pharmaceuticals

Special Issue Reprint

Computational Methods in Drug Development

Edited by
Yunjie Zhao

mdpi.com/journal/pharmaceuticals

MDPI

Computational Methods in Drug Development

Computational Methods in Drug Development

Guest Editor

Yunjie Zhao

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editor
Yunjie Zhao
Department of Physics
Central China
Normal University
Wuhan
China

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Pharmaceuticals* (ISSN 1424-8247), freely accessible at: https://www.mdpi.com/journal/pharmaceuticals/special_issues/Y0VSZ4C2YT.

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

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

ISBN 978-3-7258-9001-9 (Hbk)
ISBN 978-3-7258-9002-6 (PDF)
https://doi.org/10.3390/books978-3-7258-9002-6

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

About the Editor . . . . . **vii**

Yunjie Zhao
Computational Methods in Drug Development: From Molecular Modeling to Artificial Intelligence Technologies
Reprinted from: *Pharmaceuticals* **2026**, *19*, 1266, https://doi.org/10.3390/ph19081266 . . . . . **1**

Vera Malheiro, Beatriz Santos, Ana Figueiras and Filipa Mascarenhas-Melo
The Potential of Artificial Intelligence in Pharmaceutical Innovation: From Drug Discovery to Clinical Trials
Reprinted from: *Pharmaceuticals* **2025**, *18*, 788, https://doi.org/10.3390/ph18060788 . . . . . **5**

Jiaming Gao, Chen Zhuo, Chengwei Zeng, Haoquan Liu and Yunjie Zhao
Advances in Methods for Accurate Prediction of RNA–Small Molecule Binding Sites: From Isolated to AI-Integrated Strategies
Reprinted from: *Pharmaceuticals* **2025**, *18*, 1593, https://doi.org/10.3390/ph18101593 . . . . . **35**

Jose I. Bueso-Bordils, Gerardo M. Antón-Fos, Rafael Martín-Algarra and Pedro A. Alemán-López
Pharmacokinetic Equations Applied to Obtain New Topological Models in the Search of Antibacterial Compounds
Reprinted from: *Pharmaceuticals* **2025**, *18*, 865, https://doi.org/10.3390/ph18060865 . . . . . **63**

Shangbo Ning, Chengwei Zeng, Huiwen Wang, Junfeng Zhang, Yun Xue and Yunjie Zhao
The Allosteric Regulation of the DNA-Binding Domain of p53 by the Intrinsically Disordered C-Terminal Domain
Reprinted from: *Pharmaceuticals* **2026**, *19*, 124, https://doi.org/10.3390/ph19010124 . . . . . **78**

Antonija Jelčić, Anamarija Raspudić, Danijela Barić, Ana Ratković, Ivana Šagud, Paula Pongrac, et al.
Charged Thienobenzo-1,2,3-Triazoles as Especially Potent Non-Selective Cholinesterase Inhibitors: Design, Anti-Inflammatory Activity, and Computational Study
Reprinted from: *Pharmaceuticals* **2025**, *18*, 1032, https://doi.org/10.3390/ph18071032 . . . . . **96**

Garcia Ferreira de Souza, Airis Farias Santana, Fernanda Sanches Kuhl Antunes, Ramon Martins Cogo, Matheus Dornellas Pereira, Daniela Gonçales Galasse Rando and Carolina Passarelli Gonçalves
Exploring Brazilian Green Propolis Phytochemicals in the Search for Potential Inhibitors of B-Raf^{600}E Enzyme: A Theoretical Approach
Reprinted from: *Pharmaceuticals* **2025**, *18*, 902, https://doi.org/10.3390/ph18060902 . . . . . **113**

Shafeek Buhlak, Nadeem Abad, Jihane Akachar, Sana Saffour, Yunus Kesgun, Sevval Dik, et al.
Design, Synthesis, and Computational Evaluation of 3,4-Dihydroquinolin-2(*1H*)-One Analogues as Potential VEGFR2 Inhibitors in Glioblastoma Multiforme
Reprinted from: *Pharmaceuticals* **2025**, *18*, 233, https://doi.org/10.3390/ph18020233 . . . . . **135**

Sudhir Ghandikota and Anil G Jegga
From Gene Networks to Therapeutics: A Causal Inference and Deep Learning Approach for Drug Discovery
Reprinted from: *Pharmaceuticals* **2025**, *18*, 1304, https://doi.org/10.3390/ph18091304 . . . . . **167**

Bianca Fagan Bissacotti, Marcylene Vieira da Silveira, Charles Elias Assmann, Priscila Marquezan Copetti, André Flores dos Santos, Solange Binotto Fagan, et al.
Resveratrol Alleviates Inflammatory Response Through P2X7/NLRP3 Signaling P athway: In Silico and In Vitro Evidence from Activated Microglia
Reprinted from: *Pharmaceuticals* **2025**, *18*, 950, https://doi.org/10.3390/ph18070950 . . . . . **183**

Pranab Dev Sharma, Abdulrahman Mohammed Alhudhaibi, Abdullah Al Noman, Emad M. Abdallah, Tarek H. Taha and Himanshu Sharma
Systems Biology-Driven Discovery of Host-Targeted Therapeutics for Oropouche Virus: Integrating Network Pharmacology, Molecular Docking, and Drug Repurposing
Reprinted from: *Pharmaceuticals* **2025**, *18*, 613, https://doi.org/10.3390/ph18050613 . . . . . **212**

About the Editor

Yunjie Zhao

Yunjie Zhao, a Professor at the School of Physics Science and Technology at Central China Normal University, is also a doctoral supervisor and Deputy Dean. His research centers on condensed matter, soft matter physics, and biophysics. He has developed models for soft matter, including nucleic acid complex interaction prediction, structural module analysis, tertiary structure prediction, and targeted drug design, uncovering the physical mechanisms behind the structure–function–disease relationships in nucleic acid complexes. Zhao has authored over 70 SCI papers in journals like *Nature Immunology* and *Nature Communications*, and key publications such as '*Biomolecular Big Data Analysis*' and ' *Fundamentals of Soft Matter and Biomolecular Physics*,' which are expected to be top-selling physics books in 2025. He holds three authorized national invention patents and has been invited as a Guest Editor for journals like *Briefings in Bioinformatics*. Additionally, he reviews for more than 20 international journals, including *Nature Biotechnology*, *Nature Reviews Genetics*, *Nature Machine Intelligence*, *Nature Communications*, and *Physical Review Letters*.

pharmaceuticals

MDPI

Editorial

Computational Methods in Drug Development: From Molecular Modeling to Artificial Intelligence Technologies

Yunjie Zhao

Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan 430079, China; yjzhaowh@ccnu.edu.cn

Advances in computational methods have significantly impacted protein design and drug development by offering innovative ways to accelerate the discovery and optimization of therapeutics [1,2]. Recently, molecular modeling and artificial intelligence have advanced in their ability to predict the structures of proteins, RNA, and drug molecules, as well as their complexes; identify new drug targets; and optimize drug–receptor interactions. Computational approaches are effective alternatives because experimental methods for determining molecular structures can be expensive and time-consuming. They also enhance our collective understanding of complex biological systems, enable high-throughput screening of large chemical libraries, and help to reduce time and costs compared to traditional methods. This Special Issue, "Computational Methods in Drug Development," highlights these roles of computation in current pharmaceutical research and covers four key areas: understanding molecular mechanisms, drug discovery techniques, disease-focused therapeutic development, and future research directions.

Accurately understanding the structure and function of complex systems through experimental methods alone remains challenging [3]. Gao et al. explored techniques for predicting RNA–small-molecule binding sites [4]. Their review offers a timely summary of recent progress, systematically illustrating the history from physics-based strategies to integrated AI techniques. Traditional methods rely on limited data, whereas recent AI developments enable the integration of diverse features, leading to better predictions. The review explains the core principles behind different features, compares how features behave at binding versus non-binding sites, assesses method performance across different feature sets and test datasets, and highlights ongoing opportunities and challenges in flexibility prediction [5]. The p53 tumor suppressor protein contains intrinsically disordered N- and C-terminal domains. Ning et al. used microsecond-scale molecular dynamics simulations to investigate how the CTD and DNA binding regulate the allosteric mechanism of the p53 DNA-binding domain (DBD) [6]. Firstly, DNA binding induces conformational changes in the DBD, particularly in the S6-S7 loop and L1. Subsequently, the CTD directly interacts with DNA and influences the allosteric network, facilitating DBD binding to DNA. The allosteric mechanisms outlined here offer new insights into the functional roles of the p53 CTD and can guide the rational design of p53-targeting drugs. While a compound might show strong predicted binding, it can still fail due to unfavorable absorption, distribution, residence time, or clearance. Bueso-Bordils et al. [7] used pharmacokinetic equations to develop topological models to discover antibacterial compounds. This approach brings developability factors into the initial stages of computational screening, which is beneficial because the top-ranked compound based on target affinity is not always the best candidate for further development.

The process of developing inhibitors and prioritizing candidates involves complex computational methods. Jelčić et al. examined charged thienobenzo-1,2,3-triazoles as non-selective cholinesterase inhibitors, integrating chemical design, synthesis, anti-inflammatory testing, and computational analysis [8]. Their research emphasizes the promise of novel charged triazolinium salts as selective cholinesterase inhibitors with extra anti-inflammatory effects. De Souza et al. investigated the phytochemicals of Brazilian green propolis as potential B-Raf600E enzyme inhibitors using physicochemical analysis, ADME prediction, docking, molecular dynamics, and free-energy calculations [9]. Their work demonstrates how computational tools can help prioritize complex natural compounds for further testing. Buhlak et al. developed and evaluated 3,4-dihydroquinolin-2(1H)-one analogs as potential VEGFR2 inhibitors for glioblastoma therapy. Their approach included molecular docking and dynamics studies to analyze how these compounds interact with critical residues in the VEGFR2 binding pocket [10]. Overall, they show that AI and large-scale modeling are most effective when integrated into workflows that also consider chemical feasibility, target biology, and validation processes. AI improves candidate selection by aiding target identification, screening, optimization, repurposing, and clinical decision-making. Malheiro et al. reviewed AI's potential in pharmaceutical innovation, emphasizing that it is not a simple solution but a tool that accelerates the search and prioritization processes [11]. AI has revolutionized drug discovery and development by enabling rapid, efficient analysis of extensive biological and chemical data, helping to identify new therapeutic compounds from initial research through clinical trials. However, there are still gaps in AI deployment, primarily due to a lack of regulation.

Scientists have focused on disease-centered therapeutic discovery, recognizing that disease biology is complex and seldom involves just one molecular interaction. Therapeutic outcomes are influenced by gene networks, pathway compensation, cellular states, and patient heterogeneity. Ghandikota and Jegga introduce a new computational framework that combines network analysis, statistical mediation, and deep learning to discover causal target genes and potential small-molecule drug candidates for repurposing [12]. Incorporating causal reasoning, their framework seeks to move beyond mere correlations toward testable hypotheses. Their study illustrates how molecular signatures associated with disease can guide target and drug prioritization. Likewise, in inflammatory disease models, Bissacotti et al. examined resveratrol's anti-inflammatory effects via the P2X7/NLRP3 pathway, using in silico and in vitro evidence from activated microglia [13]. Docking analyses highlight potential molecular interactions, which cellular assays then test against biological responses. This iterative modeling and experimental process offers deeper insight than computational predictions alone, especially for compounds with multiple biological effects. Sharma et al. used a systems biology strategy that included protein–protein interaction (PPI) networks, molecular docking, and drug repurposing to identify host-targeted treatments for Oropouche virus [14]. These strategies are useful when time or resources limit direct antiviral development, but they must be carefully validated to prevent harmful disruptions to host pathways. Overall, computational drug development is viewed as a cohesive, integrated process of reasoning. Docking is most effective when combined with molecular dynamics, free-energy calculations, chemical synthesis, and experimental validation. Machine learning predictions improve when they utilize interpretable features, biological mechanisms, or validation data. Network pharmacology becomes more relevant when integrated with disease context and causal inference. Meanwhile, pharmacokinetic and developability models help to avoid early prioritization based solely on target affinity. Furthermore, RNA-binding site prediction and allosteric simulations expand the types of systems suitable for computational analysis [15,16].

https://doi.org/10.3390/ph19081266

Although progress has been achieved, benchmark quality for AI and machine learning models trained on retrospective data remains a concern. Challenges such as data leakage, biased negative samples, a narrow chemical scope, and inconsistent experimental annotations can result in overly optimistic performance evaluations. Interpretability remains challenging. Models that prioritize compounds, targets, or pathways without clarifying their reasoning can be helpful for preliminary screening but are less effective for fine-tuning based on mechanisms. Biological systems function across various levels, with molecular binding being just one aspect of the therapeutic response. Outcomes are also affected by factors like tissue exposure, metabolism, immune environment, compensatory signaling, and patient variability. As a result, validation should be incorporated into the computational workflow rather than viewed as an endpoint. Future advancements are expected to include physics-informed machine learning models that integrate molecular graphs, sequences, 3D structures, omics data, and pharmacology [17]. This Special Issue emphasizes that computation complements experiments rather than replacing them, supporting the systematic discovery process. By connecting molecular mechanisms, inhibitor design, disease contexts, and validation, computational methods help researchers to develop more accurate hypotheses and make more informed decisions.

Funding: Hubei Science Fund for Distinguished Young Scholars (grant no. 2024AFA077), the Fundamental Research Funds for the Central China Normal Universities (2026AI+), Academy of Frontier Interdisciplinary Research, Central China Normal University (JC2026PT-002).

