuolar membrane [9]. Autophagy plays an important role in maintaining cellular homeostasis. Plants are exposed to many biotic and abiotic stresses, and these stresses affect plants [10–16]. When cells face stress conditions, such as nutrient deprivation, oxygen deficiency, and endoplasmic reticulum (ER) damage,

Citation: Zhou, Y.; Manghwar, H.; Hu, W.; Liu, F. Degradation Mechanism of Autophagy-Related Proteins and Research Progress. *Int. J. Mol. Sci.* 2022, 23, 7301. https:// doi.org/10.3390/ijms23137301

Academic Editor: Wajid Zaman

Received: 2 June 2022 Accepted: 29 June 2022 Published: 30 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).



autophagy is highly induced to maintain metabolic and energy balance [16–19]. During starvation, lysosomal activity is also improved to support autophagic flux [20].

Figure 1. Morphological steps of microautophagy and macroautophagy in eukaryotes. Macroautophagy begins with the formation of a phagophore that encapsulates damaged organelles and discarded proteins. Then, through the extension of the vesicle forms a closed structure with a double membrane, which is called an autophagosome. Subsequently, the outer membrane of the autophagosome fuses with the lysosomal membrane (animals) or the tonoplast (yeast and plants) to release the autophagic body with only a single membrane. Finally, under the digestion of acid hydrolases, the cargoes were degraded into small molecular substances for recycling. Microautophagy is a process in which the lysosome/vacuole directly packages target substrates by membrane invagination to create the autophagic body.

Autophagy is a complex pathway that is tightly regulated. The expression of autophagyrelated genes has spatiotemporal specificity and is a dynamic process. Regulation of autophagy-related proteins at different levels, including transcription, post-transcription, translation, and post-translation, which helps to inhibit or activate autophagy [21–23]. At the transcriptional level, when plants face drought stress, overexpression of heat shock factor A1a (HsfA1a) can increase the mRNA levels of ATG10 and ATG18f, and the number of autophagosomes is also increased accordingly [24]. In mammals, microRNAs (miRNAs) regulate autophagy at the post-transcriptional level [25]. Post-translational modifications, such as phosphorylation, lipidation, and ubiquitination of autophagy-related proteins, are critical in regulating autophagic activity and duration [23]. For example, AMP-activated protein kinase (AMPK) phosphorylates BECN1 (beclin 1, vacuolar protein sorting (Vps) 30/Atg6 in yeast, and ATG6 in plants) at the site of Thr388 to induce autophagy [26]. ATG4a is modified by hydrogen sulfide (H2S) persulfidation, inhibiting protease activity, thereby negatively regulating autophagy [27]. During the ubiquitination modification process, autophagy-related proteins are labeled with ubiquitin, which is recognized and degraded by the 26S proteasome [28], ultimately inhibiting the production of autophagic vacuoles. In recent years, the regulatory pathways of autophagy-related protein degradation have attracted much attention. In this review, we have mainly focused on the recent advances in core autophagy-related protein degradation mechanisms.

2. ATG Complexes in Plants and Animals

ATG proteins play an indispensable role in the process of autophagy in plant and animal cells. According to their diverse functions in different stages of autophagy, core ATG proteins can be divided into four major complexes (Figure 2).



(A-C) ATGs complexes

(D) Core ATG proteins of plants and mammals

Figure 2. Core protein complexes of autophagy in plants and animals. (A) The ATG1/ATG13 protein kinase complex. When plants are under nutrient-rich conditions, TOR kinase hyperphosphorylates ATG13. While plants are placed in nutrient starved conditions, the TOR kinase is inactivated, resulting in the dephosphorylation of ATG13, which binds tightly to ATG1. Then, the ATG1 kinase activity is activated and autophosphorylation occurs to form the ATG1, ATG11, ATG13, and ATG101 complex, which results in upregulating autophagy. (B) The ATG9/2/18 transmembrane complex and PI3K complex. ATG9 delivers membrane source and mediates the extension of phagophore membrane. ATG2 and ATG18 play a synergistic role in this process. The PI3K protein complex promotes the nucleation of vesicles, which include ATG6, ATG14, VPS34, VPS15, and PI3P. (C) The ATG5-ATG12 and ATG8-PE ubiquitin-like conjugating systems. ATG12 was transferred to the target protein ATG5 with the help of ATG7 and ATG10. Subsequently, the ATG5-ATG12 complex combines with ATG16 to form an oligomeric complex, which participated in the esterification of ATG8. During the covalent binding stage of ATG8-PE, the cysteine protease ATG4 cleaves the C-terminus of ATG8. Subsequently, ATG8 is activated by ATG7 and transferred to ATG3 through a thioester bond. At last, with the help of ATG5-ATG12-ATG16 conjugate, ATG8 forms an ATG8-PE adduct with phosphatidylethanolamine. (D) Core ATG proteins of plants and mammals in four complexes.

2.1. ATG1/ATG13 Protein Kinase Complex

The ATG1/ATG13 protein kinase complex is located at the most upstream position in the ATG proteins recruitment hierarchy and initiates the formation of phagophores in response to the demand for nutrients [29]. When cells are under normal growth conditions, the target of rapamycin (TOR) kinase complex hyperphosphorylates the ATG13, reducing its affinity for ATG1 (Unc-51-like autophagy activating kinases 1, ULK1 in mammals), thereby inhibiting autophagy. When plants are placed under stress conditions, TOR kinase inactivation dephosphorylates the ATG13, promoting its tight binding to ATG1. Then, ATG1 kinase activates to autophosphorylate itself, resulting in ATG1, ATG11, ATG13, and ATG101 forming a complex and connecting to the phagophore membrane, which promotes several downstream autophagy steps [30]. In mammalian cells, the ULK complex is made up of ULK1/2, ATG13, ATG101, and FIP200/RB1CC1 (Atg11 and Atg17 in yeast) [8]. FIP200/RB1CC1 is the scaffold protein of ULK1/2 and ATG13, which is essential for the phosphorylation and stability of ULK1/2 [31]. In addition to TOR complexes, several pathways, including cAMP-dependent protein kinase A and AMP-activated protein kinase, can also regulate the ATG1/ULK complex [32].

2.2. ATG9/2/18 Transmembrane Complex

The ATG9/2/18 transmembrane complex participates in the formation of phagophore and membrane fusion. ATG9 (ATG9A in mammals) is the only transmembrane protein in the autophagy process of eukaryocytes, which provides a membrane source for forming pre-autophagosomal structure (PAS) and driving the extension of the isolation membrane. It has been found that ATG9 loses contact with the phagophore membrane in the process of autophagy, thus regulating the formation of phagophores from plant ER [33]. ATG2 (ATG2A in mammals) interacts with ATG18 (WD-repeat protein interacting with phosphoinositides, WIPI in mammals), both of which act in the later stages of phagosome formation, fixing the PAS/isolation membrane on the ER and transferring lipids, effectively completing the closing process [34].

2.3. Phosphatidylinositol-3-Kinase (PI3K) Complex

The PI3K complex mediates vesicle nucleation. The PI3K complex can be further divided into complexes I and II. In plants, complex I consists of ATG6, ATG14 (ATG14L in mammals), VPS15, and VPS34, and complex II consists of ATG6, VPS15, VPS34, and VPS38. The PI3K complex and phosphatidylinositol-3-phosphate (PI3P) are responsible for the modification of de novo synthesis of the phagophore, and PI3P recruits the ATG2-ATG18 complex to the phagophore membrane and participates in the extension of the phagophore [35].

2.4. ATG5/ATG12 and ATG8-Phosphatidylethanolamine (PE) Conjugation Systems

The ATG5-ATG12 and ATG8-PE ubiquitin-like conjugating systems not only regulate the initiation of phagophore formation but also function at a downstream step [36]. Among them, ATG4 is a key Cys-protease with important functions for ATG8 (the light chain 3, LC3 in mammals) lipidation and free ATG8 turnover [37]. During the process of autophagy, the C-terminus of ATG8 is first recognized and cleaved by ATG4 protease. Subsequently, ATG7 with ubiquitin-activating enzyme E1 activity binds to ATG8 and ATG12, activating their mature forms. The activated ATG8 and ATG12 transfer to ATG3 and ATG10 with ubiquitin-binding enzyme E2 activity, respectively, and finally link to the substrate. Among them, ubiquitination folding protein ATG8 is combined with PE, and ATG12 is coupled with ATG5 to form a complex. In addition, the ATG12-ATG5 adduct combines with ATG16 (ATG16L1 in mammals) to form oligomeric complexes, which promotes the formation of ATG8-PE adducts [38].