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Yang, W.; Wang, S.Z.; Lee, G.R.; Zhang, J.Z.; Courbet, A.; Juergens, D.; Wang, X.R.; Schlichthaerle, T.; Abedi, M.; Ragotte, R.; et al. The past, present and future of de novo protein design. *Nature* **2026**, *652*, 1139–1152. [CrossRef] [PubMed]
2. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. [CrossRef] [PubMed]
3. Zeng, C.W.; Jian, Y.R.; Vosoughi, S.; Zeng, C.; Zhao, Y.J. Evaluating native-like structures of RNA-protein complexes through the deep learning method. *Nat. Commun.* **2023**, *14*, 1060. [CrossRef] [PubMed]
4. Gao, J.; Zhuo, C.; Zeng, C.; Liu, H.; Zhao, Y. Advances in Methods for Accurate Prediction of RNA-Small Molecule Binding Sites: From Isolated to AI-Integrated Strategies. *Pharmaceuticals* **2025**, *18*, 1593. [CrossRef] [PubMed]
5. Zeng, C.; Gao, J.; Liu, H.; Zhuo, C.; Zhao, Y. Exploring ligand flexibility in nucleic acid scaffolds using graph neural networks. *Biophys. J.* 2026, *Online ahead of print*.
6. Ning, S.; Zeng, C.; Wang, H.; Zhang, J.; Xue, Y.; Zhao, Y. The Allosteric Regulation of the DNA-Binding Domain of p53 by the Intrinsically Disordered C-Terminal Domain. *Pharmaceuticals* **2026**, *19*, 124. [CrossRef] [PubMed]
7. Bueso-Bordils, J.I.; Antón-Fos, G.M.; Martín-Algarra, R.; Alemán-López, P.A. Pharmacokinetic Equations Applied to Obtain New Topological Models in the Search of Antibacterial Compounds. *Pharmaceuticals* **2025**, *18*, 865. [CrossRef] [PubMed]
8. Jelčić, A.; Raspudić, A.; Barić, D.; Ratković, A.; Šagud, I.; Pongrac, P.; Štefok, D.; Bosnar, M.; Roca, S.; Lasić, Z.; et al. Charged Thienobenzo-1,2,3-Triazoles as Especially Potent Non-Selective Cholinesterase Inhibitors: Design, Anti-Inflammatory Activity, and Computational Study. *Pharmaceuticals* **2025**, *18*, 1032. [CrossRef] [PubMed]
9. de Souza, G.F.; Santana, A.F.; Antunes, F.S.K.; Cogo, R.M.; Pereira, M.D.; Rando, D.G.G.; Gonçalves, C.P. Exploring Brazilian Green Propolis Phytochemicals in the Search for Potential Inhibitors of B-Raf(600)E Enzyme: A Theoretical Approach. *Pharmaceuticals* **2025**, *18*, 902. [CrossRef] [PubMed]
10. Buhlak, S.; Abad, N.; Akachar, J.; Saffour, S.; Kesgun, Y.; Dik, S.; Yasin, B.; Bati-Ayaz, G.; Hanashalshahaby, E.; Türkez, H.; et al. Design, Synthesis, and Computational Evaluation of 3,4-Dihydroquinolin-2(1H)-One Analogues as Potential VEGFR2 Inhibitors in Glioblastoma Multiforme. *Pharmaceuticals* **2025**, *18*, 233. [CrossRef] [PubMed]
11. Malheiro, V.; Santos, B.; Figueiras, A.; Mascarenhas-Melo, F. The Potential of Artificial Intelligence in Pharmaceutical Innovation: From Drug Discovery to Clinical Trials. *Pharmaceuticals* **2025**, *18*, 788. [CrossRef] [PubMed]
12. Ghandikota, S.; Jegga, A.G. From Gene Networks to Therapeutics: A Causal Inference and Deep Learning Approach for Drug Discovery. *Pharmaceuticals* **2025**, *18*, 1304. [CrossRef] [PubMed]

https://doi.org/10.3390/ph19081266

13. Bissacotti, B.F.; da Silveira, M.V.; Assmann, C.E.; Copetti, P.M.; Santos, A.F.D.; Fagan, S.B.; da Rocha, J.A.P.; Schetinger, M.R.C.; Morsch, V.M.M.; Bottari, N.B.; et al. Resveratrol Alleviates Inflammatory Response Through P2X7/NLRP3 Signaling Pathway: In Silico and In Vitro Evidence from Activated Microglia. *Pharmaceuticals* **2025**, *18*, 950. [CrossRef] [PubMed]
14. Dev Sharma, P.; Alhudhaibi, A.M.; Al Noman, A.; Abdallah, E.M.; Taha, T.H.; Sharma, H. Systems Biology-Driven Discovery of Host-Targeted Therapeutics for Oropouche Virus: Integrating Network Pharmacology, Molecular Docking, and Drug Repurposing. *Pharmaceuticals* **2025**, *18*, 613. [CrossRef] [PubMed]
15. Liu, H.Q.; Jian, Y.R.; Zeng, C.; Zhao, Y.J. RNA-protein interaction prediction using network-guided deep learning. *Commun. Biol.* **2025**, *8*, 247. [CrossRef] [PubMed]
16. Liu, H.Q.; Jian, Y.R.; Hou, J.X.; Zeng, C.; Zhao, Y.J. RNet: A network strategy to predict RNA binding preferences. *Brief. Bioinform.* **2024**, *25*, bbad482. [CrossRef] [PubMed]
17. Liu, H.Q.; Zhu, Y.N.; Gao, J.M.; Zhuo, C.; Zeng, C.W.; Zhao, Y.J. RNA regulation mechanisms study using physics-informed machine learning. *Commun. Phys.* **2025**, *9*, 19. [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.

pharmaceuticals

MDPI

Review

The Potential of Artificial Intelligence in Pharmaceutical Innovation: From Drug Discovery to Clinical Trials

Vera Malheiro [1,†], Beatriz Santos [1,†], Ana Figueiras [1,2,*] and Filipa Mascarenhas-Melo [3,4,*]

[1] Laboratory of Drug Development and Technologies, Faculty of Pharmacy, University of Coimbra, 3000-548 Coimbra, Portugal; veracatarina.am13@gmail.com (V.M.); beatrizcsantos.trabalho@gmail.com (B.S.)
[2] REQUIMTE/LAQV, Group of Pharmaceutical Technology, University of Coimbra, 3000-548 Coimbra, Portugal
[3] Higher School of Health, Polytechnic Institute of Guarda, Rua da Cadeia, 6300-307 Guarda, Portugal
[4] BRIDGES—Biotechnology Research, Innovation and Design for Health Products, Polytechnic University of Guarda, Avenida Dr. Francisco Sá Carneiro, n.° 50, 6300-559 Guarda, Portugal
* Correspondence: rfigueiras@ff.uc.pt (A.F.); filipamelo@ff.uc.pt (F.M.-M.)
† These authors contributed equally to this work.

Abstract: Artificial intelligence (AI) is a subfield of computer science focused on developing systems that can execute tasks traditionally associated with human intelligence. AI systems work through algorithms based on rules or instructions that enable the machine to make decisions. With the advancement of science, more sophisticated AI techniques, such as machine learning and deep learning, have been developed, allowing machines to learn from large amounts of data and improve their performance over time. The pharmaceutical industry has greatly benefited from the development of this technology. AI has revolutionized drug discovery and development by enabling rapid and effective analysis of vast volumes of biological and chemical data during the identification of new therapeutic compounds. The algorithms developed can predict the efficacy, toxicity, and possible adverse effects of new drugs, optimize the steps involved in clinical trials, reduce associated time and costs, and facilitate the implementation of innovative drugs in the market, making it easier to develop precise therapies tailored to the individual genetic profile of patients. Despite significant advancements, there are still gaps in the application of AI, particularly due to the lack of comprehensive regulation. The constant evolution of this technology requires ongoing and in-depth legislative oversight to ensure its use remains safe, ethical, and free from bias. This review explores the role of AI in drug development, assessing its potential to enhance formulation, accelerate discovery, and repurpose existing medications. It highlights AI's impact across all stages, from initial research to clinical trials, emphasizing its ability to optimize processes, drive innovation, and improve therapeutic outcomes.

Keywords: artificial intelligence; machine and deep learning; drug discovery and development; drug repurposing; clinical trials; regulatory issues

1. Introduction

The development of medicines has always been a long, costly process with a low success rate [1]. In recent years, there has been significant progress in the pharmaceutical industry, requiring the adoption of more efficient and systematic drug production practices [1]. Drug discovery is driven by the fact that there are no treatments available for some diseases, or existing drugs have low efficacy or high toxicity [1]. To address these challenges, advancements

in technology have provided a wide range of techniques widely used to accelerate all stages of drug development, which have been radically transformed by the rapid progress of AI [2].

AI was described in the 1950s as a science and engineering field capable of developing intelligent machines. It quickly evolved into a neural network model similar to the human brain, enabling the execution of tasks requiring human intelligence, such as solving complex problems and making complex decisions. AI has evolved exponentially and is becoming a highly important and revolutionary component in many areas, particularly in drug discovery [2,3].

The exponential growth of AI has raised several ethical issues. The ability of these technologies to make autonomous decisions raises concerns about human dignity and inherent values. For this reason, it has become necessary to establish ethical guidelines to prevent significant consequences for human life. Ethics is a practical necessity to ensure that technologies are used in a fair and responsible way [4].

In the drug development process, AI is not limited to discovering new drugs. It is also characterized by the optimization of existing treatments tailored to the individual profile of each patient. This approach improves treatment efficacy, minimizes side effects, and provides safer and more effective therapeutic solutions [5,6].

In this context, the main objective of this review is to evaluate the potential of AI in drug formulation and how this technology can effectively contribute to the discovery and development of new medications, as well as the adaptation of existing medications to other diseases, thus providing a safer and more effective therapeutic experience.

2. Artificial Intelligence

Artificial intelligence (AI) is increasingly integrated into pharmaceutical research, revolutionizing processes across drug discovery, development, and clinical decision-making. AI encompasses technologies that simulate human intelligence to perform tasks such as learning, reasoning, and pattern recognition [7]. AI is an interdisciplinary field combining computer science, statistics, and domain-specific knowledge. It is often described as a set of technologies, processes, and methods that enable machines to perform cognitive functions, such as reasoning and decision-making [8]. In the context of pharmaceutical research, AI models can accelerate compound screening, predict molecular interactions, and optimize clinical trial designs. While traditional definitions of AI cover a wide scope, from narrow to general intelligence, this work emphasizes those tools and frameworks with demonstrated relevance to biomedical and pharmaceutical applications [9].

Talking about AI inevitably involves referencing the concepts of machine learning (ML) and deep learning (DL). While these concepts are related, they are not interchangeable. AI focuses on intellectual activities typically performed by humans, such as solving well-defined logical problems, whereas ML and DL are specific methods that help achieve this, such as voice recognition or image classification [10]. In other words, both ML and DL operate within the broader scope of AI (Figure 1).

2.1. From the Beginning of AI to the Present Day

Charles Babbage became one of the leading mathematicians of his time and is often referred to as the "father of the computer" for creating the first modern analytical machine, which he described as the first thinking machine. Later, he conceived the analytical machine with the intention of developing precise mathematical tables [7]. Subsequently, Alan Turing was recognized as the "father of computation and AI" when, in 1950, he published the article *Computers and Intelligence*, in which he described the ability of computers to exhibit intelligence similar to human intelligence, an idea that became known as the "Turing Test" [3]. This test is based on an "imitation game", which assesses a person's ability to

distinguish between a conversation with a human and one with a machine [11]. Following Turing, the field became more complex, requiring the exploration of new challenges, which contributed to advancements in computer science and mathematics. This journey leading to the modern AI we know today is concisely outlined in Table 1.

Table 1. AI progress: major breakthroughs over time.

Year	Progress Information	References
1955	Allen Newell, J.C. Shaw, and Herbert A. Simon created a program called the "Logic Theorist", the first program capable of performing automated reasoning—the first AI software in operation. In the following year, the program was demonstrated for the first time in Pittsburgh.	[7]
1956	John McCarthy first introduced the term AI at a conference on the subject. Later, in 1958, McCarthy developed the programming language *Lisp*, which became fundamental to the advancement of computer science and AI. Alongside Marvin Minsky, in 1959, he founded the AI laboratories.	[7,12]
1958	Arthur Samuel coined the term "Machine Learning" when he wrote about teaching a machine to play chess better than the humans who created it.	[7]
1965	Edward Feigenbaum initiated *DENDRAL*, the first "expert system", capable of extracting molecular structures from organic compounds using scientific data.	[7]
1966	Joseph Weizenbaum developed the first "chatterbot", *ELIZA*, a computer program designed to simulate a fictional psychotherapist. ELIZA used natural language processing to communicate with humans.	[7]
1979	The first American Association for AI was founded, currently known as the Association for the Advancement of AI (AAAI).	[7]
1980s	The beginning of this decade became known as the "AI Boom", as rapid growth and interest in AI emerged. John Hopfield and David Rumelhart introduced deep learning techniques, enabling computers to learn from experiences. Edward Feigenbaum developed specialized systems to imitate human decision-making processes. These systems were widely adopted in industry. The year 1984 became known as the "AI winter" because consumer, public, and private sectors began to show little interest in AI, leading to a decline in funding and research.	[7]
1995	Stuart Russell and Peter Norvig published *"AI: A Modern Approach"*, which became one of the key textbooks in the field. Richard Wallace developed the chatbot *ALICE*, which established more natural conversations with users.	[7]
1997	IBM's specialized chess-playing system, *Deep Blue*, defeated Garry Kasparov, the reigning world chess champion and grandmaster, in a historic chess match. Dragon Systems developed voice recognition software, which was implemented in Windows.	[7,13]
2000	Cynthia Breazeal developed *Kismet*, the first robot capable of simulating human emotions, mimicking human facial expressions.	[7]
2002	iRobot introduced the Roomba, the first robotic vacuum cleaner that revolutionized the way homes were cleaned. Equipped with intelligent sensors, the Roomba navigates autonomously, detecting and avoiding obstacles, making cleaning easier and more efficient.	[7]
2010	Microsoft introduced Kinect for the Xbox 360, the first gaming device capable of tracking body movements and converting them into commands within the game.	[7]
2011	Apple launched Siri, the first virtual assistant.	[7]
2014	Amazon created the virtual assistant Alexa, allowing users to interact with electronic devices using voice commands.	[7]
2016	Hanson Robotics developed Sophia, the first humanoid robot with a realistic appearance, capable of communicating, seeing, and replicating human emotions.	[7]
2020	DeepMind released an updated version of AlphaFold, a program capable of deciphering the three-dimensional structure of any protein.	[7]
2022	OpenAI launched the chatbot *ChatGPT*	[7]

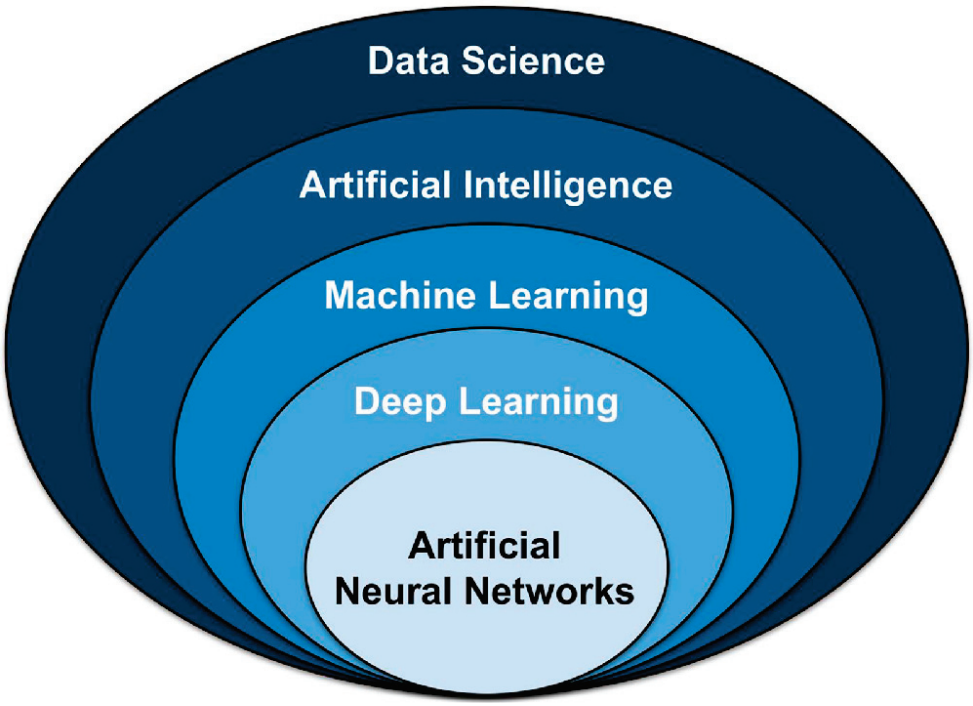


Figure 1. Overview of key data science techniques. AI (AI) is a subset of data science that encompasses both classical programming and machine learning (ML). Within ML, various models and approaches exist, including deep learning (DL) and artificial neural networks (ANNs). Adapted from [14].