3. Protein Degradation Pathways in Eukaryotic Cells

To ensure that the autophagy machinery can respond in time to each stimulus, regulating the rapid inactivation/degradation of autophagy-related proteins through negative feedback would be an ideal mechanism. In eukaryotes, autophagy and the ubiquitinproteasome system (UPS) maintain proteome homeostasis by degrading redundant or misfolded proteins [39]. Among them, autophagy mainly degrades long-lived proteins and damaged or redundant organelles, and the UPS mainly impairs short-lived proteins [40,41]. Ubiquitination of proteins involves multiple steps. First, ubiquitin-activating enzyme E1 activates the C-terminus of ubiquitin molecules by consuming adenosine triphosphate (ATP) to form an acyl-adenylate complex. Second, the ubiquitin-conjugating enzyme E2 transduces activated ubiquitin molecules through cysteine residues. The ubiquitin ligase E3 then links the ubiquitin molecule to the amino group of the lysine side chain of the target protein by catalyzing the formation of isopeptide bonds. The above steps are repeated to form ubiquitinated target proteins with oligomeric ubiquitin chains in order to form ubiquitination tags. Finally, the proteasome cap recognizes the ubiquitination tag and uses ATP hydrolysis to provide energy to drive ubiquitin molecular excision and target protein unfolding. Then, the unfolded target protein is transferred to the proteasome core for degradation [42].

Ubiquitination-modified proteins can be deubiquitinated by the family of deubiquitinases (DUBs), which hydrolyze the isopeptide bonds formed by target protein lysine residues and ubiquitin molecules, thus inversely regulating the degradation of protein [43]. Thus far, six structurally distinct DUB families have been described, namely, the ubiquitinspecific proteases (USPs), the Josephin family, the ubiquitin C-terminal hydrolases (UCHs), ovarian tumor-related proteases (OTUs), a family of Zn-dependent JAB1/MPN/MOV34 metalloprotease DUBs (JAMMs), and the motif interacting with ubiquitin (MIU)-containing novel DUB family (MINDYs) [44]. Therefore, ubiquitination is a reversible and controllable process. Autophagy and the UPS work together to control intracellular protein quality at the stage of individual development. Plant autophagy-related proteins can be degraded by the UPS pathway [45], while components of the UPS, such as the 26S proteasome, can also be degraded by autophagy [46].

4. Degradation of Yeast Autophagy-Related Proteins

Among yeast autophagy-related proteins, the degradation mechanisms of the transmembrane proteins Atg9 and Atg32 have been discovered. The content of Atg9 in cells is known to be a key factor in determining the number of phagosomes [47]. Under normal growth conditions, Atg9 is ubiquitinated at the Lys113, Lys121, and Lys138 sites and subsequently degraded by the proteasome at peripheral sites away from PAS by Met30, which is one of the substrate recognition components of the ubiquitin ligase SCF complex (Skp, Cullin, F-box containing) (Table 1) [48,49]. When cells are under nutrient-deficient conditions, Atg9 is stabilized by Atg1 phosphorylation, thus inhibiting the degradation process, and activating the autophagy process [50].

In yeast mitophagy, the mitochondrial outer membrane protein Atg32 is the only specific receptor [51], which determines mitochondrial turnover. Overexpression of Atg32 can effectively activate mitophagy, while the deletion of Atg32 leads to a defect in mitophagy [52,53]. Interestingly, two published papers have different findings on the degradation mechanism of Atg32. Levchenko et al. [54] reported that there are two forms of Atg32 in Saccharomyces cerevisiae, namely, unmodified Atg32 and post-translation modified Atg32, and their degradation pathways are different. Post-translationally modified Atg32 is activated when mitophagy is induced to target mitochondria for degradation. Experiments show that this modification is neither phosphorylation nor ubiquitination. In contrast, the turnover of unmodified Atg32 is mediated by an unknown protease independent of proteasome and vacuolar proteases [54]. However, studies by Camougrand et al. [55] demonstrated that the degradation of Atg32 is associated with the UPS. Ubiquitination of Atg32 occurs at least on Lys282 residues and is mediated by several E3 ligases, including Rsp5, which is a relatively complicated process [55]. To sum up, apart from the different strains used in the experimental materials, these two contradictory observations may be attributed to the differences in growth conditions or mitophagy induction, leading to the coordinated regulation of Atg32 levels by multiple pathways. Therefore, further research is needed to determine the details of different degradation pathways of Atg32.

Eukaryotes	ATGs	E3 Ligases	Ubiquitin Chain Types	References
Yeast Mammalian	Atg9	Met30	Unknown	[49]
	Atg32	Rsp5	Unknown	[55]
	Ū	CUL3-KLHL20	K48	[56]
	ULK1	BTRC	K48	[57]
		TRAF3	K48	[58]
		NEDD4L	K27 and K29	[59]
	ATG101	HUWE1	K48	[60]
	WIPI2	HUWE1	Unknown	[61]
	ATG14L	ZBTB16-CUL3-Roc1	Unknown	[62]
		NEDD4	K11	[63]
		RNF216	K48	[64]
	BECN1	CUL3-KLHL20	Unknown	[65]
		CUL3-KLHL38	K48	[65]
Plants		TRAF6	K63	[66]
	VPS34	UBE3C	K29/K48 branched	[67]
		FBXL20-CUL1-SKP1	Unknown	[68]
	ATG16L1	Gigaxonin	K48	[69]
	LC3B	BIRC6	Single ubiquitin	[70]
	ATG3	PTK2	Unknown	[71]
	ATG6	SINAT1/2	Unknown	[45]
	ATG13	SINAT1/2	K48	[72]
	SH3P2	XopL	Unknown	[73]

Table 1. E3 ligases and ubiquitin chain types involved in ATGs degradation in yeast, mammalian, and plants.

5. Degradation of Autophagy-Related Proteins in Metazoa

5.1. Degradation of ULK Complex Components

In mammals, the ULK complex includes four parts: ULK1/2, ATG13, RB1CC1/FIP200, and ATG101. Among them, ULK1 plays a major role in autophagy [74]. The level of ULK1 is regulated by E3 ubiquitin ligase and USPs. Under prolonged starvation, autophosphorylation of ULK1 at the Ser1042/Thr1046 site facilitates its recruitment to Kelch-like (KLHL) 20 for ubiquitination and protein degradation as a substrate for the Cullin3 (Cul3)-KLHL20 ubiquitin ligase, thereby preventing cell unrestricted activate autophagy. In addition, KLHL20 also coordinates the degradation of ATG13 through an indirect mechanism [56]. Cancer can hijack the protective function of autophagy to promote tumorigenesis [75]. In breast cancer cells, mitogen-activated protein kinase (MAPK) 1/3 kinases promote its binding to the E3 ligase BTRC by phosphorylating ULK1 at multiple sites, triggering proteasomal degradation of K48-linked ULK1 ubiquitination, ultimately attenuating mitophagy and promoting breast cancer bone metastases [57]. Tumor necrosis factor receptor-associated factor 3 (TRAF3) is a member of the TRAF family with E3 ligase activity. In macrophages, TRAF3 mediates K48-linked ULK1 ubiquitination and proteasomal degradation, and the TRK-fused gene (TFG)-TRAF3 complex is able to interfere with TRAF3-ULK1 interaction to stabilize ULK1 in response to lipopolysaccharide (LPS)-induced pyroptosis [58]. Neural precursor cell expressed developmentally downregulated 4-like (NEDD4L) is an E3 ubiquitin ligase containing a HECT domain. When cells need to endure starvation, NEDD4L ubiquitinates ULK1 at the Lys925 and Lys933 sites. It induces its degradation by the proteasome, thus autophagy is calibrated to optimal levels to ensure the survival of cells. Interestingly, NEDD4L-mediated ubiquitination of ULK1 is not a canonical K48 linkage, but a K27 and K29 linkage [59].

The degradation of ULK1 was also regulated by deubiquitination. USP20 mediates ULK1 deubiquitination in the basal state, thus interfering with lysosome-dependent degradation of ULK1 and playing a key role in autophagy initiation [76]. ATG101 is a scaffold protein that maintains the stability of ATG13 within the ULK1/ATG1 complex [77,78]. The HECT, UBA, and WWE domain containing E3 ubiquitin-protein ligase 1 (HUWE1) regu-

lates cell proliferation and cell death and is an important factor affecting tumorigenesis [79]. In cancer cells, HUWE1 targets the ubiquitination and degradation of the C-terminal region of ATG101, inhibiting autophagy activity to reduce cancer cell survival [60].

5.2. Degradation of ATG9 Complex Components

The mammalian ATG9 complex contains four ATG18 homologs, namely, WIPI 1-4. Studies have shown that WIPI2 can associate the phagophore with ER [80] and play a role in cellular antibacterial autophagy [81]. Under basal cellular conditions, mTORC1 mediates the phosphorylation of WIPI2 at the Ser395 site, thereby enhancing the specific interaction of WIPI2 with HUWE1, promoting the ubiquitination and proteasomal degradation of WIPI2, and inhibiting the degree of autophagy [61].