2.2. Machine Learning

Machine learning is recognized as a subset of AI, involving algorithms with the ability to define their own rules based on input data without explicit programming. It primarily uses two methods for data analysis [10]. In the supervised method, the system is programmed to recognize specific patterns previously discovered, meaning the system has access to the desired responses, facilitating the process of classification and grouping ideas. On the other hand, the unsupervised method detects an unknown set of data and categorizes it into specific combinations independently [11].

It can also be divided into another method, known as reinforcement learning. This is a trial-and-error approach, driven mainly by decision-making within a specific environment and the execution of actions to maximize performance. The goal of this method is to guide situations through its actions [12,14].

Common AI Models in Pharmaceutical Applications

Recent years have seen a marked increase in the use of machine learning models to tackle complex problems in drug development, toxicology, and personalized medicine. Among the machine learning techniques applied in the pharmaceutical field, several models stand out due to their versatility and performance. One such model is random forest, an ensemble method that builds multiple decision trees and combines their outputs to improve prediction accuracy and minimize overfitting. This approach has proven effective in tasks such as classifying toxicity profiles and identifying potential biomarkers in preclinical research [15].

Another widely adopted model is the artificial neural network (ANN), inspired by the structure of the human brain. ANNs consist of layers of interconnected nodes (or "neurons") capable of recognizing complex patterns in large datasets. In pharmaceutical research, they have been successfully used in areas ranging from molecular property prediction to clinical

decision support and image-based drug screening [16]. Recent advances in generative models, particularly Generative Adversarial Networks (GANs), have introduced new possibilities in drug design. By leveraging a dual-network system—comprising a generator and a discriminator—GANs are able to produce novel molecular structures that meet predefined pharmacological criteria, thus accelerating early-stage drug discovery [17,18].

In parallel, Transformer architectures, originally developed for natural language processing, are now being repurposed to handle biomedical data. Their self-attention mechanism allows them to process large-scale biological sequences, predict drug–target interactions, and extract insights from unstructured clinical data with remarkable precision. These models are increasingly being integrated into platforms aimed at personalized medicine and advanced therapeutic development [19].

2.3. Deep Learning

Deep learning is seen as an extension of the ML method, as while the latter relies on more traditional approaches to reach its answers, DL is more complex and consists of a set of algorithms aimed at mimicking the capabilities of the human brain [11,13]. It builds on the evolution of the artificial neural network (ANN)—a set of multilayered units of nonlinear processing used for data representation, which includes sophisticated computational elements connected to one another, resembling "perceptions" comparable to human neurons. These systems replicate the transmission of electrical signals observed in the human brain [11]. These models are particularly effective in image analysis (e.g., histopathology, radiology), drug–target interaction prediction, and modeling protein structures.

2.4. Ethical Concerns in the Use of AI in Healthcare

The integration of artificial intelligence (AI) into pharmaceutical research and healthcare systems introduces significant ethical considerations, especially in areas such as data privacy, algorithmic bias, transparency, and human oversight. AI technologies are capable of analyzing and processing vast amounts of data, supporting decisions in clinical trial design, patient stratification, and the prediction of drug response [4,9,20]. While these capabilities promise major advances, they also raise questions about how to ensure the fair and responsible use of AI.

Ethical principles, understood as duties that promote human dignity in the presence of intelligent systems, must guide the development and deployment of AI in drug development contexts. Although these principles are universal in nature, their application can vary across different cultural, legal, and social settings [14]. As healthcare becomes more data driven and patient centered, AI is increasingly seen as a valuable support tool. However, maintaining human control over clinical decisions remains essential. Human dignity and fundamental values—such as the obligation to avoid harm, ensure fairness, and prevent discrimination—must remain central to the use of AI [4]. This is particularly crucial when AI is involved in sensitive areas like patient selection for trials or prediction of adverse events, where biases in training data can exacerbate existing health inequities.

To ensure ethical integrity, values such as accountability, transparency, explainability, and privacy should be embedded into AI systems from the outset [4]. Table 2 outlines widely accepted ethical principles designed to complement existing human rights standards and guide responsible AI deployment in healthcare.

Table 2. Ethical principles of AI in health identified by the WHO [14,21].

Ethical Principle	Ethical Foundations
Autonomy	- The use of machines must not undermine human autonomy. - Design AI systems that help humans make informed decisions. - Humans can decide whether to use an AI system to reach a final decision.
Public interest	- Ensure safety, accuracy, and effectiveness requirements are met before implementing any AI technology. - Measures must ensure ongoing control and improvement. - Monitor AI algorithm performance to prevent patients or groups from suffering mental or physical harm. - Protect individuals from any form of stigmatization or discrimination due to their health status.
Transparency	- Implementation of documents with sufficient information for AI technology implementation. - Improve system quality and protect patients and public health. - Identify errors. - Adequate and effective supervision. - Include accurate information on the limitations and assumptions of the technology, algorithm development, operational protocols, and data properties. - Educational information adapted to the population.
Responsibility	- Minimize or avoid errors. - Use of technology in a responsible manner. - Identify all parties involved in the development and implementation of the technology and assign collective responsibilities. - Support justice in resolving issues.
Inclusion and equity	- Hiring staff from diverse backgrounds, cultures, and disciplines. - Design and evaluation of technologies by those who will be directly affected by them. - Adoption of open-source software available in both high and low-income countries. - Diversity in languages and communication methods. - Unbiased technologies.
Privacy	- Protect the confidentiality and privacy of individuals and ensure informed consent.

Regulatory bodies, such as the European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA), are actively working to establish frameworks that assess both the technical and ethical robustness of AI-based tools. For instance, the EMA's 2023 Reflection Paper on Artificial Intelligence emphasizes the importance of ethical oversight across the entire AI lifecycle, from data collection to real-world deployment.

Ultimately, ethical governance must involve a multi-stakeholder approach, including developers, clinicians, industry, and regulators, ensuring that AI technologies serve the collective good while safeguarding individual rights.

3. Challenges and Transformations in R&D Through AI Integration

Pharmaceutical companies face significant challenges in the development of new medications, primarily due to the extremely high costs and low success rates associated with the process. Since 2001, drug development expenses have risen by several billion dollars, making cost the leading factor behind inefficiencies in research and development [22,23]. Despite these investments, the return remains low, largely because of high failure rates—especially regarding toxicity and efficacy.

The clinical trial success rate is under 15%, and approximately half of drug discovery failures are due to poor pharmacokinetic properties in the preclinical phase. Although

Phase I trials usually confirm safety, many candidates fail in Phase II because of insufficient efficacy. Additionally, population heterogeneity contributes to this lack of success [22,23].

Traditionally, drug discovery has followed a linear and highly regulated pathway, from target identification through preclinical testing to clinical trials, often taking 10 to 15 years and billions of dollars to bring a single drug to market [24]. This approach was historically centered on single-target drug development, but modern strategies have since evolved to enhance efficiency [24,25]. Consequently, there is a growing need to adopt innovative technologies that can improve the overall R&D process.

The integration of AI (AI) into drug development is explored in Figure 2, detailing the strategies and outcomes of its key components. Additionally, AI applications at each stage of the process are highlighted.

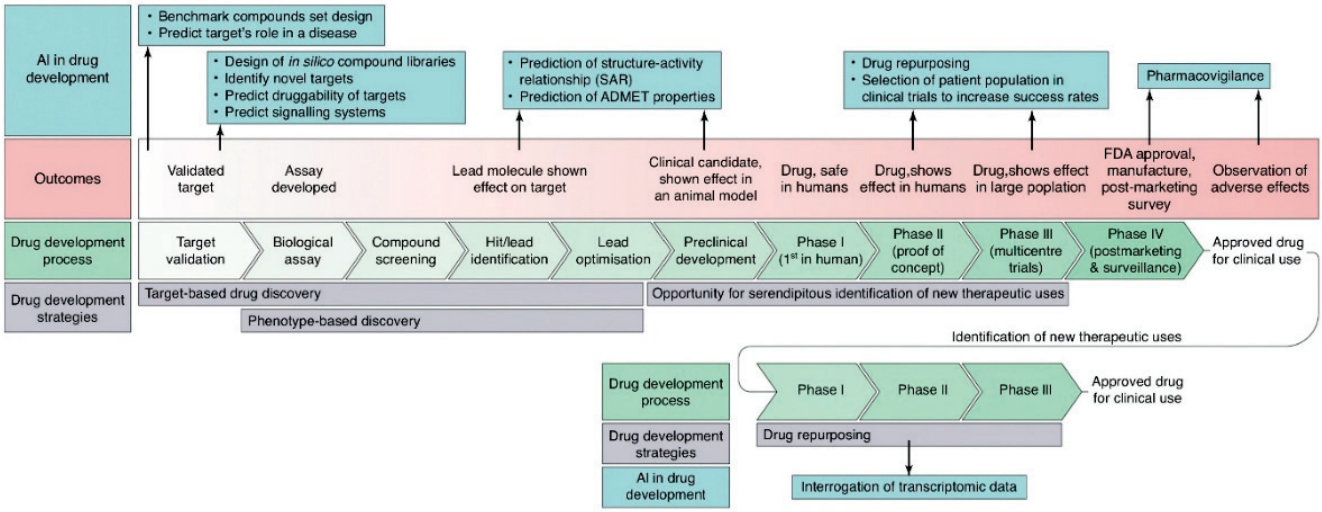


Figure 2. Schematic representation of the application of AI in each stage of the drug development process. Adapted with permission from Ref. [12]. Copyright 2019 Elsevier.

These paradigm shifts in drug development, especially the move toward data-intensive approaches, such as high-throughput screening and drug repurposing, have set the stage for the integration of artificial intelligence. The following sections explore how AI is being applied across these evolving domains, from molecule design to clinical trial execution.

3.1. Application of AI in Formulation

AI has begun to drive its application across various sectors of society, with the pharmaceutical industry emerging as a primary beneficiary. Advancements in AI have shown the potential to discover and develop medications more quickly and at a lower cost [5]. Over the years, there has been a dramatic increase in data input within the pharmaceutical sector, motivating the use of AI to automate and optimize slow processes, thereby significantly accelerating the research and development of medications due to its capability to manage large volumes of data [5].

When applied in the formulation of pharmaceutical products, AI can be utilized at various stages of drug development, including the identification and validation of potential targets, drug design and screening, drug repositioning, improving R&D efficacy, analyzing biomedical information, quality control, pharmaceutical product management, and assisting in the selection of patients for clinical trials [5,12]. The search for new drugs that will be successful in the future is perhaps the most challenging part of the drug development process,

as the estimated number of viable molecules with potential interest in developing new medications ranges between 10^{60} and 10^{100} [12,26]. A viable molecule is defined as one that meets Lipinski's rule of five, which is a computational procedure used to estimate the solubility, permeability, and pharmacological efficacy in drug discovery. According to this rule, molecules must have a molecular weight of less than 500 Da, a number of hydrogen bond donors of less than 5, a number of hydrogen bond acceptors of less than 10, and a LogP of less than 5 [26]. However, the lack of advanced technology limits the process of selecting viable molecules and, consequently, the development of medications, making this process slow and costly, which could be resolved through the use of AI [26].

The virtual chemical space is vast and provides a geographical map of molecules, illustrating their properties and interactions between molecules, facilitating the synthesis of information about compounds that could produce desired therapeutic effects, and also selecting molecules with properties suitable for further testing [5,27]. There are several databases available, such as PubChem, which is freely accessible and provides information on chemical molecules and their activities in biological assays, and ZINC, which contains a vast number of chemical compounds and is designed to facilitate virtual screening. It is essentially organized to assist techniques like molecular docking—virtual simulations of interactions between molecules and protein binding sites, helping to identify new molecules with pharmacological activity [27].

The data available in pharmaceutical companies for developing medications can involve various compounds, and ML tools may struggle to handle this type of data. For this reason, a computational model called QSAR (Quantitative Structure–Activity Relationship) was developed, which can quickly predict a large number of compounds and their physicochemical parameters [5,28]. However, this model cannot predict complex biological properties, such as efficacy and potential adverse effects of compounds. Moreover, QSAR models encounter several challenges, including limited training datasets, inaccuracies in the experimental data, and insufficient experimental validation. To address these limitations, more sophisticated AI methodologies, such as deep learning combined with relevant modeling studies, have been incorporated into the evaluation of drug safety and efficacy through the application of big data analytics [5,28].