5.3. Degradation of PI3K Complex Components

In mammals, the autophagy-specific PI3K complex I consists of four proteins: VPS34, VPS15, BECN1, and ATG14L [82]. Ubiquitination and proteasomal degradation of ATG14L are regulated by the E3 ubiquitin ligase complex zinc finger and BTB domain containing 16 (ZBTB16)-CUL3-Regulator of cullins 1 (Roc1). Under normal nutritional conditions, inhibition of G-protein-coupled receptors (GPCRs) activates glycogen synthase kinase-3β (GSK-3β). It mediates ZBTB16 phosphorylation to promote its autoubiquitination, which in turn upregulates ATG14L levels to activate autophagy [62]. In host cells, *Streptococcus pneumoniae* (Sp) releases choline binding protein C (CbpC) to form the ATG14L-CbpC-sequestosome 1 (SQSTM1)/p62 intracellular complex, promoting autophagy-dependent degradation of ATG14L, which results in inhibiting autophagosome-lysosome fusion and bactericidal autophagic degradation, increasing bacterial survival [83].

BECN1, the first tumor-associated ATG protein discovered in mammals, plays a role in the formation, elongation, and maturation of autophagosomes [84,85]. Different E3 ligases play various roles in regulating BECN1 activity by binding to specific types of ubiquitin chains. NEDD4 and Ring finger protein 216 (RNF216) modify BECN1 through K11, K63 and K48-linked polyubiquitination chains, respectively, which mediates the BECN1 degradation in a proteasome-dependent manner and negatively regulates autophagy [63,64]. Similarly, CUL3-KLHL20-mediated ubiquitination of BECN1 is primarily responsible for the termination of autophagy during prolonged starvation, whereas CUL3-KLHL38-mediated BECN1 K48-linked ubiquitination prevents autophagy under normal conditions [65]. TRAF6 acts as an E3 ligase to trigger the K63-linked ubiquitination of BECN1, while the deubiquitinating enzyme A20 inhibits the ubiquitination of BECN1 [66]. Furthermore, tripartite motif 59 (TRIM59) mediates K48-linked ubiquitination and proteasomal degradation of TRAF6, which results in affecting TRAF6 to ubiquitinate BECN1, regulating the initiation of autophagy [86]. In contrast, three ubiquitin-specific peptidases, USP10, USP13, and USP19, act as positive regulators of autophagy and maintain BECN1 by mediating deubiquitination of BECN1 in the VPS34 complex stability [87,88]. By binding to BECN1, solute carrier family 9 isoform 3 regulator 1 (SLC9A3R1) inhibits BECN1 ubiquitination and proteasomal degradation, which stimulates the formation of the autophagic core lipid kinase complex [89].

VPS34 is the only mammalian phosphatidylinositol 3-kinase class III (PI3KC3) that is activated on phagophore and endosomal membranes [90,91]. The level of VPS34 in cells is controlled by a variety of regulatory factors. For example, the ubiquitin-protein ligase E3C (UBE3C) and the deubiquitinating enzyme TRABID act antagonistically in regulating the levels of VPS34 to balance autophagic activity. UBE3C assembles K29/K48-branched ubiquitin chains on VPS34, enhancing VPS34 binding to the proteasome for degradation. TRABID stabilizes VPS34 by reducing the ubiquitination of VPS34 K29/K48, which in turn promotes phagophore formation [67]. Similarly, auto-ubiquitination of NEDD4-1 enables it to act as a scaffolding protein to recruit USP13 to form the NEDD4-1-USP13 deubiquitination complex. This complex then promotes autophagy by removing K48-linked polyubiquiti-

nation of VPS34 mediated by the E2 enzyme UBC-13/UEV-1 and the E3 enzyme CHN-1 can stabilize the level of VPS34, promoting phagophore maturation [93]. In addition, phosphorylation of substrates can enhance their recognition by SCF complex [94]. Phosphorylation of VPS34 at the Thr159 site promotes F-Box and leucine-rich repeat protein 20 (FBXL20) as an adaptor for the SCF complex, regulating the ubiquitination and proteasomal degradation of VPS34 [68]. Notably, Tang et al. highlighted that the ubiquitination modification of VPS34 is not directly linked to its degradation process [95].

5.4. Degradation of the ATG12-Conjugation System Components

The ATG12-conjugation system includes ATG5, ATG7, ATG10, ATG12, and ATG16L1. ATG12 is a member of the ubiquitin-like protein (UBL) family, which consists of about 20 members [96]. Free ATG12 is highly labile and can be targeted for proteasomal degradation through ubiquitin-dependent or ubiquitin-independent mechanisms [97]. ATG5 is a ubiquitin-like ligase that is conjugated by ATG12 [98]. In Lepidoptera, there is an interaction between ATG1 and ATG5, and *Spodoptera litura* ATG1 (SIATG1) promotes the degradation of ATG5 [99]. In cardiomyocytes, the immunoproteasome β5i subunit interacts with ATG5 to promote the ubiquitination and degradation of ATG5, thereby inhibiting autophagy and leading to cardiac hypertrophy [100]. Moreover, saturated fatty acid palmitate induces ER stress and degrades ATG5 protein through the ER-associated protein degradation (ERAD) pathway, consequently inhibiting autophagy and inducing apoptosis [101].

ATG16L1 interacts with ATG5, which stimulates the ATG8-PE coupling reaction [31]. Gigaxonin (GAN)-E3 ligase ubiquitinates ATG16L1 through K48-type ubiquitin chain polymerization, driving its degradation, thereby controlling the steady-state level of ATG16L1 to ensure fine-tuning of autophagy activation. Loss of GAN leads to the formation of ATG16L1 aggregates that impair phagophore elongation, inhibiting autophagic flux [69].

5.5. Degradation of the LC3-Conjugation System Components

In mammalian cells, LC3, GABA type A receptor-associated proteins (GABARAPs), GATE16, and ATG8L exist as yeast Atg8 homologs [102,103]. All of them can be used as autophagosome markers. Under normal conditions, BRUCE acts as an autophagy inhibitor, promoting proteasomal degradation of LC3-I, which results in reducing LC3-II levels and autophagy [104]. Similarly, UBA6 and BIRC6 act synergistically as E1 and E2/E3 enzymes, respectively, for the monoubiquitination and proteasomal degradation of LC3B, which protects cells from cell death caused by excessive autophagy [70,105]. In addition, circumsporozoite protein (CSP) downregulates LC3B through the UPS pathway, but the mechanism involved is unclear [106].

ATG3 acts as an E2-like enzyme in LC3 lipidation. It can autocatalyze itself to form a complex with ATG12 to promote mitochondrial homeostasis [107,108]. Under DNA damage conditions, protein tyrosine kinase 2 (PTK2) phosphorylates ATG3 at the Tyr203 site and promotes the degradation of ATG3 through the ubiquitin-dependent proteasome pathway, resulting in positively regulating the activity of cancer cells [71].

6. Degradation of Plant Autophagy-Related Proteins

Up to now, more than 40 kinds of ATG proteins and their related regulatory proteins have been identified in plants [109], mainly from the analysis of yeast autophagy-deficient mutants, but their degradation mechanisms are poorly understood. In *Arabidopsis thaliana*, the turnover of ATG1 and ATG13 is strongly and specifically regulated by nutrition, which closely links autophagy with plant nutritional conditions. During the period of fixed carbon/nitrogen limitation, the levels of ATG1 and ATG13 decreased sharply, but they could be reversed again during refeeding. Studies on *A. thaliana* mutants that damage autophagy or 26S proteasome show that both degradation pathways are involved in their degradation, but autophagy is more directly involved. Under the condition of limited nutrition, ATG1 and ATG13 combine with autophagic-like cytolytic structures and finally transfer to vacuoles together with autophagosomes for degradation [110].

In addition, Liu et al. [35] found a similar phenomenon in their study of ATG14 and its associated PI3K complex. By expressing GFP-ATG14b in *atg7* and *atg14a atg14b* mutants, it was found that there was still free GFP in *atg14a atg14b* mutants, which indicated that ATG14 was degraded, unlike the intact GFP-ATG14b fusion in *atg7* mutants with autophagy defects. Next, after N starvation treatment, co-localization of *atg14a atg14b* mutant roots expressing both mCherry-ATG8a and GFP-ATG14b revealed that ATG14 was bound to autophagic membranes [35]. Thus, similar to ATG1 and ATG13, ATG14 is degraded by the autophagic pathway through associating with the autophagic bodies.

The effect of ubiquitination of autophagy-related protein on autophagy has been widely studied in mammals, but less in plant cells. In plants, a single ubiquitin molecule can form a polyubiquitin chain at a certain site in a target protein (polyubiquitination) or be attached to multiple lysine residues (multi-monoubiquitination). In *A. thaliana*, K48 is commonly used to form polyubiquitin chains [111]. Qi et al. [45,72] successfully studied the turnover of ubiquitination modification of ATG6 and ATG13 in *A. Thaliana* (Figure 3). Under different nutritional conditions, as molecular adaptors, *A. thaliana* TRAF1a and TRAF1b interacted with Ring finger E3 ligase seven in absentia of *Arabidopsis thaliana* 1 (SINAT1)/SINAT2 and SINAT6 to regulate the turnover of ATG6 and ATG13. Among them, SINAT6 contains only a short truncated Ring finger domain compared to SINAT1 and SINAT2 [112]. Under nutrient-rich conditions, *A. thaliana* TRAF1a and TRAF1b interact with SINAT1 and SINAT2 to mediate ubiquitination and degradation of ATG6. However, under conditions of nutrient deprivation, starvation-induced accumulation of SINAT6 reduces the binding of SINAT1 and SINAT2 to ATG6, stabilizes ATG6 levels, and activates autophagy [45].

Similarly, at the Lys607 and Lys609 sites of the ATG13 protein, TRAF1-SINAT1/ SINAT2-ATG13 TRAFasome was degraded by ubiquitination modification linked by K48, resulting in the dissociation of ATG1-ATG13 complex to maintain a relatively low autophagy level. However, TRAF1-SINAT6-ATG13 TRAFasome promotes the stability of ATG13, which induces the biogenesis of autophagy in response to nutritional deficiency. Besides, under starvation conditions, ATG1 kinase phosphorylates TRAF1a and promotes its protein stability in vivo, indicating the feedback regulation of autophagy [72]. In short, SINAT1/SINAT2 and SINAT6 play negative and positive roles in regulating the stability of ATG6 and ATG13, thus playing opposite roles in autophagy. Furthermore, at the regulatory level, the turnover of ATG8, which interacts with various adaptor/receptor proteins to recruit specific cargos for degradation, is affected by acyl-coa-binding protein 3 (ACBP3) [113]. In *A. thaliana*, ACBP3, as a phospholipid-binding protein, is involved in the regulation of leaf senescence by regulating membrane phospholipid metabolism and the stability of ATG8. Overexpression of ACBP3 promotes the degradation of ATG8 and disrupts the formation of autophagic vesicles, which results in inhibiting autophagy [114].

Except for maintaining cell homeostasis, autophagy also plays a vital role in plant immunity against pathogens. However, bacteria have evolved the mechanism of evading autophagic clearance in order to better parasitize in the host [115]. Type III effector proteins (T3E) of plant pathogens are present in host cells, and these effectors are capable of manipulating host defense response [116]. Src homology 3 (SH3) domain-containing protein-2 (SH3P2) is a novel membrane-associated protein involved in the formation of autophagosomes [117]. When *Xanthomonas campestris* pv. *vesicatoria* (*Xcv*) invades plant cells; it utilizes the bacterial effector E3 ligase XopL to mediate the ubiquitination and degradation of SH3P2 in a proteasome-dependent manner, and inhibits host autophagy, thus enhancing the virulence of *Xanthomonas*. Intriguingly, in host cells, XopL is recognized and degraded by NBR1/Joka2-mediated selective autophagy related to defense. Hence, the mutual targeting of pathogen effector XopL and plant protein SH3P2 reveals the complex antagonism between pathogen and plant autophagy mechanism [73].



Autophagosome formation

Figure 3. Degradation mechanisms of ATG6 and ATG13 in plants. During nutrient-rich conditions, the TRAF1s-SINAT1/SINAT2-ATG6 and TRAF1s-SINAT1/SINAT2-ATG13 TRAFasomes regulate the ubiquitination and proteasomal degradation of ATG6 and ATG13, which results in inhibiting autophagy. Under the condition of nutrient starvation, SINAT6 accumulates to form the TRAF1s-SINAT6-ATG6 and TRAF1s-SINAT6-ATG13 TRAFasomes, which maintain the stability of ATG6 and ATG13. Furthermore, ATG1 kinase phosphorylates TRAF1s to increase its stability.

7. Concluding Remarks and Future Perspectives

Autophagy is an important regulatory factor for eukaryotic cells to cope with various stresses. In plants, it is precisely regulated by environmental changes and developmental stages [118]. In this review, the degradation mechanism of autophagy-related proteins in eukaryotes has been discussed. There are many studies on the degradation mechanism of animal autophagy-related proteins, probably because the regulation of autophagy can be used as an effective intervention for the treatment of metabolic and neurodegenerative diseases [119,120]. However, there is little information about how autophagy-related proteins are degraded in plant cells, and only the degradation mechanisms of ATG1, ATG6, ATG13, ATG14, and SH3P2 have been discovered thus far [45,72,110].

As far as the known degradation mechanism of animal autophagy-related proteins is concerned, ubiquitin-proteasome mainly plays a role in it. What is more, some regulatory factors that affect the degradation of autophagy-related proteins were also found. Both autophagy and the UPS are involved in the degradation of the plant ATG1-ATG13 complex. Then, how plants themselves coordinate the effects of these two pathways on autophagy still needs to be further explored. Whenever plants are under conditions that are not conducive to their own growth, such as drought, salt stress, and virus invasion, autophagy will be activated, thereby improving plant resistance [121–123]. Under suboptimal growth conditions, autophagy is beneficial for improving crop yield [124], and the turnover of ATG proteins plays a decisive role in the level of autophagy. Therefore, the degradation mechanism of plant autophagy-related proteins needs to be studied more extensively in order that they can make a beneficial contribution to agricultural production.