3.1.1. Drug Screening

In the preclinical phase, high-throughput screening, omics technologies, and computational tools are now routinely used to evaluate thousands of compounds in parallel. These approaches allow for the simultaneous testing of multiple targets, greatly expanding the number of viable candidates and accelerating early discovery stages. The inclusion of AI in drug discovery has proven to be more effective than in vitro screening. Virtual screening is the most commonly used method to aid in the discovery of bioactive molecules, selecting a viable compound from the vast pool available in databases. Therefore, drug screening involves two main components: the preparation of the compound library and the selection of desirable compounds [29,30].

The SMACC (Small Molecule Antiviral Compound Collection) is a highly curated and meticulously annotated library of compounds tested in phenotypic assays and antiviral-based screening for viruses [29].

Moreover, AI plays a crucial role in the screening process, as it can predict toxicity, bioactivity, ADME properties, and the physicochemical properties of molecules under investigation [30].

One particularly challenging domain where AI has shown promise is the prediction of interactions between conventional drugs and herbal medicinal products, which are in-

creasingly used as dietary supplements. These natural products typically consist of complex mixtures with poorly characterized pharmacokinetics, making traditional interaction screening methods less effective. AI algorithms can integrate large-scale data, including chemical structures, molecular pathways, pharmacological properties, and known interaction patterns, to infer potential risks and mechanistic pathways of drug–herb interactions. A recent review highlighted how AI tools can aid in identifying high-risk combinations, offering valuable support for pharmacovigilance and experimental validation. As the use of herbal supplements grows globally, the incorporation of AI in this context represents a crucial advance in drug safety assessment [31]. AI-driven methods have proven particularly useful in optimizing diagnostic parameters, including thresholds for biomarker-based screening tools. A recent study investigating the aldosterone-to-renin ratio (ARR) for screening primary aldosteronism (PA) demonstrated that with appropriate adjustment of the ARR threshold, screening remains accurate even in patients taking interfering antihypertensive medications. As illustrated in Figure 3, the pharmacological context significantly alters the underlying biomarker values (PAC, PRC, and ARR), necessitating adaptive analytical frameworks. Through receiver operating characteristic (ROC) curve analysis (Figure 4), the authors showed that lowering the ARR threshold in treated patients preserved sensitivity and specificity comparable to standard protocols. These results illustrate the value of data-centric, adaptive approaches, often implemented using AI, in refining clinical decision-making and ensuring diagnostic robustness across real-world patient conditions [32].

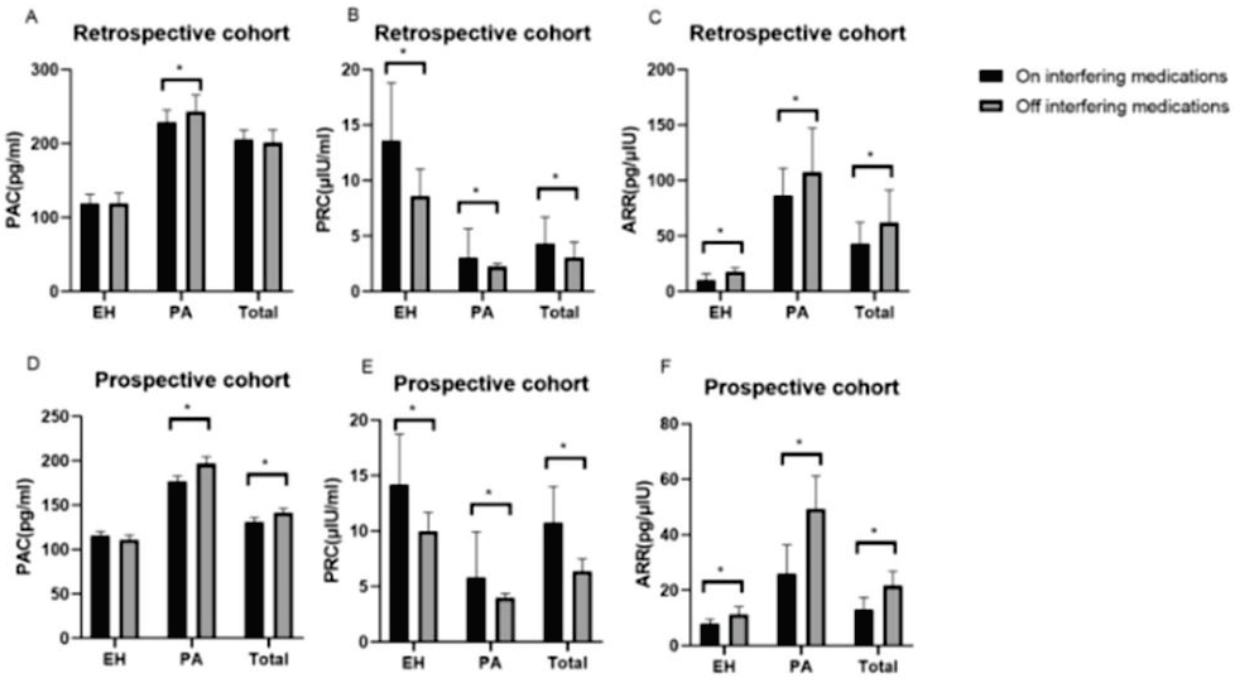


Figure 3. Variations in plasma aldosterone concentration (PAC), plasma renin concentration (PRC), and the aldosterone-to-renin ratio (ARR) observed in patients with essential hypertension (EH) and primary aldosteronism (PA) before and after the discontinuation of interfering antihypertensive medications. (**A**) PAC levels in the retrospective cohort, on and off interfering medications; (**B**) PRC levels in the retrospective cohort, on and off interfering medications; (**C**) ARR values in the retrospective cohort, on and off interfering medications; (**D**) PAC levels in the prospective cohort, on and off interfering medications; (**E**) PRC levels in the prospective cohort, on and off interfering medications; (**F**) ARR values in the prospective cohort, on and off interfering medications. * Statistically significant difference ($p < 0.05$). Reprinted with permission from Ref. [32]. Copyright 2023 Springer Nature.

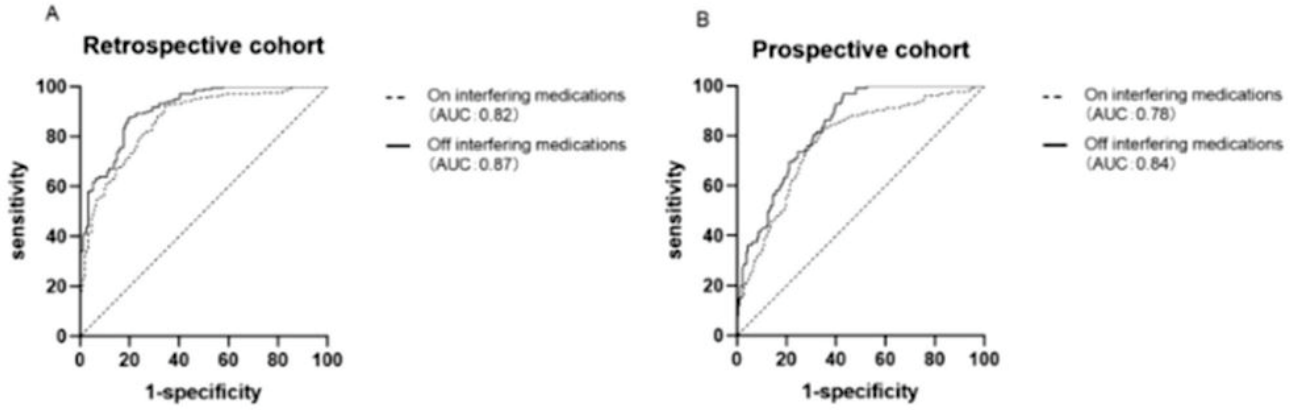


Figure 4. Receiver operating characteristic (ROC) curves comparing the diagnostic performance of the aldosterone-to-renin ratio (ARR) with and without interfering with antihypertensive medications in retrospective (**A**) and prospective (**B**) cohorts. The analysis illustrates how data-driven threshold optimization—analogous to AI-driven calibration—can maintain diagnostic accuracy despite clinical variability. Reprinted with permission from Ref. [32]. Copyright 2023 Springer Nature.

While small molecules have traditionally dominated early-stage drug discovery efforts, particularly due to their favorable physicochemical properties and oral bioavailability, recent advances in AI have increasingly been extended to the design and optimization of biotechnological products, including monoclonal antibodies, therapeutic peptides, and nucleic acid-based therapies. These complex molecules pose unique challenges, such as stability, immunogenicity, and production scalability, which are now being addressed through machine learning algorithms trained on structural, genomic, and pharmacokinetic datasets [33–35].

In this context, traditional heuristics, such as Lipinski's Rule of Five, remain useful for the initial filtering of small molecules but have limited applicability in the design of larger or more complex entities. Moreover, AI-driven discovery platforms are beginning to surpass these empirical rules by identifying candidates that fall outside conventional boundaries yet still demonstrate clinical efficacy. For example, deep learning models that consider 3D conformations, protein–ligand interactions, and dynamic behavior are revealing new "druggable" spaces [36].

The integration of tools like AlphaFold, which accurately predicts protein structures, and the rise of explainable AI (XAI) approaches, which provide interpretability and mechanistic insights, further indicate a shift from traditional rule-based paradigms toward data-driven and structure-informed discovery frameworks. These innovations are expected to expand the range of viable therapeutic candidates and improve the rational design of both small and large molecules [36].

Toxicity Prediction

The toxicity of a drug is a crucial parameter in the drug development process, referring to adverse effects caused by the drug's mechanism of action or metabolites in the body [28,30]. Toxicity tests are initially conducted in vitro on cells, followed by studies in animals to identify the compound's toxicity in a living organism. These methods increase the cost and extend the time required for drug development [37].

Several tools and pipelines are used to predict the toxicity of compounds. For instance, eToxPred predicts bioavailability and toxicity through molecular fingerprints [30]. The Tox21 Data Challenge is a competition where researchers tested thousands of environmental chemicals and drugs for twelve different toxic effects. This led to the creation of the DeepTox algorithm, which first normalizes the chemical representations of the products and calculates

numerous molecular properties used as input for machine learning methods. Then, these chemical properties are grouped, and DeepTox predicts the toxicity of new compounds [37].

Bioactivity Prediction

The effectiveness of a drug hinges on its binding affinity to the target protein or receptor. Molecules that fail to interact or attach to the target protein are unlikely to trigger a therapeutic response [5]. Sometimes, the drugs being studied may interact with unintended proteins or receptors, leading to toxicity. Therefore, it is essential to understand the drug's affinity for its receptor (DTBA—Drug–Target Binding Affinity) in order to predict drug–receptor interactions. Algorithms like ChemMapper are designed to predict such interactions [5,30]. The contribution of AI in this field has been demonstrated through advancements in the analysis of molecular pairs (MMP—Matched Molecular Pair), where the impact of a single molecular change on molecular properties and bioactivity is investigated. This approach is widely used in QSAR testing [28]. A candidate molecule is chemically defined by a static core and two fragments that are subsequently encoded. Machine learning methods, applied after the MMP, can be used again to extrapolate possible new transformations, fragments, and modifications of the static core [28]. DL approaches outperform ML methods in some cases since they do not rely on the availability of 3D protein structures. Examples of such DL methods include DeepDTA, which predicts and analyzes interactions between drugs and target proteins, and DeepAffinity, which considers amino acid sequences and molecules for structural and physicochemical properties, among others [5,30].

Prediction of Physicochemical Properties

Physicochemical properties—solubility, logP (partition coefficient), ionization degree, intrinsic permeability, lipophilicity, and melting point—must be considered when designing a new medication, as they can impact pharmacokinetic properties and interaction with receptors [5,37]. LogP, which is the ratio of a substance's concentration between an organic solvent and water, helps estimate the drug's cellular absorption. On the other hand, the melting point indicates the ease with which a drug dissolves in a medium [37].

Considering the above properties, molecular representations used in AI algorithms for drug design include "fingerprint", a method that represents the molecule as a sequence of numbers, as well as a simplified molecular input line entry system (SMILES) for predicting molecular properties, molecular graphs where atoms are represented as points and bonds as edges, and others [38]. Applying algorithms facilitates the prediction of molecular properties and allows for the identification of potential relationships between molecular properties and their structures. For example, the EPI (Estimation Program Interface) program allows for the evaluation of certain physicochemical properties through the application of the QSPR (Quantitative Structure–Property Relationship) algorithm [39].

3.1.2. Drug Design

The vast availability of data presents a significant challenge for drug design, as it involves navigating an expansive chemical information space associated with various targets to identify the optimal solution [40]. Several milestones have been achieved in the drug design process, establishing it as the primary approach for the present and future. The first major advancement was recognizing and understanding drug–receptor interactions. Daniel Koshland demonstrated that both molecules undergo conformational changes during their interaction, adopting the most suitable conformations to bind effectively [41]. With technological advancements and the integration of AI, the primary breakthrough in this field has been the development of in silico modeling technologies applied to virtual

screening, compound design, SAR and QSAR analyses, pharmacokinetic modeling, and drug–drug interaction studies [41]. AI plays a critical role in drug design by anticipating protein–drug interactions and predicting the structure of target proteins [39].

Prediction of Drug–Protein Interaction

For a drug to have a therapeutic effect, it must first bind to a protein to ensure effective treatment of a disease. Understanding drug–protein interactions is crucial for optimizing drug efficacy and effectiveness. This knowledge is instrumental in avoiding unintended drug repurposing and mitigating polypharmacology-related issues [5]. These interactions are fundamental and are often directly linked to various diseases through distinct biological processes. They involve numerous protein binding sites beyond the conventional targets, such as ion channels, kinases, and nuclear receptors, facilitating the development of small-molecule drugs. However, existing AI methods in this area face limitations due to the availability and diversity of structural data, as only a small fraction of known structures include detailed insights into yeasts, microorganisms, and human proteins [42].

Prediction of Target Protein 3D Structure

When the goal is to develop a molecule that will later be integrated into a medication targeting a specific disease, predicting the structure of the target protein is crucial for ensuring a successful therapeutic outcome [5]. The 3D structure of the target protein is critical because new drug molecules are designed based on the chemical environment of the target protein's binding site [37]. However, predicting the three-dimensional structure of a protein is challenging, as it relies solely on its amino acid sequence, making this task highly complex [43].