Author Contributions: The authors confirm their contributions to this work: Y.Z. in conceptualization, original draft preparation, and writing; and H.M., W.H., and F.L. in revising and review. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the National Natural Science Foundation of China (No. 32100297 to Fen Liu).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Because of space limitations, we apologize to the authors whose works are not cited.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Sarika, C.; Sreedevi, P.; Mishra, S.; Singh, S.N. Multifaceted Housekeeping Functions of Autophagy. J. Indian Inst. Sci. 2017, 97, 79–94.
- 2. Huang, H.; Ayaz, A.; Zheng, M.; Yang, X.; Zaman, W.; Zhao, H.; Lü, S. Arabidopsis KCS5 and KCS6 Play Redundant Roles in Wax Synthesis. *Int. J. Mol. Sci.* 2022, 23, 4450. [CrossRef]
- 3. van Doorn, W.G.; Papini, A. Ultrastructure of autophagy in plant cells: A review. *Autophagy* **2013**, *9*, 1922–1936. [CrossRef] [PubMed]
- 4. Bolliet, V.; Labonne, J.; Olazcuaga, L.; Panserat, S.; Seiliez, I. Modeling of autophagy-related gene expression dynamics during long term fasting in European eel (*Anguilla anguilla*). Sci. Rep. **2017**, *7*, 17896. [CrossRef] [PubMed]
- 5. Lamark, T.; Johansen, T. Mechanisms of selective autophagy. Annu. Rev. Cell Dev. Biol. 2021, 37, 143–169. [CrossRef]
- 6. Li, F.; Chung, T.; Vierstra, R.D. AUTOPHAGY-RELATED11 plays a critical role in general autophagy- and senescence-induced mitophagy in Arabidopsis. *Plant Cell* **2014**, *26*, 788–807. [CrossRef]
- 7. Marion, J.; Le Bars, R.; Besse, L.; Batoko, H.; Satiat-Jeunemaitre, B. Multiscale and Multimodal Approaches to Study Autophagy in Model Plants. *Cells* **2018**, *7*, 5. [CrossRef]
- 8. Nakatogawa, H. Mechanisms governing autophagosome biogenesis. Nat. Rev. Mol. Cell Biol. 2020, 21, 439–458. [CrossRef]
- 9. Sieńko, K.; Poormassalehgoo, A.; Yamada, K.; Goto-Yamada, S. Microautophagy in Plants: Consideration of Its Molecular Mechanism. *Cells* **2020**, *9*, 887. [CrossRef]
- 10. Liaquat, F.; Munis, M.F.H.; Arif, S.; Haroon, U.; Shi, J.; Saqib, S.; Zaman, W.; Che, S.; Liu, Q. PacBio single-molecule long-read sequencing reveals genes tolerating manganese stress in Schima superba saplings. *Front. Genet.* **2021**, *12*, 635043. [CrossRef]
- Ayaz, A.; Huang, H.; Zheng, M.; Zaman, W.; Li, D.; Saqib, S.; Zhao, H.; Lü, S. Molecular cloning and functional analysis of GmLACS2-3 reveals its involvement in cutin and suberin biosynthesis along with abiotic stress tolerance. *Int. J. Mol. Sci.* 2021, 22, 9175. [CrossRef] [PubMed]
- 12. Manghwar, H.; Hussain, A.; Ali, Q.; Liu, F. Brassinosteroids (BRs) role in plant development and coping with different stresses. *Int. J. Mol. Sci.* **2022**, 23, 1012. [CrossRef] [PubMed]
- Manghwar, H.; Hussain, A.; Ullah, A.; Gul, S.; Shaban, M.; Khan, A.H.; Ali, M.; Sani, S.G.A.S.; Chaudhary, H.J.; Munis, M.F.H. Expression analysis of defense related genes in wheat and maize against Bipolaris sorokiniana. *Physiol. Mol. Plant Pathol.* 2018, 103, 36–46. [CrossRef]
- Ali, Q.; Yu, C.; Hussain, A.; Ali, M.; Ahmar, S.; Sohail, M.A.; Riaz, M.; Ashraf, M.F.; Abdalmegeed, D.; Wang, X. Genome engineering technology for durable disease resistance: Recent progress and future outlooks for sustainable agriculture. *Front. Plant Sci.* 2022, 13, 860281. [CrossRef]
- Ali, Q.; Zheng, H.; Rao, M.J.; Ali, M.; Hussain, A.; Saleem, M.H.; Nehela, Y.; Sohail, M.A.; Ahmed, A.M.; Kubar, K.A. Advances, limitations, and prospects of biosensing technology for detecting phytopathogenic bacteria. *Chemosphere* 2022, 296, 133773. [CrossRef]
- Liaquat, F.; Qunlu, L.; Arif, S.; Haroon, U.; Saqib, S.; Zaman, W.; Jianxin, S.; Shengquan, C.; Li, L.X.; Akbar, M. Isolation and characterization of pathogen causing brown rot in lemon and its control by using ecofriendly botanicals. *Physiol. Mol. Plant Pathol.* 2021, 114, 101639. [CrossRef]
- 17. Senft, D.; Ronai, Z.A. UPR, autophagy, and mitochondria crosstalk underlies the ER stress response. *Trends Biochem. Sci.* **2015**, 40, 141–148. [CrossRef]
- Manghwar, H.; Li, J. Endoplasmic Reticulum Stress and Unfolded Protein Response Signaling in Plants. Int. J. Mol. Sci. 2022, 23, 828. [CrossRef]
- 19. Manghwar, H.; Hussain, A. Mechanism of tobacco osmotin gene in plant responses to biotic and abiotic stress tolerance: A brief history. *Biocell* **2022**, *46*, 623. [CrossRef]
- 20. Yim, W.W.-Y.; Mizushima, N. Lysosome biology in autophagy. Cell Discov. 2020, 6, 6. [CrossRef]
- Yang, M.; Bu, F.; Huang, W.; Chen, L. Multiple Regulatory Levels Shape Autophagy Activity in Plants. *Front. Plant Sci.* 2019, 10, 532. [CrossRef] [PubMed]
- Yang, C.; Luo, M.; Zhuang, X.; Li, F.; Gao, C. Transcriptional and epigenetic regulation of autophagy in plants. *Trends Genet.* 2020, 36, 676–688. [CrossRef] [PubMed]
- Ayaz, A.; Saqib, S.; Huang, H.; Zaman, W.; Lü, S.; Zhao, H. Genome-wide comparative analysis of long-chain acyl-CoA synthetases (LACSs) gene family: A focus on identification, evolution and expression profiling related to lipid synthesis. *Plant Physiol. Biochem.* 2021, 161, 1–11. [CrossRef] [PubMed]
- 24. Wang, Y.; Cai, S.; Yin, L.; Shi, K.; Xia, X.; Zhou, Y.; Yu, J.; Zhou, J. Tomato HsfA1a plays a critical role in plant drought tolerance by activating ATG genes and inducing autophagy. *Autophagy* **2015**, *11*, 2033–2047. [CrossRef]
- 25. Feng, Y.; Yao, Z.; Klionsky, D.J. How to control self-digestion: Transcriptional, post-transcriptional, and post-translational regulation of autophagy. *Trends Cell Biol.* **2015**, *25*, 354–363. [CrossRef]
- Zhang, D.; Wang, W.; Sun, X.; Xu, D.; Wang, C.; Zhang, Q.; Wang, H.; Luo, W.; Chen, Y.; Chen, H.; et al. AMPK regulates autophagy by phosphorylating BECN1 at threonine 388. *Autophagy* 2016, 12, 1447–1459. [CrossRef]
- Laureano-Marín, A.M.; Aroca, Á.; Pérez-Pérez, M.E.; Yruela, I.; Jurado-Flores, A.; Moreno, I.; Crespo, J.L.; Romero, L.C.; Gotor, C. Abscisic Acid-Triggered Persulfidation of the Cys Protease ATG4 Mediates Regulation of Autophagy by Sulfide. *Plant Cell* 2020, 32, 3902–3920. [CrossRef]
- Zientara-Rytter, K.; Sirko, A. To deliver or to degrade-an interplay of the ubiquitin-proteasome system, autophagy and vesicular transport in plants. FEBS J. 2016, 283, 3534–3555. [CrossRef]
- 29. Nanji, T.; Liu, X.; Chew, L.H.; Li, F.K.; Biswas, M.; Yu, Z.Q.; Lu, S.; Dong, M.Q.; Du, L.L.; Klionsky, D.J.; et al. Conserved and unique features of the fission yeast core Atg1 complex. *Autophagy* **2017**, *13*, 2018–2027. [CrossRef]
- Marshall, R.S.; Vierstra, R.D. Autophagy: The Master of Bulk and Selective Recycling. Annu. Rev. Plant Biol. 2018, 69, 173–208. [CrossRef]
- 31. Rehman, N.U.; Zeng, P.; Mo, Z.; Guo, S.; Liu, Y.; Huang, Y.; Xie, Q. Conserved and Diversified Mechanism of Autophagy between Plants and Animals upon Various Stresses. *Antioxidants* **2021**, *10*, 1736. [CrossRef] [PubMed]
- 32. Budovskaya, Y.V.; Stephan, J.S.; Reggiori, F.; Klionsky, D.J.; Herman, P.K. The Ras/cAMP-dependent protein kinase signaling pathway regulates an early step of the autophagy process in Saccharomyces cerevisiae. *J. Biol. Chem.* **2004**, *279*, 20663–20671. [CrossRef] [PubMed]
- 33. Zhuang, X.; Chung, K.P.; Cui, Y.; Lin, W.; Gao, C.; Kang, B.H.; Jiang, L. ATG9 regulates autophagosome progression from the endoplasmic reticulum in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E426–E435. [CrossRef] [PubMed]
- Nair, U.; Cao, Y.; Xie, Z.; Klionsky, D.J. Roles of the lipid-binding motifs of Atg18 and Atg21 in the cytoplasm to vacuole targeting pathway and autophagy. J. Biol. Chem. 2010, 285, 11476–11488. [CrossRef] [PubMed]
- Liu, F.; Hu, W.; Li, F.; Marshall, R.S.; Zarza, X.; Munnik, T.; Vierstra, R.D. AUTOPHAGY-RELATED14 and Its Associated Phosphatidylinositol 3-Kinase Complex Promote Autophagy in Arabidopsis. *Plant Cell* 2020, *32*, 3939–3960. [CrossRef] [PubMed]
 Minuching M. The ATC and institution at tanking Complex Promote Cell Pick 2020, *1*, 10 [CrossRef] [PubMed]
- 36. Mizushima, N. The ATG conjugation systems in autophagy. Curr. Opin. Cell Biol. 2020, 63, 1–10. [CrossRef]
- 37. Pérez-Pérez, M.E.; Lemaire, S.D.; Crespo, J.L. The ATG4 protease integrates redox and stress signals to regulate autophagy. *J. Exp. Bot.* **2021**, *72*, 3340–3351. [CrossRef]
- 38. Chung, T.; Phillips, A.R.; Vierstra, R.D. ATG8 lipidation and ATG8-mediated autophagy in Arabidopsis require ATG12 expressed from the differentially controlled ATG12A AND ATG12B loci. *Plant J.* **2010**, *62*, 483–493. [CrossRef]
- 39. Balchin, D.; Hayer-Hartl, M.; Hartl, F.U. In vivo aspects of protein folding and quality control. *Science* **2016**, *353*, aac4354. [CrossRef]
- 40. Rubinsztein, D.C. The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature* **2006**, *443*, 780–786. [CrossRef]
- 41. Nedelsky, N.B.; Todd, P.K.; Taylor, J.P. Autophagy and the ubiquitin-proteasome system: Collaborators in neuroprotection. *Biochim. Biophys. Acta* **2008**, 1782, 691–699. [CrossRef] [PubMed]
- 42. Kwon, Y.T.; Ciechanover, A. The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy. *Trends Biochem. Sci.* 2017, 42, 873–886. [CrossRef] [PubMed]
- 43. Komander, D. Mechanism, specificity and structure of the deubiquitinases. *Subcell. Biochem.* 2010, 54, 69–87. [PubMed]
- 44. Mevissen, T.E.T.; Komander, D. Mechanisms of Deubiquitinase Specificity and Regulation. *Annu. Rev. Biochem.* **2017**, *86*, 159–192. [CrossRef]