Advancements in technology have led to the development of useful methods and tools to assist researchers in addressing this issue. Among these is AlphaFold, the first computational method capable of predicting protein structures with atomic precision, even when no similar structure is known [37,43]. AlphaFold utilizes deep neural networks (DNNs) to predict protein properties from its primary amino acid sequence. It analyzes the distance between adjacent amino acid pairs and the corresponding peptide bond angles to accurately predict the protein's three-dimensional structure. AlphaFold successfully predicted 25 out of 43 protein structures with remarkable precision, surpassing other methods by a significant margin [37,43,44].

3.1.3. Drug Repurposing

In parallel to the new drug development, drug repurposing, the identification of new therapeutic uses for existing or previously shelved drugs, has become an increasingly strategic and cost-effective alternative. This approach circumvents early preclinical testing and leverages existing safety and pharmacokinetic data to advance candidates more rapidly into clinical evaluation. When combined with AI technologies, drug repurposing becomes even more powerful, as machine learning models can mine large-scale biomedical datasets, molecular structures, and disease networks to identify novel indications [45].

Drug repurposing involves discovering new therapeutic uses for already approved drugs (Table 3) [45]. The main advantages of this approach include pre-existing knowledge of toxicity, efficacy, and pharmacokinetic and pharmacodynamic profiles. This allows the drug to proceed directly to Phase 2 or Phase 3 clinical trials, significantly reducing associated costs and development time (Figure 5) [6,46].

Table 3. Examples of drugs subject to the mechanism of repurposing using AI techniques [47].

Molecule	Original Therapeutic Indication	New Therapeutic Indication
Ormeloxifene	Oral contraceptive	Prostate Cancer
Metformin	Diabetes	Cancer
Cimetidine	Ulcer	Lung Cancer
Bazedoxifene	Osteoporosis	Breast Cancer

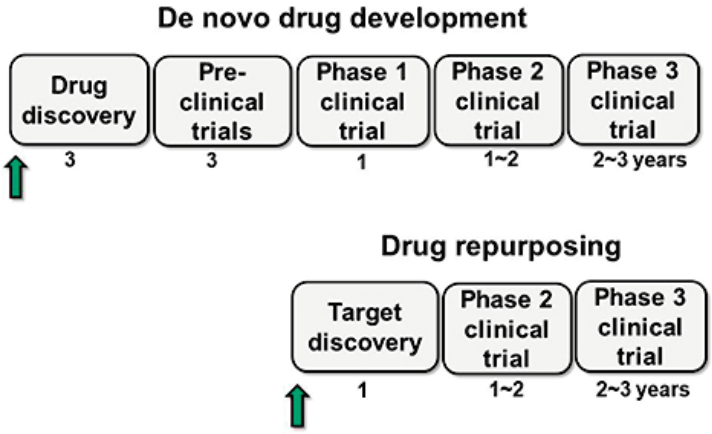


Figure 5. Schematic representation of the new drug development and drug repurposing. The arrow indicates the starting point for each approach, while the numbers represent the duration in years. In this figure, drug repurposing is considered the process of identifying or discovering new therapeutic targets for an already marketed drug. Unlike the new development, repurposing begins with target discovery and proceeds directly to Phase 2 and 3 clinical trials, bypassing animal studies and Phase 1 trials, as existing data from the original drug can be leveraged. Adapted from [46].

Drug repurposing can be carried out experimentally or through computers, with the latter method being referred to as in silico drug repurposing, which can be classified according to two approaches: discovering new indications for an existing drug—drug centered—or identifying effective drugs for diseases—disease centered [46].

The concept of in silico drug repurposing is currently widely used, and its expansion was made possible due to two types of technologies, such as the high-throughput data collection from various sources, including genomics, proteomics, chemoproteomics, and phenomics, which were generated and stored, and due to advances in computational and data sciences that enabled the development of repositioning algorithms along with retrospective analyses and maintenance of experimental data in a database [46].

This strategy includes **target-based drug repurposing,** which screens almost all drug compounds with known chemical structures and directly links targets to disease mechanisms; **pathway-based repurposing**, which utilizes metabolic and signaling pathways and information about protein interactions to predict the similarity or connection between drugs and diseases; and **mechanism-based repurposing**, which interlinks information about signaling pathways, treatment-related omics data, and protein interaction networks to discover new mechanisms of action for drugs [46].

ML techniques applied to drug repurposing include logistic regression, support vector machine (SVM), random forest, neural forest (NN), and DL [46].

Ormeloxifene, originally developed as a non-steroidal oral contraceptive, has been repurposed for cancer treatment using AI-driven approaches. Machine learning algorithms analyzing gene expression data and molecular docking studies have identified its potential in targeting signaling pathways involved in cancer progression, particularly in breast and

ovarian cancers. By leveraging deep learning models trained on drug–target interaction networks, researchers have found that ormeloxifene exhibits anti-cancer properties by modulating cell cycle regulation and apoptosis [47].

Widely used for type 2 diabetes, metformin has gained attention as a potential anti-cancer agent through deep learning models that analyze large-scale patient datasets, gene expression profiles, and cancer cell line responses. AI-based techniques, particularly unsupervised clustering and neural networks, have linked metformin to the inhibition of the mTOR and AMPK pathways—key regulators of cancer metabolism and proliferation. These findings have led to multiple clinical trials investigating metformin's role in preventing and treating various cancers, including breast and colorectal cancer [47].

Initially developed as an H2 receptor antagonist for gastric ulcers and acid reflux, cimetidine was identified as a potential cancer therapy through AI-based drug repurposing pipelines. Computational methods, including network-based drug discovery and virtual screening, have demonstrated that cimetidine can modulate immune response and inhibit cancer cell adhesion, particularly in colorectal and gastric cancers. AI-driven molecular docking studies suggest that cimetidine disrupts tumor immune evasion by blocking E-selectin-mediated cancer metastasis, making it a promising candidate for combination cancer therapies [47].

Finally, originally approved for osteoporosis treatment, bazedoxifene has been repurposed for oncology using deep learning techniques that analyze molecular binding affinities and structural similarities with known anti-cancer agents. AI-driven screening of protein–ligand interactions has shown that bazedoxifene acts as a potent STAT3 inhibitor, a key target in various cancers, including breast and pancreatic cancers. By leveraging reinforcement learning and generative models, researchers have further optimized its anti-cancer potential, leading to preclinical studies exploring its role in targeted cancer therapies [47].

These examples highlight how AI techniques—ranging from machine learning on biological data to deep learning-based molecular simulations—are accelerating drug repurposing efforts, offering new therapeutic possibilities for cancer treatment [47].

3.1.4. Polypharmacology

The concept of polypharmacology shifts from the traditional "one drug, one target" approach and explores how different drugs can interact with multiple pharmacological targets, potentially producing side effects or therapeutic benefits [48,49]. Polypharmacology involves addressing several challenging issues in drug discovery, including identifying synergistic targets, designing chemical molecules with the desired multi-target profile, and evaluating their bioactivity [48,49]. The emergence of the concept of multifactorial diseases required the involvement of multi-target (MT) approaches, becoming a key focus in various therapeutic areas. Drug interactions with different targets could lead to unexpected side effects, involving "anti-targets", such as ion channels or metabolizing enzymes. However, not all these unintended effects are undesirable, as they provide a basis for drug repurposing [48].

AI has helped revolutionize this concept of polypharmacology by providing extensive knowledge about drugs and their interactions. SynPhayhy, an ML-based approach, uses a neural network based on drug chemical structures, offers insights into drug synergy, and identifies relevant biological pathways. DeepMDS also predicts synergy between multiple drugs based on therapeutic targets and gene expression information [49].

3.1.5. Integration of AI with PBPK and Population PK Models

The integration of artificial intelligence with Model-Informed Drug Development (MIDD) offers a powerful paradigm shift in pharmaceutical R&D. MIDD approaches, such as physiologically based pharmacokinetic (PBPK) and population pharmacokinetic (Pop-PK) modeling,

rely on mechanistic simulations to predict drug behavior under various physiological conditions. When enhanced by AI techniques, including machine learning, deep learning, and generative algorithms, these models can incorporate complex, real-world datasets to refine predictions, simulate virtual populations, and optimize dosing strategies. This synergistic integration allows for more accurate hypothesis generation, derisking of candidates, and acceleration of development timelines, particularly in early-phase trials. However, successful implementation depends on the availability of high-quality data, robust validation processes, and attention to interpretability and algorithmic fairness. Regulatory agencies have begun adapting their evaluation frameworks to accommodate this hybrid approach, acknowledging its transformative potential for drug discovery and personalized medicine [50].

Pharmacokinetic modeling has long been essential for understanding drug absorption, distribution, metabolism, and excretion (ADME) characteristics. Among the most widely used approaches are physiologically based pharmacokinetic (PBPK) models and population pharmacokinetic (Pop-PK) models. PBPK models simulate drug behavior using mathematical representations of physiological systems, incorporating variables such as tissue volumes, blood flow, and enzyme activity. Pop-PK models, by contrast, analyze variability in drug kinetics across diverse patient populations to support dose optimization and safety assessments [51,52]. The integration of AI and machine learning techniques into PBPK and Pop-PK modeling has significantly enhanced predictive accuracy and efficiency. AI algorithms can process vast clinical and preclinical datasets to automatically calibrate model parameters, identify nonlinear patterns, and generate simulations that would be unfeasible using conventional methods alone. This hybrid modeling approach allows for more accurate dose prediction, better inter-subject variability characterization, and faster regulatory submissions [53,54]. Moreover, AI-enhanced PBPK models have been used to anticipate drug–drug interactions, simulate scenarios in pediatric or geriatric populations, and guide first-in-human (FIH) dose selection. In combination with real-world data and omics-based biomarkers, these models are also instrumental in precision dosing strategies. By reducing the number of empirical studies required and enabling earlier go/no-go decisions, the fusion of AI with pharmacokinetic modeling contributes directly to cost reduction, faster development, and improved clinical outcomes [55,56].

3.2. Integrating AI in Clinical Trials

The data generated from clinical trials form the basis for evidence-based medicine, guiding healthcare professionals in making informed decisions about patient care. Through these rigorous tests in diverse populations, clinical trials not only validate the effectiveness of new interventions but also uncover potential side effects and long-term impacts. Ultimately, the impact of clinical trials extends beyond individual patients, shaping the landscape of healthcare and improving outcomes on a global scale [57]. However, while clinical trials are indispensable for advancing medical research and establishing the safety and efficacy of new interventions, they come with inherent limitations that can impact their efficiency, generalizability, and ability to address certain complex healthcare challenges. The limited patient diversity, stringent inclusion criteria, short follow-up durations, and ethical concerns are some of the limitations that pose obstacles to the comprehensive understanding of treatments and their real-world implications [58].

In recent years, the incorporation of AI technologies has become a transformative force, providing innovative solutions to address these challenges. This intersection of clinical trials and AI holds tremendous promise, providing tools and methodologies to enhance patient recruitment, optimize trial designs, streamline data management, and address ethical considerations [59–62].

Artificial intelligence has increasingly demonstrated its utility in enhancing the efficiency and precision of clinical trial processes, particularly in patient recruitment, trial design, and adaptive protocols. In the recruitment phase, AI models can analyze vast datasets—such as electronic health records, genomic data, and clinical notes—to identify individuals who match specific inclusion and exclusion criteria. This targeted approach not only accelerates recruitment but also improves population diversity and trial representativeness [63].

In trial design, AI-powered analytics enable the simulation of multiple study scenarios by integrating historical clinical data, disease progression models, and demographic variability. This allows researchers to optimize key parameters such as sample size, endpoint selection, and stratification strategies. Furthermore, adaptive trial protocols have benefited from real-time AI monitoring systems capable of analyzing incoming data streams during the trial. These systems support dynamic adjustments—such as dose modification, cohort expansion, or early termination—based on predefined statistical thresholds. By doing so, AI enhances both the ethical and scientific rigor of trials, minimizing patient exposure to ineffective interventions while maximizing data relevance and trial success rates [10,59,64].

In fact, a recent study analyzed the success of AI-discovered drugs in clinical trials, highlighting their growing impact on pharmaceutical research. The study found that AI-driven biotech companies have introduced 75 drug candidates into clinical trials since 2015, with 67 still progressing as of 2023. Notably, AI-discovered molecules demonstrated an 80–90% success rate in Phase 1 trials, significantly surpassing traditional drug development benchmarks. In Phase 2 trials, these drugs exhibited a 40% success rate, aligning with historical industry averages. The study categorized AI-discovered drugs into key areas, including small molecules, biologics, vaccines, and repurposed drugs, with oncology being a primary focus. Additionally, the rapid expansion of AI in drug discovery has accelerated development timelines and fostered collaborations between biotech startups and major pharmaceutical companies [65].

AI algorithms can analyze vast amounts of patient data from electronic healthcare records, the medical literature, and other sources to identify suitable candidates for clinical trials. Natural language processing (NPL) techniques enable the extraction of relevant information from unstructured data, helping researchers find individuals who meet specific criteria quickly and accurately [66].

Designing an effective clinical trial protocol is crucial for its success. AI-powered predictive analytics can analyze historical trial data to identify patterns, predict potential outcomes, and optimize study protocols. By considering various factors, such as patient demographics, disease characteristics, treatment protocols, and endpoints, AI algorithms can suggest adjustments to enhance trial effectiveness and efficiency [59,60]. In addition, AI algorithms can analyze patient data to predict treatment responses and outcomes. By integrating data from multiple sources, including genomics, proteomics, and clinical parameters, AI can identify biomarkers and patient subgroups that are more likely to respond favorably to a particular treatment. This predictive capability helps in personalizing treatment regimens and improving patient stratification in clinical trials [67].

AI-powdered monitoring systems can continuously analyze incoming data from clinical trials in real time. By detecting deviations from expected patterns and identifying safety concerns or efficacy signals early, AI enables researchers to make timely adjustments to trial protocols. This capability is particularly valuable in adaptative trials, where trial parameters can be modified dynamically based on accumulating data.

AI algorithms can assess patient risk factors and predict adverse events, allowing researchers to proactively mitigate risks and ensure patient safety during clinical trials. By

analyzing patient characteristics, medical history, and treatment-related factors, AI can identify individuals who may be at higher risk of experiencing adverse events or treatment failures, enabling tailored interventions and monitoring strategies [68–70].

One of the most significant bottlenecks in clinical trials is patient recruitment. In fact, finding eligible patients who meet specific criteria can delay trials for months or even years, which drives up costs and postpones potentially life-saving treatments from reaching the public. AI addresses this challenge by analyzing vast datasets, such as electronic health records (EHRs), to identify potential participants based on factors like age, medical history, and genetic makeup. Using AI-powered tools, researchers can quickly match patients to trials, improving the chances of filling participant quotas efficiently. Furthermore, AI systems enhance retention rates by offering personalized outreach, sending tailored reminders, and providing ongoing support. These platforms keep participants engaged, reducing the risk of dropout and ensuring that trials can continue without disruptions.

To optimize patient selection for osteoarthritis (OA) clinical trials, a machine learning (ML) strategy was developed to identify individuals with a higher likelihood of disease progression. This "progression-enriched" approach aims to better evaluate the effectiveness of disease-modifying treatments by focusing on patients most likely to exhibit OA progression. The recruitment strategy was structured in two stages. In the first stage, ML models were trained on data from existing OA cohorts to rank candidates by their probability of progression. This model narrowed down the pool of potential participants for an initial screening visit. In the second stage, a refined ML model incorporated data gathered during screening to make the final patient selection. The effectiveness of this strategy was validated in the IMI-APPROACH knee OA study. Of 3500 candidates, 433 patients underwent screening, 297 were enrolled, and 247 completed a 24-month follow-up. The ML model demonstrated high accuracy in predicting pain-related progression (AUC 0.61). The progression rates in the selected population varied, with 30% showing pain-related progression, 13% showing structural changes, and 5% showing both. Furthermore, progressors were ranked higher than non-progressors, indicating the model's success in identifying those more likely to experience disease progression. This ML-assisted recruitment strategy effectively enriched the trial population with progressive OA patients, demonstrating its potential to enhance trial efficacy. However, further refinement is needed to improve predictions for structural progression, and additional research is necessary to validate this approach in an interventional setting [71].

Figure 6 schematically presents the application of AI in drug development, illustrating the interaction between the different stages mentioned earlier.

Beyond structured clinical trial data, the integration of real-world data (RWD) and natural language processing (NLP) has significantly broadened the analytical capabilities of AI in pharmaceutical research. RWD encompass information collected from a variety of sources outside traditional clinical trials, such as electronic health records (EHRs), claims and billing data, mobile health apps, wearable sensors, patient registries, and even social media platforms. These data offer insights into patient behavior, adherence, comorbidities, and long-term treatment outcomes, providing a more comprehensive view of real-world clinical practice [72,73]. AI algorithms can process these large, heterogeneous datasets to identify patient subpopulations with specific clinical characteristics, predict progression trajectories, and simulate trial outcomes under various scenarios. For instance, machine learning models trained on EHR data can detect hidden phenotypes or patterns indicative of eligibility, thus enhancing patient stratification and reducing recruitment time. Moreover, RWD supports external control arms, allowing researchers to compare intervention groups with historical real-world cohorts, which is particularly valuable in rare diseases or ethical situations where placebo use is limited [74–76].

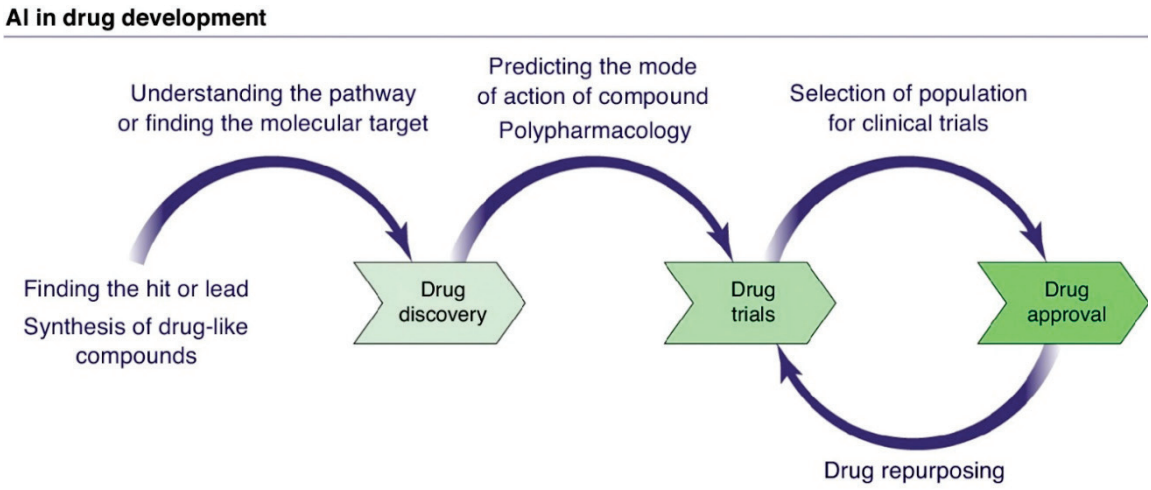


Figure 6. Schematic representation of AI application in drug development. Adapted with permission from Ref. [12]. Copyright 2019 Elsevier.

Complementing this, natural language processing (NLP) enables the extraction of structured information from vast amounts of unstructured text, such as clinical notes, pathology reports, radiology descriptions, and biomedical literature. NLP has been used to automatically identify adverse drug reactions, patient-reported outcomes, and even subtle clinical indicators not captured in standard coding systems. By integrating NLP outputs into AI models, researchers can uncover associations and patient characteristics that would otherwise remain inaccessible, enriching trial data and supporting adaptive protocols. Furthermore, NLP facilitates the automated screening of trial eligibility by parsing narrative health records, thereby streamlining recruitment workflows [77,78].

Artificial intelligence methods have also shown remarkable promise in the area of adverse drug reaction (ADR) detection, a critical component of post-marketing surveillance. A recent study introduced a novel deep learning framework known as Neural Self-Controlled Case Series (NSCCS), designed specifically for ADR discovery using large-scale electronic health records (EHRs). By following a self-controlled case series methodology, the model implicitly adjusts for time-invariant confounders, enhancing the reliability of the detected associations between drugs and adverse conditions. The framework was validated on real-world clinical data and demonstrated superior performance in identifying ADRs compared to traditional methods, illustrating the potential of deep neural networks to extract clinically actionable insights from complex longitudinal datasets [75]. Together, the synergistic use of RWD and NLP enhances the relevance, inclusiveness, and efficiency of clinical trials, bridging the gap between controlled experimental settings and real-world clinical environments.

Recent studies also demonstrate the potential of AI tools beyond trial design and recruitment, particularly in the context of post-treatment patient management and supportive care. For instance, a three-year clinical follow-up study in patients with stage III–IV nasopharyngeal carcinoma (NPC) showed that AI-assisted home enteral nutrition (HEN) management significantly improved long-term health outcomes. Patients who received AI-guided nutritional support exhibited higher survival rates, lower recurrence and metastasis, better nutritional profiles, and reduced psychological stress and physical discomfort when compared to standard care. These effects were clearly illustrated through objective physical and biochemical indicators, including body mass composition (Figure 7), and improved laboratory values, such as albumin and hemoglobin levels (Figure 8). These findings reinforce the relevance of AI not only in accelerating drug development but also in enhancing quality of life and treatment recovery trajectories [79].

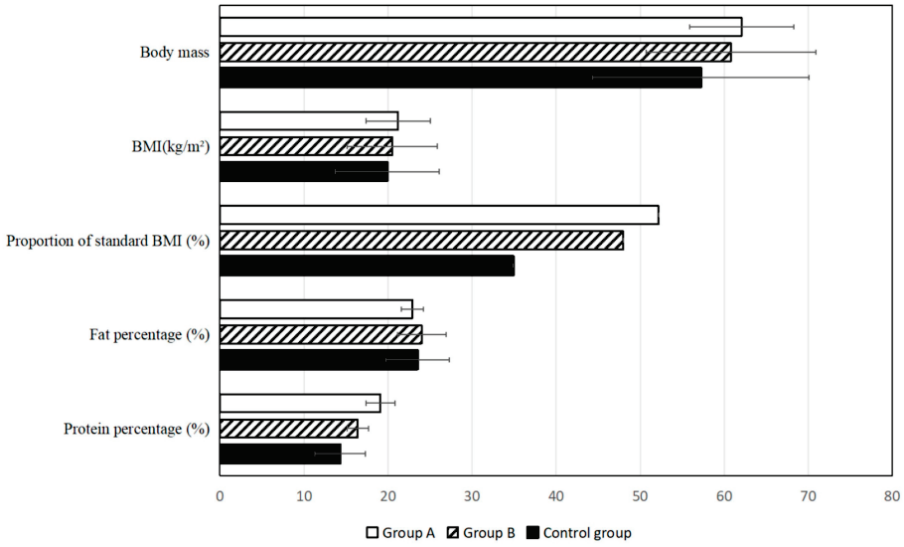


Figure 7. Comparative analysis of basic body quality indicators, including body mass index (BMI) and body protein levels, among patients with stage III–IV nasopharyngeal carcinoma (NPC) three years after treatment. Patients receiving AI-assisted home enteral nutrition (HEN) management (Group A) exhibited significantly better nutritional status than those under traditional HEN (Group B) or no HEN (control group) [79].

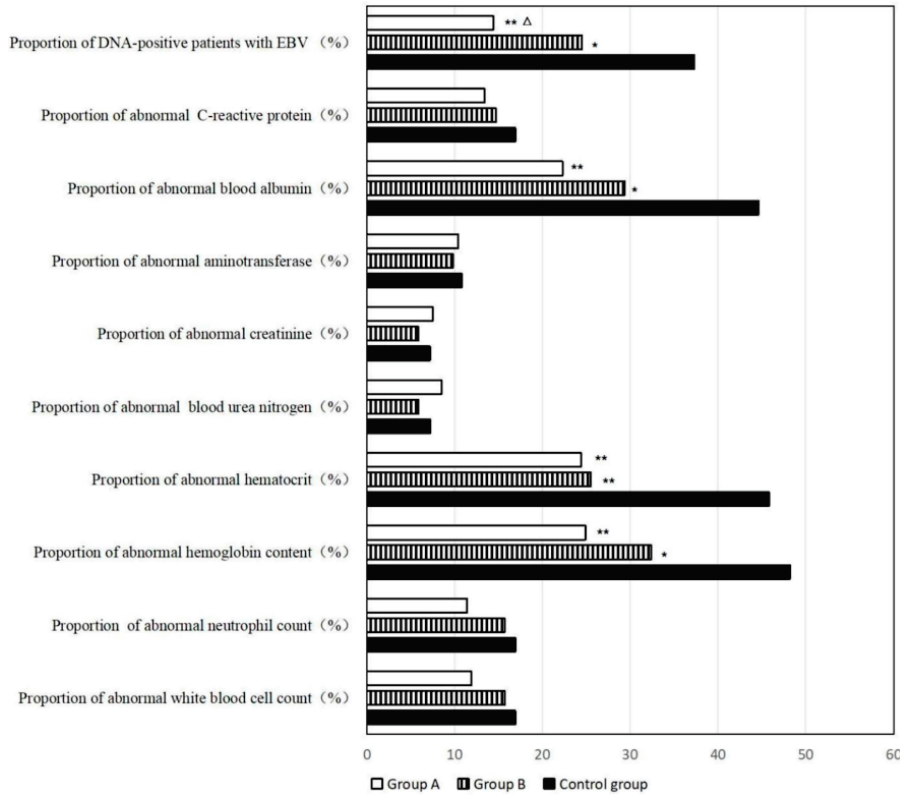


Figure 8. Laboratory results assessing hematologic and biochemical parameters among NPC patients three years post-treatment. AI-assisted HEN (Group A) was associated with higher hemoglobin and albumin levels and a lower Epstein–Barr virus (EBV) DNA positivity rate compared to the control group. * Compared with control group, $0.01 < p < 0.05$; ** compared with control group, $p < 0.01$; △ compared with Group B, $0.01 < p < 0.05$ [79].

3.3. AI-Developed Drugs: Highlighted Cases

Computational methods have played a crucial role in drug discovery. The use of medication design with the aid of computers has contributed to the development of several drugs that have already been approved, such as the initial examples of captopril, saquinavir, ritonavir, indinavir, and tirofiban [80].

In January 2020, the British pharmaceutical company Exscientia published that DSP-1181, a potent long-lasting 5-HT1A serotonin receptor agonist developed to treat OCD (Obsessive–Compulsive Disorder), had entered Phase 1 clinical trials. An AI platform was created to identify relevant compounds from chemical libraries. In partnership with the Japanese pharmaceutical company Sumitomo Dainippon Pharma, the drug took just 12 months to develop, from screening to the end of preclinical trials [81,82]. Exscientia has developed other drugs using AI, including DSP-0038, which entered Phase 1 clinical trials in the UK in May 2021. It was also developed in collaboration with Sumitomo Dainippon Pharma and aims to alleviate the behavioral and psychological symptoms of dementia, such as agitation, anxiety, and aggression, in individuals suffering from Alzheimer's disease, as it has a low affinity for dopamine D2 receptors, improving its tolerability compared to conventional antipsychotics. EXS-21546, developed in December 2020, also entered Phase 1 clinical trials and acts as an immuno-oncology agent for various types of tumors. It was created in collaboration with Evotec [83].

Another example is Baricitinib, which was defined by BenevolentAI, a company that uses AI to discover new uses for existing drugs and to predict their effectiveness in treating SARS-CoV-2. It was found to be effective in reducing the virus's ability to infect lung cells due to its anti-inflammatory and antiviral properties [84].

It is also relevant to mention a study that employed AI to screen 7684 molecules, training a predictive model to identify potential antibacterial compounds. The model then analyzed 6680 additional molecules and, within hours, successfully pinpointed 240 candidates for experimental testing. Among them, Abaucin emerged as a potent and highly selective antibiotic against *Acinetobacter baumannii*, a dangerous pathogen responsible for hospital-acquired infections. Unlike broad-spectrum antibiotics, Abaucin specifically targets *A. baumannii* by disrupting lipoprotein transport via the LolE protein, minimizing off-target effects on beneficial bacteria. In animal models, Abaucin demonstrated strong antibacterial activity, significantly reducing infection and inflammation in mice with *A. baumannii*-infected wounds. This breakthrough illustrates AI's potential to accelerate drug discovery, optimize target specificity, and improve treatment efficacy, marking a shift toward more efficient and precise antibiotic development [85].