- Qi, H.; Xia, F.N.; Xie, L.J.; Yu, L.J.; Chen, Q.F.; Zhuang, X.H.; Wang, Q.; Li, F.; Jiang, L.; Xie, Q.; et al. TRAF Family Proteins Regulate Autophagy Dynamics by Modulating AUTOPHAGY PROTEIN6 Stability in Arabidopsis. *Plant Cell* 2017, 29, 890–911. [CrossRef]
- 46. Su, T.; Yang, M.; Wang, P.; Zhao, Y.; Ma, C. Interplay between the Ubiquitin Proteasome System and Ubiquitin-Mediated Autophagy in Plants. *Cells* **2020**, *9*, 2219. [CrossRef]
- 47. Jin, M.; He, D.; Backues, S.K.; Freeberg, M.A.; Liu, X.; Kim, J.K.; Klionsky, D.J. Transcriptional regulation by Pho23 modulates the frequency of autophagosome formation. *Curr. Biol.* **2014**, *24*, 1314–1322. [CrossRef]
- 48. Smothers, D.B.; Kozubowski, L.; Dixon, C.; Goebl, M.G.; Mathias, N. The abundance of Met30p limits SCF(Met30p) complex activity and is regulated by methionine availability. *Mol. Cell Biol.* 2000, 20, 7845–7852. [CrossRef]
- 49. Feng, Y.; Ariosa, A.R.; Yang, Y.; Hu, Z.; Dengjel, J.; Klionsky, D.J. Downregulation of autophagy by Met30-mediated Atg9 ubiquitination. *Proc. Natl. Acad. Sci. USA* 2021, *118*, e2005539118. [CrossRef]
- 50. Hu, G.; Rios, L.; Yan, Z.; Jasper, A.M.; Luera, D.; Luo, S.; Rao, H. Autophagy regulator Atg9 is degraded by the proteasome. *Biochem. Biophys. Res. Commun.* **2020**, *522*, 254–258. [CrossRef]
- 51. Kanki, T.; Wang, K.; Cao, Y.; Baba, M.; Klionsky, D.J. Atg32 is a mitochondrial protein that confers selectivity during mitophagy. *Dev. Cell* **2009**, *17*, 98–109. [CrossRef] [PubMed]
- Liu, L.; Feng, D.; Chen, G.; Chen, M.; Zheng, Q.; Song, P.; Ma, Q.; Zhu, C.; Wang, R.; Qi, W.; et al. Mitochondrial outer-membrane protein FUNDC1 mediates hypoxia-induced mitophagy in mammalian cells. *Nat. Cell Biol.* 2012, 14, 177–185. [CrossRef] [PubMed]
- 53. Wu, W.; Tian, W.; Hu, Z.; Chen, G.; Huang, L.; Li, W.; Zhang, X.; Xue, P.; Zhou, C.; Liu, L.; et al. ULK1 translocates to mitochondria and phosphorylates FUNDC1 to regulate mitophagy. *EMBO Rep.* **2014**, *15*, 566–575. [CrossRef] [PubMed]
- 54. Levchenko, M.; Lorenzi, I.; Dudek, J. The Degradation Pathway of the Mitophagy Receptor Atg32 Is Re-Routed by a Posttranslational Modification. *PLoS ONE* **2016**, *11*, e0168518. [CrossRef] [PubMed]
- 55. Camougrand, N.; Vigié, P.; Gonzalez, C.; Manon, S.; Bhatia-Kiššová, I. The yeast mitophagy receptor Atg32 is ubiquitinated and degraded by the proteasome. *PLoS ONE* **2020**, *15*, e0241576. [CrossRef]
- Liu, C.C.; Lin, Y.C.; Chen, Y.H.; Chen, C.M.; Pang, L.Y.; Chen, H.A.; Wu, P.R.; Lin, M.Y.; Jiang, S.T.; Tsai, T.F.; et al. Cul3-KLHL20 Ubiquitin Ligase Governs the Turnover of ULK1 and VPS34 Complexes to Control Autophagy Termination. *Mol. Cell* 2016, 61, 84–97. [CrossRef]
- 57. Deng, R.; Zhang, H.L.; Huang, J.H.; Cai, R.Z.; Wang, Y.; Chen, Y.H.; Hu, B.X.; Ye, Z.P.; Li, Z.L.; Mai, J.; et al. MAPK1/3 kinasedependent ULK1 degradation attenuates mitophagy and promotes breast cancer bone metastasis. *Autophagy* **2021**, *17*, 3011–3029. [CrossRef]
- Shi, J.H.; Ling, C.; Wang, T.T.; Zhang, L.N.; Liu, W.W.; Qin, Y.; Tan, Y.H.; Cui, N.P.; Ni, Z.Y. TRK-fused gene (TFG) regulates ULK1 stability via TRAF3-mediated ubiquitination and protects macrophages from LPS-induced pyroptosis. *Cell Death Dis.* 2022, 13, 93. [CrossRef]
- Nazio, F.; Carinci, M.; Valacca, C.; Bielli, P.; Strappazzon, F.; Antonioli, M.; Ciccosanti, F.; Rodolfo, C.; Campello, S.; Fimia, G.M.; et al. Fine-tuning of ULK1 mRNA and protein levels is required for autophagy oscillation. J. Cell Biol. 2016, 215, 841–856. [CrossRef]
- 60. Lee, J.; Kim, J.; Shin, J.; Kang, Y.; Choi, J.; Cheong, H. ATG101 Degradation by HUWE1-Mediated Ubiquitination Impairs Autophagy and Reduces Survival in Cancer Cells. *Int. J. Mol. Sci.* **2021**, *22*, 9182. [CrossRef]
- 61. Wan, W.; You, Z.; Zhou, L.; Xu, Y.; Peng, C.; Zhou, T.; Yi, C.; Shi, Y.; Liu, W. mTORC1-Regulated and HUWE1-Mediated WIPI2 Degradation Controls Autophagy Flux. *Mol. Cell* **2018**, *72*, 303–315.e6. [CrossRef]
- Zhang, T.; Dong, K.; Liang, W.; Xu, D.; Xia, H.; Geng, J.; Najafov, A.; Liu, M.; Li, Y.; Han, X.; et al. G-protein-coupled receptors regulate autophagy by ZBTB16-mediated ubiquitination and proteasomal degradation of Atg14L. *eLife* 2015, 4, e06734. [CrossRef] [PubMed]
- 63. Platta, H.W.; Abrahamsen, H.; Thoresen, S.B.; Stenmark, H. Nedd4-dependent lysine-11-linked polyubiquitination of the tumour suppressor Beclin 1. *Biochem. J.* 2012, 441, 399–406. [CrossRef] [PubMed]
- 64. Xu, C.; Feng, K.; Zhao, X.; Huang, S.; Cheng, Y.; Qian, L.; Wang, Y.; Sun, H.; Jin, M.; Chuang, T.H.; et al. Regulation of autophagy by E3 ubiquitin ligase RNF216 through BECN1 ubiquitination. *Autophagy* **2014**, *10*, 2239–2250. [CrossRef] [PubMed]
- 65. Li, X.; Yang, K.B.; Chen, W.; Mai, J.; Wu, X.Q.; Sun, T.; Wu, R.Y.; Jiao, L.; Li, D.D.; Ji, J.; et al. CUL3 (cullin 3)-mediated ubiquitination and degradation of BECN1 (beclin 1) inhibit autophagy and promote tumor progression. *Autophagy* **2021**, *17*, 4323–4340. [CrossRef] [PubMed]
- 66. Shi, C.S.; Kehrl, J.H. TRAF6 and A20 regulate lysine 63-linked ubiquitination of Beclin-1 to control TLR4-induced autophagy. *Sci. Signal.* **2010**, *3*, ra42. [CrossRef] [PubMed]
- Chen, Y.H.; Huang, T.Y.; Lin, Y.T.; Lin, S.Y.; Li, W.H.; Hsiao, H.J.; Yan, R.L.; Tang, H.W.; Shen, Z.Q.; Chen, G.C.; et al. VPS34 K29/K48 branched ubiquitination governed by UBE3C and TRABID regulates autophagy, proteostasis and liver metabolism. *Nat. Commun.* 2021, 12, 1322. [CrossRef] [PubMed]
- 68. Xiao, J.; Zhang, T.; Xu, D.; Wang, H.; Cai, Y.; Jin, T.; Liu, M.; Jin, M.; Wu, K.; Yuan, J. FBXL20-mediated Vps34 ubiquitination as a p53 controlled checkpoint in regulating autophagy and receptor degradation. *Genes Dev.* **2015**, *29*, 184–196. [CrossRef] [PubMed]
- 69. Scrivo, A.; Codogno, P.; Bomont, P. Gigaxonin E3 ligase governs ATG16L1 turnover to control autophagosome production. *Nat. Commun.* **2019**, *10*, 780. [CrossRef]

- 70. Jia, R.; Bonifacino, J.S. Negative regulation of autophagy by UBA6-BIRC6-mediated ubiquitination of LC3. *eLife* **2019**, *8*, e50034. [CrossRef]
- 71. Ma, K.; Fu, W.; Tang, M.; Zhang, C.; Hou, T.; Li, R.; Lu, X.; Wang, Y.; Zhou, J.; Li, X.; et al. PTK2-mediated degradation of ATG3 impedes cancer cells susceptible to DNA damage treatment. *Autophagy* **2017**, *13*, 579–591. [CrossRef] [PubMed]
- 72. Qi, H.; Li, J.; Xia, F.N.; Chen, J.Y.; Lei, X.; Han, M.Q.; Xie, L.J.; Zhou, Q.M.; Xiao, S. Arabidopsis SINAT Proteins Control Autophagy by Mediating Ubiquitylation and Degradation of ATG13. *Plant Cell* **2020**, *32*, 263–284. [CrossRef] [PubMed]
- Leong, J.X.; Raffeiner, M.; Spinti, D.; Langin, G.; Franz-Wachtel, M.; Guzman, A.R.; Kim, J.G.; Pandey, P.; Minina, A.E.; Macek, B.; et al. A bacterial effector counteracts host autophagy by promoting degradation of an autophagy component. *EMBO J.* 2022, e110352. [CrossRef]
- Hosokawa, N.; Hara, T.; Kaizuka, T.; Kishi, C.; Takamura, A.; Miura, Y.; Iemura, S.; Natsume, T.; Takehana, K.; Yamada, N.; et al. Nutrient-dependent mTORC1 association with the ULK1-Atg13-FIP200 complex required for autophagy. *Mol. Biol. Cell* 2009, 20, 1981–1991. [CrossRef]
- 75. Mizushima, N.; White, E.; Rubinsztein, D.C. Breakthroughs and bottlenecks in autophagy research. *Trends Mol. Med.* **2021**, 27, 835–838. [CrossRef]
- 76. Kim, J.H.; Seo, D.; Kim, S.J.; Choi, D.W.; Park, J.S.; Ha, J.; Choi, J.; Lee, J.H.; Jung, S.M.; Seo, K.W.; et al. The deubiquitinating enzyme USP20 stabilizes ULK1 and promotes autophagy initiation. *EMBO Rep.* **2018**, *19*, e44378. [CrossRef]
- 77. Hosokawa, N.; Sasaki, T.; Iemura, S.; Natsume, T.; Hara, T.; Mizushima, N. Atg101, a novel mammalian autophagy protein interacting with Atg13. *Autophagy* **2009**, *5*, 973–979. [CrossRef]
- 78. Mercer, C.A.; Kaliappan, A.; Dennis, P.B. A novel, hu...uman Atg13 binding protein, Atg101, interacts with ULK1 and is essential for macroautophagy. *Autophagy* **2009**, *5*, 649–662. [CrossRef]
- 79. Kao, S.H.; Wu, H.T.; Wu, K.J. Ubiquitination by HUWE1 in tumorigenesis and beyond. J. Biomed. Sci. 2018, 25, 67. [CrossRef] [PubMed]
- Zhao, Y.G.; Chen, Y.; Miao, G.; Zhao, H.; Qu, W.; Li, D.; Wang, Z.; Liu, N.; Li, L.; Chen, S.; et al. The ER-Localized Transmembrane Protein EPG-3/VMP1 Regulates SERCA Activity to Control ER-Isolation Membrane Contacts for Autophagosome Formation. *Mol. Cell* 2017, 67, 974–989.e6. [CrossRef]
- Thurston, T.L.; Boyle, K.B.; Allen, M.; Ravenhill, B.J.; Karpiyevich, M.; Bloor, S.; Kaul, A.; Noad, J.; Foeglein, A.; Matthews, S.A.; et al. Recruitment of TBK1 to cytosol-invading Salmonella induces WIPI2-dependent antibacterial autophagy. *EMBO J.* 2016, 35, 1779–1792. [CrossRef] [PubMed]
- 82. Dikic, I.; Elazar, Z. Mechanism and medical implications of mammalian autophagy. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 349–364. [CrossRef] [PubMed]
- 83. Shizukuishi, S.; Ogawa, M.; Ryo, A.; Ohnishi, M. Streptococcus pneumoniae promotes its own survival via choline-binding protein CbpC-mediated degradation of ATG14. *Autophagy* 2020, *16*, 1529–1531. [CrossRef] [PubMed]
- 84. Kang, R.; Zeh, H.J.; Lotze, M.T.; Tang, D. The Beclin 1 network regulates autophagy and apoptosis. *Cell Death Differ.* 2011, 18, 571–580. [CrossRef]
- Funderburk, S.F.; Wang, Q.J.; Yue, Z. The Beclin 1-VPS34 complex–at the crossroads of autophagy and beyond. *Trends Cell Biol.* 2010, 20, 355–362. [CrossRef]
- 86. Han, T.; Guo, M.; Gan, M.; Yu, B.; Tian, X.; Wang, J.B. TRIM59 regulates autophagy through modulating both the transcription and the ubiquitination of BECN1. *Autophagy* **2018**, *14*, 2035–2048. [CrossRef]
- Liu, J.; Xia, H.; Kim, M.; Xu, L.; Li, Y.; Zhang, L.; Cai, Y.; Norberg, H.V.; Zhang, T.; Furuya, T.; et al. Beclin1 controls the levels of p53 by regulating the deubiquitination activity of USP10 and USP13. *Cell* 2011, 147, 223–234. [CrossRef]
- 88. Jin, S.; Tian, S.; Chen, Y.; Zhang, C.; Xie, W.; Xia, X.; Cui, J.; Wang, R.F. USP19 modulates autophagy and antiviral immune responses by deubiquitinating Beclin-1. *EMBO J.* **2016**, *35*, 866–880. [CrossRef]
- Liu, H.; Ma, Y.; He, H.W.; Wang, J.P.; Jiang, J.D.; Shao, R.G. SLC9A3R1 stimulates autophagy via BECN1 stabilization in breast cancer cells. *Autophagy* 2015, 11, 2323–2334. [CrossRef]
- 90. Hurley, J.H.; Young, L.N. Mechanisms of Autophagy Initiation. Annu. Rev. Biochem. 2017, 86, 225–244. [CrossRef]
- 91. Mizushima, N. A brief history of autophagy from cell biology to physiology and disease. *Nat. Cell Biol.* **2018**, *20*, 521–527. [CrossRef] [PubMed]
- 92. Xie, W.; Jin, S.; Wu, Y.; Xian, H.; Tian, S.; Liu, D.A.; Guo, Z.; Cui, J. Auto-ubiquitination of NEDD4-1 Recruits USP13 to Facilitate Autophagy through Deubiquitinating VPS34. *Cell Rep.* 2020, *30*, 2807–2819.e4. [CrossRef] [PubMed]
- 93. Liu, J.; Li, M.; Li, L.; Chen, S.; Wang, X. Ubiquitination of the PI3-kinase VPS-34 promotes VPS-34 stability and phagosome maturation. *J. Cell Biol.* **2018**, 217, 347–360. [CrossRef] [PubMed]
- Skaar, J.R.; Pagan, J.K.; Pagano, M. Mechanisms and function of substrate recruitment by F-box proteins. *Nat. Rev. Mol. Cell Biol.* 2013, 14, 369–381. [CrossRef]
- 95. Tang, J.; Sun, F.; Deng, X.J.; Ma, Y.L.; Li, K.Y.; Tang, Y.; Chen, G.J. Ubiquitination status does not affect Vps34 degradation. *Genes Dis.* 2020, 7, 401–407. [CrossRef]
- 96. van der Veen, A.G.; Ploegh, H.L. Ubiquitin-like proteins. Annu. Rev. Biochem. 2012, 81, 323–357. [CrossRef]
- 97. Haller, M.; Hock, A.K.; Giampazolias, E.; Oberst, A.; Green, D.R.; Debnath, J.; Ryan, K.M.; Vousden, K.H.; Tait, S.W. Ubiquitination and proteasomal degradation of ATG12 regulates its proapoptotic activity. *Autophagy* **2014**, *10*, 2269–2278. [CrossRef]
- 98. Li, X.; He, S.; Ma, B. Autophagy and autophagy-related proteins in cancer. *Mol. Cancer* **2020**, *19*, 12. [CrossRef]