Another notable example is GENTRL, an AI-driven generative reinforcement learning model that successfully designed DDR1 kinase inhibitors, a promising drug class for treating fibrosis and cancer. Similarly, DeepMalaria, a deep learning-based model, identified compounds with Plasmodium falciparum inhibitory activity, contributing to malaria drug discovery. Another breakthrough is ReLeaSE, a reinforcement learning model capable of generating new drug-like molecules while predicting their biological activity with high accuracy. Additionally, AlphaFold, an AI-powered protein structure prediction tool, has transformed drug target identification by accurately predicting 3D protein structures, significantly improving drug design strategies. These advancements highlight AI's ability to streamline drug development, optimize molecular properties, and repurpose existing drugs for new therapeutic applications, paving the way for more efficient and targeted treatments [86].

INS018_055 is an innovative drug candidate for idiopathic pulmonary fibrosis (IPF), a severe lung disease with limited treatment options. Developed by Insilico Medicine, this small molecule was designed using deep learning and generative AI models, which screened millions of chemical structures to predict an optimal candidate. AI significantly accelerated

the drug's development, allowing it to progress from discovery to Phase 1 clinical trials in record time. INS018_055 represents a breakthrough in AI-driven drug discovery, showcasing how machine learning can rapidly identify novel treatments for complex diseases [87].

Halicin is an antibiotic discovered through AI by a team at MIT. It was identified by a deep learning model known as Chem net, which was trained to predict antimicrobial activity across various compounds. The AI system searched chemical databases to find molecules with the potential to combat drug-resistant bacteria, and Halicin emerged as a promising candidate. Unlike conventional antibiotics, Halicin operates by disrupting the proton gradient across bacterial cell membranes, a mechanism that was previously unrecognized. The drug demonstrated broad-spectrum antimicrobial activity, effectively targeting a wide range of bacteria, including those resistant to multiple existing antibiotics. In preclinical tests, Halicin was shown to be effective against pathogens such as *Clostridium difficile*, *E. coli*, and *Mycobacterium tuberculosis*. Halicin represents a major breakthrough in the fight against antibiotic resistance, showcasing the potential of AI to accelerate drug discovery and provide novel solutions to longstanding global health challenges. Although still in the early stages of development, Halicin holds significant promise for tackling the growing threat of antibiotic-resistant infection [88].

Several artificial intelligence platforms have already demonstrated a concrete impact on pharmaceutical innovation, providing real-world examples of how these technologies are reshaping traditional drug discovery paradigms [36]. One of the most prominent is AlphaFold, developed by DeepMind, which revolutionized protein structure prediction. This platform employs deep learning techniques to accurately predict the three-dimensional structure of proteins based solely on their amino acid sequences, overcoming one of the major bottlenecks in structural biology. The wide accessibility of AlphaFold's predictions has significantly accelerated the identification of novel drug targets, facilitating structure-based drug design [89,90]. Another noteworthy example is BenevolentAI, a platform that integrates machine learning with biomedical data mining to identify new therapeutic applications for existing drugs. During the COVID-19 pandemic, BenevolentAI successfully identified Baricitinib—a drug originally approved for rheumatoid arthritis—as a potential treatment for SARS-CoV-2 infection. The repurposed use of Baricitinib was rapidly validated through clinical trials and received emergency use authorization, demonstrating the practical relevance and speed that AI-based approaches can bring to drug development [91,92].

Atomwise is also at the forefront of AI in drug discovery. Using deep convolutional neural networks, their AtomNet platform performs structure-based virtual screening to predict binding affinities between small molecules and biological targets. This technology has been applied across multiple therapeutic areas, including oncology, neurology, and infectious diseases. Through partnerships with academic institutions and pharmaceutical companies, Atomwise has contributed to the identification of new candidate molecules, significantly reducing the time and resources required for early-stage discovery [93,94].

These platforms illustrate how AI can produce clinically meaningful results and support the development of innovative therapies with enhanced precision and efficiency.

3.4. AI in Biomedical Systems Modeling and Network Dynamics

In complex biomedical and pharmacological systems, where feedback loops and time-delayed interactions are common, identifying the underlying network structure in a timely manner is crucial for accurate modeling and intervention. While traditional topology identification techniques often rely on asymptotic convergence, recent advances in finite-time topology identification offer faster and more robust alternatives. A novel method based on finite-time stability theory has demonstrated effective synchronization and topology reconstruction of de-

layed dynamical networks, with potential implications for biological systems where dynamic connectivity influences drug response. Although initially applied to power grids for rapid detection of line outages, the proposed method proved capable of fast and accurate topology identification, even in second-order nonlinear systems. Future developments—including the incorporation of switching topologies and resilience against cyber attacks—may further enhance its applicability in healthcare-related network modeling [95].

3.5. Limitations, Risks, and Real-World Performance of AI Applications

Despite the growing number of AI-driven tools in pharmaceutical R&D, several critical limitations continue to hinder their effective translation into real-world success. One of the major challenges is data heterogeneity. Pharmaceutical data originate from highly diverse sources, including omics datasets, clinical records, imaging, and chemical structures, which often differ in format, quality, scale, and completeness. These discrepancies can impair model performance, reduce reproducibility, and require extensive preprocessing and standardization [96].

Another major concern is algorithmic bias, which arises when training datasets are unbalanced or not representative of the target population. This can lead to biased predictions, poor generalizability, and potential safety risks, especially when AI models are applied to underrepresented groups in clinical settings [96,97]. Model interpretability also remains a significant barrier. Many deep learning approaches operate as "black-box" systems, making it difficult for researchers and regulators to understand the rationale behind specific predictions. The emergence of explainable AI (XAI) seeks to mitigate this by providing interpretable outputs and highlighting key variables that influence model decisions. However, implementation in regulated environments still faces resistance due to the complexity of the models [91,96,97]. Equally important is the lack of real-world validation for many AI-derived insights. While models may perform well in retrospective or simulated datasets, prospective validation in live clinical or regulatory contexts is rare. This limits confidence in AI-generated drug candidates and slows adoption.

Although there have been notable successes, such as *Baricitinib*—repurposed for COVID-19 treatment using AI-based target prediction—and the rapid design of *DSP-1181*, failures are less often reported. Several AI-predicted candidates have shown promise in silico but failed to replicate efficacy in vitro or in early clinical phases, often due to overfitting, flawed assumptions, or lack of biological plausibility [91,92]. For AI to become a dependable tool in drug development, robust data governance, external validation, and transparent reporting standards must be established, ensuring that AI models are not only innovative but also reproducible, equitable, and clinically meaningful.

4. Regulatory Issues

The integration of AI and ML into pharmaceutical research and development presents significant regulatory challenges. Existing regulatory frameworks were not initially designed to evaluate AI-driven therapeutics, and their rapid evolution has outpaced the development of specific guidelines tailored to these tools [98,99]. However, regulatory agencies worldwide are actively adapting their frameworks to the unique challenges posed by AI and ML in pharmaceutical research and healthcare. For example, the U.S. Food and Drug Administration (FDA) has launched targeted programs, such as the Digital Health Software Precertification (Pre-Cert) Program and the Artificial Intelligence/Machine Learning-Based Software as a Medical Device (SaMD) Action Plan. These initiatives aim to streamline the evaluation and approval of AI-driven tools while ensuring patient safety and efficacy [100,101]. Notably, AI-based medical devices like IDx-DR, an autonomous system

for diabetic retinopathy diagnosis, have already received FDA clearance, exemplifying the agency's proactive stance toward validated AI technologies [101].

Recent regulatory reviews emphasize the growing consensus on the need for risk-based oversight frameworks to ensure the ethical, safe, and scalable use of AI in drug and biological product development. A multidisciplinary review reflecting on FDA workshops has highlighted the importance of practical regulatory guidelines, cross-border harmonization, and the integration of human oversight in AI-assisted decision-making processes. These reflections underscore that beyond technical validation, regulatory approaches must also address broader issues such as data privacy, transparency, governance, and equitable access. As global agencies navigate differing regulatory perspectives, a unified ethical and legal framework for AI in drug development is becoming increasingly vital [97].

One of the central concerns lies in the lack of clear, standardized guidance on how to evaluate and validate AI-derived predictions and computational models. Unlike conventionally developed molecules—typically assessed through well-established experimental paradigms—AI-generated candidates often arise from complex, opaque algorithms, particularly deep learning models, which may lack interpretability and mechanistic transparency. This "black-box" nature can hinder regulatory confidence and pose challenges to the assessment of safety, efficacy, and reproducibility. To address these issues, agencies such as the FDA have introduced several initiatives. For instance, the FDA's "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan" emphasizes the importance of flexibility in oversight while maintaining rigorous standards of patient safety. A key element of this initiative is the development of "Predetermined Change Control Plans" (PCCPs), which allow for software updates without resubmission, provided that modifications are pre-specified and controlled. Complementing this, the FDA is developing Good Machine Learning Practice (GMLP) guidelines, in collaboration with organizations like IEEE and ISO, to ensure algorithmic quality, transparency, and reliability [98,100].

Additionally, the FDA has launched the Digital Health Software Precertification (Pre-Cert) Program and the Emerging Technology Program to accelerate the evaluation and approval of novel AI-based tools [101]. Real-world regulatory approvals, such as the clearance of IDx-DR—an autonomous AI diagnostic system for diabetic retinopathy—demonstrate the agency's openness to validated AI applications. Similarly, the FDA's Real-World Performance (RWP) Monitoring program seeks to track the post-market performance of AI/ML-based tools to ensure ongoing safety and efficacy [102].

In Europe, the EMA has taken similar steps through its Innovation Task Force, which provides early dialogue with developers of AI-driven health technologies. The EMA has stressed the importance of transparency, robust data quality, and post-market surveillance in submissions that involve AI components. Under the EU's Medical Device Regulation (MDR), AI-based medical devices must undergo stringent classification and conformity assessments. However, both the FDA and EMA continue to face challenges in regulating continuously learning AI systems, particularly with respect to algorithmic drift and the need for dynamic risk assessment [103].

The successful integration of physiologically based pharmacokinetic (PBPK) models into regulatory submissions provides a relevant precedent for the future acceptance of AI tools. Originally limited to exploratory simulations, PBPK models are now routinely used to inform dosing strategies, predict drug–drug interactions, and guide extrapolations to special populations. Their journey to regulatory acceptance—through systematic validation, standardization, and inclusion in guidance documents—highlights a viable pathway for future AI technologies [104].

Another emerging regulatory frontier is the incorporation of in silico clinical trials and virtual patient populations. These AI-driven simulations, powered by real-world data and mechanistic models, allow researchers to predict drug responses, optimize protocols, and explore therapeutic scenarios without direct patient involvement. Although not a replacement for traditional clinical trials, they are increasingly viewed as valuable complements that can enhance trial design and improve resource efficiency [105,106]. At the same time, several risks require careful consideration. Data quality and bias represent significant regulatory hurdles, as training data may be unbalanced or non-representative, potentially leading to discriminatory outcomes. As such, regulators are emphasizing the need for fairness audits, diverse training datasets, and post-market surveillance to detect algorithmic biases [107].

The issue of interpretability is also central. Regulatory bodies require explainable AI (XAI) approaches that can justify molecular selection and decision-making processes, especially in contexts involving patient safety. Companies are responding by incorporating XAI methods into development pipelines, aiming to build models that are both powerful and interpretable [108]. Another complex challenge is the question of intellectual property (IP) rights for AI-generated drug candidates. Current patent systems were designed around human inventorship, and legal frameworks are still evolving to determine how ownership is assigned when AI plays a central role in molecular design. These raise questions about attribution, responsibility, and protection of innovation.

Finally, the lack of harmonized global standards remains a barrier. Regional disparities in classification, approval, and surveillance practices can hinder the efficient development and deployment of AI-enabled pharmaceutical products. Collaborative initiatives and international regulatory convergence will be essential to foster a balanced, globally coherent approach [96].

In summary, the regulatory landscape for AI in drug discovery is rapidly evolving. While substantial challenges persist—such as interpretability, bias, and lack of harmonization—regulatory bodies are increasingly proactive in addressing these through flexible oversight models, stakeholder collaboration, and adaptive guidance frameworks. Continued dialogue between developers, regulators, and academia will be key to ensuring that AI technologies are safely and effectively integrated into the pharmaceutical innovation pipeline.

5. Conclusions and Future Perspectives

These investments are forming a clear vision that AI will play a significant role in the future discovery and development of drugs. The fast advancement of AI (AI) is contributing to innovation across numerous industries, with the pharmaceutical sector being one of the most impacted. AI aims to automate, optimize, and personalize various areas within the pharmaceutical industry, particularly in pharmacological research, which shows the most interest among researchers [28,29].

Virtual screening techniques will help accelerate the drug discovery process by analyzing vast chemical libraries and finding the best therapeutic candidate based on the required characteristics. This process will significantly reduce the time spent discovering potential therapeutic compounds and also cut down on associated costs [13].

Patient categorization through AI will facilitate the development of personalized therapies by analyzing genomes, proteomes, and clinical records. This will allow for the creation of a more effective medication for each specific patient, with fewer adverse reactions and a tailored dose. With AI algorithms, it will be possible to improve therapeutic outcomes by considering factors such as age, weight, genetics, and disease status. Using

devices to collect real-time data from patients will enable the proposal of personalized therapies, helping in the development of suitable and effective medications [13].

Clinical trials recruit volunteer patients to aid in the drug development process. For the results to be as reliable as possible, a patient selection process must be in place. AI will allow access to records of biomarkers, genetic profiles, and patient health status, enabling the identification of suitable patients, thus reducing costs and speeding up the process [13].

However, the continued use of AI in drug formulation raises ethical and regulatory concerns that must be carefully examined. Key issues include the transparency of algorithms, the protection of patient personal data, and ensuring that technologies are used in a fair and safe manner.

As AI evolves and integrates with the pharmaceutical sector, it is crucial that the industry and regulatory bodies work together to harness this technology and minimize potential risks, ensuring that everyone benefits from the proposed innovations in a fair and safe way.