- Zhang, N.; Yang, Y.; Lu, H.; Xiang, Y.; Huang, X.; Hu, R.; Chen, Z.; Yuan, W.; Peng, R.; Peng, J.; et al. Spodoptera litura autophagy-related protein 1 interacts with autophagy-related protein 5 and enhances its degradation. *Insect Mol. Biol.* 2017, 26, 190–203. [CrossRef]
- 100. Xie, X.; Bi, H.L.; Lai, S.; Zhang, Y.L.; Li, N.; Cao, H.J.; Han, L.; Wang, H.X.; Li, H.H. The immunoproteasome catalytic β5i subunit regulates cardiac hypertrophy by targeting the autophagy protein ATG5 for degradation. *Sci. Adv.* **2019**, *5*, eaau0495. [CrossRef]
- 101. Mallik, A.; Yammani, R.R. Saturated fatty acid palmitate negatively regulates autophagy by promoting ATG5 protein degradation in meniscus cells. *Biophys. Res. Commun.* **2018**, *502*, 370–374. [CrossRef] [PubMed]
- 102. Kabeya, Y.; Mizushima, N.; Yamamoto, A.; Oshitani-Okamoto, S.; Ohsumi, Y.; Yoshimori, T. LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *J. Cell Sci.* 2004, *117 Pt* 13, 2805–2812. [CrossRef] [PubMed]
- 103. Tanida, I.; Sou, Y.S.; Minematsu-Ikeguchi, N.; Ueno, T.; Kominami, E. Atg8L/Apg8L is the fourth mammalian modifier of mammalian Atg8 conjugation mediated by human Atg4B, Atg7 and Atg3. FEBS J. 2006, 273, 2553–2562. [CrossRef] [PubMed]
- 104. Jiang, T.X.; Zou, J.B.; Zhu, Q.Q.; Liu, C.H.; Wang, G.F.; Du, T.T.; Luo, Z.Y.; Guo, F.; Zhou, L.M.; Liu, J.J.; et al. SIP/CacyBP promotes autophagy by regulating levels of BRUCE/Apollon, which stimulates LC3-I degradation. *Proc. Natl. Acad. Sci. USA* 2019, 116, 13404–13413. [CrossRef] [PubMed]
- 105. Liu, Y.; Levine, B. Autosis and autophagic cell death: The dark side of autophagy. Cell Death Differ. 2015, 22, 367–376. [CrossRef]
- 106. Lu, X.; Zhang, J.; Li, Y.Q.; Liu, Q.X.; Zhou, D.; Deng, X.F.; Qiu, Y.; Chen, Q.; Li, M.Y.; Liu, X.Q.; et al. Plasmodium Circumsporozoite Protein Enhances the Efficacy of Gefitinib in Lung Adenocarcinoma Cells by Inhibiting Autophagy via Proteasomal Degradation of LC3B. Front. Cell Dev. Biol. 2022, 10, 830046. [CrossRef]
- 107. Besteiro, S.; Brooks, C.F.; Striepen, B.; Dubremetz, J.F. Autophagy protein Atg3 is essential for maintaining mitochondrial integrity and for normal intracellular development of Toxoplasma gondii tachyzoites. *PLoS Pathog.* **2011**, *7*, e1002416. [CrossRef]
- 108. Metlagel, Z.; Otomo, C.; Takaesu, G.; Otomo, T. Structural basis of ATG3 recognition by the autophagic ubiquitin-like protein ATG12. *Proc. Natl. Acad. Sci. USA* 2013, *110*, 18844–18849. [CrossRef]
- 109. Liu, Y.; Bassham, D.C. Autophagy: Pathways for self-eating in plant cells. Annu. Rev. Plant Biol. 2012, 63, 215–237. [CrossRef]
- 110. Suttangkakul, A.; Li, F.; Chung, T.; Vierstra, R.D. The ATG1/ATG13 protein kinase complex is both a regulator and a target of autophagic recycling in Arabidopsis. *Plant Cell* **2011**, *23*, 3761–3779. [CrossRef]
- 111. Kim, D.Y.; Scalf, M.; Smith, L.M.; Vierstra, R.D. Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. *Plant Cell* **2013**, *25*, 1523–1540. [CrossRef] [PubMed]
- 112. Bao, Y.; Wang, C.; Jiang, C.; Pan, J.; Zhang, G.; Liu, H.; Zhang, H. The tumor necrosis factor receptor-associated factor (TRAF)-like family protein SEVEN IN ABSENTIA 2 (SINA2) promotes drought tolerance in an ABA-dependent manner in Arabidopsis. *New Phytol.* **2014**, 202, 174–187. [CrossRef] [PubMed]
- 113. Bu, F.; Yang, M.; Guo, X.; Huang, W.; Chen, L. Multiple Functions of ATG8 Family Proteins in Plant Autophagy. *Front. Cell Dev. Biol.* 2020, *8*, 466. [CrossRef] [PubMed]
- 114. Xiao, S.; Gao, W.; Chen, Q.F.; Chan, S.W.; Zheng, S.X.; Ma, J.; Wang, M.; Welti, R.; Chye, M.L. Overexpression of Arabidopsis acyl-CoA binding protein ACBP3 promotes starvation-induced and age-dependent leaf senescence. *Plant Cell* 2010, 22, 1463–1482. [CrossRef]
- 115. Leary, A.Y.; Sanguankiattichai, N.; Duggan, C.; Tumtas, Y.; Pandey, P.; Segretin, M.E.; Linares, J.S.; Savage, Z.D.; Yow, R.J.; Bozkurt, T.O. Modulation of plant autophagy during pathogen attack. *J. Exp. Bot.* **2018**, *69*, 1325–1333. [CrossRef]
- 116. Khan, M.; Seto, D.; Subramaniam, R.; Desveaux, D. Oh, the places they'll go! A survey of phytopathogen effectors and their host targets. *Plant J.* **2018**, *93*, 651–663. [CrossRef]
- 117. Zhuang, X.; Jiang, L. Autophagosome biogenesis in plants: Roles of SH3P2. Autophagy 2014, 10, 704–705. [CrossRef]
- 118. Wang, P.; Wang, T.; Han, J.; Li, M.; Zhao, Y.; Su, T.; Ma, C. Plant Autophagy: An Intricate Process Controlled by Various Signaling Pathways. *Front. Plant Sci.* 2021, 12, 754982. [CrossRef]
- 119. Rabinowitz, J.D.; White, E. Autophagy and metabolism. Science 2010, 330, 1344–1348. [CrossRef]
- 120. Nixon, R.A. The role of autophagy in neurodegenerative disease. Nat. Med. 2013, 19, 983–997. [CrossRef]
- 121. Sedaghatmehr, M.; Thirumalaikumar, V.P.; Kamranfar, I.; Marmagne, A.; Masclaux-Daubresse, C.; Balazadeh, S. A regulatory role of autophagy for resetting the memory of heat stress in plants. *Plant Cell Environ.* **2019**, *42*, 1054–1064. [CrossRef] [PubMed]
- 122. Luo, L.; Zhang, P.; Zhu, R.; Fu, J.; Su, J.; Zheng, J.; Wang, Z.; Wang, D.; Gong, Q. Autophagy Is Rapidly Induced by Salt Stress and Is Required for Salt Tolerance in Arabidopsis. *Front. Plant Sci.* **2017**, *8*, 1459. [CrossRef]
- 123. Deretic, V. Autophagy as an innate immunity paradigm: Expanding the scope and repertoire of pattern recognition receptors. *Curr. Opin. Immunol.* **2012**, *24*, 21–31. [CrossRef] [PubMed]
- 124. Li, F.; Chung, T.; Pennington, J.G.; Federico, M.L.; Kaeppler, H.F.; Kaeppler, S.M.; Otegui, M.S.; Vierstra, R.D. Autophagic recycling plays a central role in maize nitrogen remobilization. *Plant Cell* **2015**, *27*, 1389–1408. [CrossRef] [PubMed]

MDPI St. Alban-Anlage 66 4052 Basel Switzerland Tel. +41 61 683 77 34 Fax +41 61 302 89 18 www.mdpi.com

International Journal of Molecular Sciences Editorial Office E-mail: ijms@mdpi.com www.mdpi.com/journal/ijms



MDPI St. Alban-Anlage 66 4052 Basel Switzerland

Tel: +41 61 683 77 34 Fax: +41 61 302 89 18

www.mdpi.com



ISBN 978-3-0365-6996-3