Author Contributions: Conceptualization, A.F. and F.M.-M.; methodology, A.F. and F.M.-M.; validation, A.F. and F.M.-M.; formal analysis, A.F. and F.M.-M.; investigation, V.M. and B.S.; resources and software, V.M. and B.S.; writing—original draft preparation, V.M. and B.S.; writing—review and editing, A.F. and F.M.-M.; supervision, A.F. and F.M.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research is sponsored by national funds through FCT, Fundação para a Ciência e a Tecnologia, under project DOI 10.54499/PTDC/NAN-MAT/1431/2021.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors acknowledge the support from the BRIDGES—Biotechnology Research, Innovation and Design for Health Products, Polytechnic University of Guarda, 6300-559 Guarda, Portugal, and REQUIMTE/LAQV, Group of Pharmaceutical Technology, University of Coimbra, 3000-548 Coimbra, Portugal.

Conflicts of Interest: The authors declare no competing interests in this work.

Abbreviations

ADME	Absorption, Distribution, Metabolism, Excretion
ANN	Artificial Neural Network
DL	Deep Learning
DNNs	Deep Neural Networks
DTBA	Drug–Target Binding Affinity
EPI	Estimation Program Interface
FDA	Food and Drug Administration
R&D	Research and Development
AI	Artificial Intelligence
ML	Machine Learning
MMP	Matched Molecular Pair
NN	Neural Forest
QSPR	Quantitative Structure–Property Relationship
SaMD	Software as a Medical Device
SMILE	Simplified Molecular Input Line Entry System
SVM	Support Vector Machine
WHO	World Health Organization

References

1. Deng, J.; Yang, Z.; Ojima, I.; Samaras, D.; Wang, F. Artificial intelligence in drug discovery: Applications and techniques. *Brief. Bioinform.* **2021**, *23*, bbab430. [CrossRef] [PubMed]
2. Kaul, V.; Enslin, S.; Gross, S.A. History of artificial intelligence in medicine. *Gastrointest. Endosc.* **2020**, *92*, 807–812. [CrossRef] [PubMed]
3. Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H.J. Artificial intelligence in radiology. *Nat. Rev. Cancer* **2018**, *18*, 500–510. [CrossRef] [PubMed]
4. Denecke, K.; Baudoin, C.R. A Review of Artificial Intelligence and Robotics in Transformed Health Ecosystems. *Front. Med.* **2022**, *9*, 795957. [CrossRef]
5. Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R.K. Artificial intelligence in drug discovery and development. *Drug Discov. Today* **2021**, *26*, 80–93. [CrossRef]
6. Wang, L.; Ding, J.; Pan, L.; Cao, D.; Jiang, H.; Ding, X. Artificial intelligence facilitates drug design in the big data era. *Chemom. Intell. Lab. Syst.* **2019**, *194*, 103850. [CrossRef]
7. Grzybowski, A.; Pawlikowska-Łagód, K.; Lambert, W.C. A History of Artificial Intelligence. *Clin. Dermatol.* **2024**, *42*, 221–229. [CrossRef]
8. Bredt, S. Artificial Intelligence (AI) in the Financial Sector—Potential and Public Strategies. *Front. Artif. Intell.* **2019**, *2*, 16. [CrossRef]
9. Stahl, B.C.; Antoniou, J.; Ryan, M.; Macnish, K.; Jiya, T. Organisational responses to the ethical issues of artificial intelligence. *AI Soc.* **2022**, *37*, 23–37. [CrossRef]
10. Theodosiou, A.A.; Read, R.C. Artificial intelligence, machine learning and deep learning: Potential resources for the infection clinician. *J. Infect.* **2023**, *87*, 287–294. [CrossRef]
11. Gawehn, E.; Hiss, J.A.; Schneider, G. Deep Learning in Drug Discovery. *Mol. Inform.* **2016**, *35*, 3–14. [CrossRef] [PubMed]
12. Mak, K.-K.; Pichika, M.R. Artificial intelligence in drug development: Present status and future prospects. *Drug Discov. Today* **2019**, *24*, 773–780. [CrossRef] [PubMed]
13. Vora, L.K.; Gholap, A.D.; Jetha, K.; Thakur, R.R.S.; Solanki, H.K.; Chavda, V.P. Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design. *Pharmaceutics* **2023**, *15*, 1916. [CrossRef] [PubMed]
14. Choi, R.Y.; Coyner, A.S.; Kalpathy-Cramer, J.; Chiang, M.F.; Campbell, J.P. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl. Vis. Sci. Technol.* **2020**, *9*, 14.
15. Cabrera, A.; Bouterse, A.; Nelson, M.; Razzouk, J.; Ramos, O.; Chung, D.; Cheng, W.; Danisa, O. Use of random forest machine learning algorithm to predict short term outcomes following posterior cervical decompression with instrumented fusion. *J. Clin. Neurosci.* **2023**, *107*, 167–171. [CrossRef]
16. Rebollo, R.; Oyoun, F.; Corvis, Y.; El-Hammadi, M.M.; Saubamea, B.; Andrieux, K.; Mignet, N.; Alhareth, K. Microfluidic Manufacturing of Liposomes: Development and Optimization by Design of Experiment and Machine Learning. *ACS Appl. Mater. Interfaces* **2022**, *14*, 39736–39745. [CrossRef]
17. Lin, E.; Lin, C.-H.; Lane, H.-Y. De Novo Peptide and Protein Design Using Generative Adversarial Networks: An Update. *J. Chem. Inf. Model.* **2022**, *62*, 761–774. [CrossRef]
18. Gupta, R.R. Application of Artificial Intelligence and Machine Learning in Drug Discovery. In *Artificial Intelligence in Drug Design*; Heifetz, A., Ed.; Springer: New York, NY, USA, 2022; pp. 113–124.
19. Svensson, H.G.; Tyrchan, C.; Engkvist, O.; Chehreghani, M.H. Utilizing reinforcement learning for de novo drug design. *Mach. Learn.* **2024**, *113*, 4811–4843. [CrossRef]
20. Pearson, G.S. Artificial Intelligence and Publication Ethics. *J. Am. Psychiatr. Nurses Assoc.* **2024**, *30*, 453–455. [CrossRef]
21. Rogers, W.A.; Draper, H.; Carter, S.M. Evaluation of artificial intelligence clinical applications: Detailed case analyses show value of healthcare ethics approach in identifying patient care issues. *Bioethics* **2021**, *35*, 623–633. [CrossRef]
22. Nagarajan, N.; Yapp, E.K.Y.; Le, N.Q.K.; Kamaraj, B.; Al-Subaie, A.M.; Yeh, H.-Y. Application of Computational Biology and Artificial Intelligence Technologies in Cancer Precision Drug Discovery. *BioMed Res. Int.* **2019**, *2019*, 8427042. [CrossRef] [PubMed]
23. Zhang, Y.; Luo, M.; Wu, P.; Wu, S.; Lee, T.-Y.; Bai, C. Application of Computational Biology and Artificial Intelligence in Drug Design. *Int. J. Mol. Sci.* **2022**, *23*, 13568. [CrossRef]
24. Deore, A.B.; Dhumane, J.R.; Wagh, R.; Sonawane, R. The Stages of Drug Discovery and Development Process. *Asian J. Pharm. Res. Dev.* **2019**, *7*, 62–67. [CrossRef]
25. Parvathaneni, V.; Kulkarni, N.S.; Muth, A.; Gupta, V. Drug repurposing: A promising tool to accelerate the drug discovery process. *Drug Discov. Today* **2019**, *24*, 2076–2085. [CrossRef]
26. Dana, D.; Gadhiya, S.V.; St. Surin, L.G.; Li, D.; Naaz, F.; Ali, Q.; Paka, L.; Yamin, M.A.; Narayan, M.; Goldberg, I.D.; et al. Deep Learning in Drug Discovery and Medicine; Scratching the Surface. *Molecules* **2018**, *23*, 2384. [CrossRef] [PubMed]

27. Sarkar, C.; Das, B.; Rawat, V.S.; Wahlang, J.B.; Nongpiur, A.; Tiewsoh, I.; Lyngdoh, N.M.; Das, D.; Bidarolli, M.; Sony, H.T. Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development. *Int. J. Mol. Sci.* **2023**, *24*, 2026. [CrossRef] [PubMed]
28. Ciallella, H.L.; Zhu, H. Advancing Computational Toxicology in the Big Data Era by Artificial Intelligence: Data-Driven and Mechanism-Driven Modeling for Chemical Toxicity. *Chem. Res. Toxicol.* **2019**, *32*, 536–547. [CrossRef]
29. Parvatikar, P.P.; Patil, S.; Khaparkhuntikar, K.; Patil, S.; Singh, P.K.; Sahana, R.; Kulkarni, R.V.; Raghu, A.V. Artificial intelligence: Machine learning approach for screening large database and drug discovery. *Antivir. Res.* **2023**, *220*, 105740. [CrossRef]
30. Singh, S.; Gupta, H.; Sharma, P.; Sahi, S. Advances in Artificial Intelligence (AI)-assisted approaches in drug screening. *Artif. Intell. Chem.* **2024**, *2*, 100039. [CrossRef]
31. Spanakis, M.; Tzamali, E.; Tzedakis, G.; Koumpouzi, C.; Pediaditis, M.; Tsatsakis, A.; Sakkalis, V. Artificial Intelligence Models and Tools for the Assessment of Drug–Herb Interactions. *Pharmaceuticals* **2025**, *18*, 282. [CrossRef]
32. Li, X.; Liang, J.; Hu, J.; Ma, L.; Yang, J.; Zhang, A.; Jing, Y.; Song, Y.; Yang, Y.; Feng, Z.; et al. Screening for primary aldosteronism on and off interfering medications. *Endocrine* **2024**, *83*, 178–187. [CrossRef]
33. Rocco, P.; Musazzi, U.M.; Franzè, S.; Minghetti, P. Copies of nonbiological complex drugs: Generic, hybrid or biosimilar? *Drug Discov Today* **2019**, *24*, 250–255. [CrossRef] [PubMed]
34. Zagalo, D.M.; Sousa, J.; Simões, S. Quality by design (QbD) approach in marketing authorization procedures of Non-Biological Complex Drugs: A critical evaluation. *Eur. J. Pharm. Biopharm.* **2022**, *178*, 1–24. [CrossRef]
35. Saha, E.; Rathore, P.; Parida, R.; Rana, N.P. The interplay of emerging technologies in pharmaceutical supply chain performance: An empirical investigation for the rise of Pharma 4.0. *Technol. Forecast. Soc. Change* **2022**, *181*, 121768. [CrossRef]
36. Malheiro, V.; Duarte, J.; Veiga, F.; Mascarenhas-Melo, F. Exploiting Pharma 4.0 Technologies in the Non-Biological Complex Drugs Manufacturing: Innovations and Implications. *Pharmaceutics* **2023**, *15*, 2545. [CrossRef]
37. Chan, H.C.S.; Shan, H.; Dahoun, T.; Vogel, H.; Yuan, S. Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol. Sci.* **2019**, *40*, 592–604. [CrossRef]
38. Shiammala, P.N.; Duraimutharasan, N.K.B.; Vaseeharan, B.; Alothaim, A.S.; Al-Malki, E.S.; Snekaa, B.; Safi, S.Z.; Singh, S.K.; Velmurugan, D.; Selvaraj, C. Exploring the artificial intelligence and machine learning models in the context of drug design difficulties and future potential for the pharmaceutical sectors. *Methods* **2023**, *219*, 82–94. [CrossRef] [PubMed]
39. Singh, S.; Kaur, N.; Gehlot, A. Application of artificial intelligence in drug design: A review. *Comput. Biol. Med.* **2024**, *179*, 108810. [CrossRef]
40. Rashid, M.B.M.A. Artificial Intelligence Effecting a Paradigm Shift in Drug Development. *SLAS Technol.* **2021**, *26*, 3–15. [CrossRef]
41. Doytchinova, I. Drug Design—Past, Present, Future. *Molecules* **2022**, *27*, 1496. [CrossRef]
42. Gholap, A.D.; Uddin, J.; Faiyazuddin; Omri, A.; Gowri, S.; Khalid, M. Advances in artificial intelligence for drug delivery and development: A comprehensive review. *Comput. Biol. Med.* **2024**, *178*, 108702. [CrossRef] [PubMed]
43. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef] [PubMed]
44. Hu, W.; Ohue, M. SpatialPPI: Three-dimensional space protein-protein interaction prediction with AlphaFold Multimer. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 1214–1225. [CrossRef] [PubMed]
45. Ahmed, F.; Samantasinghar, A.; Soomro, A.M.; Kim, S.; Choi, K.H. A systematic review of computational approaches to understand cancer biology for informed drug repurposing. *J. Biomed. Inform.* **2023**, *142*, 104373. [CrossRef]
46. Park, K. A review of computational drug repurposing. *Transl. Clin. Pharmacol.* **2019**, *27*, 59–63. [CrossRef]
47. Issa, N.T.; Stathias, V.; Schürer, S.; Dakshanamurthy, S. Machine and deep learning approaches for cancer drug repurposing. *Semin. Cancer Biol.* **2021**, *68*, 132–142. [CrossRef]
48. Bajorath, J. Origins and progression of the polypharmacology concept in drug discovery. *Artif. Intell. Life Sci.* **2024**, *5*, 100094. [CrossRef]
49. Cichońska, A.; Ravikumar, B.; Rahman, R. AI for targeted polypharmacology: The next frontier in drug discovery. *Curr. Opin. Struct. Biol.* **2024**, *84*, 102771. [CrossRef]
50. Raman, K.; Kumar, R.; Musante, C.J.; Madhavan, S. Integrating Model-Informed Drug Development With AI: A Synergistic Approach to Accelerating Pharmaceutical Innovation. *Clin. Transl. Sci.* **2025**, *18*, e70124. [CrossRef]
51. Peters, S.A. Applications of Physiologically Based Pharmacokinetic Models Integrated with Drug Effect models (PBPK/PD). In *Physiologically Based Pharmacokinetic (PBPK) Modeling and Simulations: Principles, Methods, and Applications in the Pharmaceutical Industry, Second Edition*; John Wiley & Sons: Hoboken, NJ, USA, 2021; pp. 353–374.
52. Saeheng, T.; Karbwang, J.; Na-Bangchang, K. Population-pharmacokinetic/pharmacodynamic model of atractylodes lancea (Thunb.) DC. administration in patients with advanced-stage intrahepatic cholangiocarcinoma: A dosage prediction. *BMC Complement. Med. Ther.* **2024**, *24*, 384. [CrossRef]

53. Chou, W.-C.; Chen, Q.; Yuan, L.; Cheng, Y.-H.; He, C.; Monteiro-Riviere, N.A.; Riviere, J.E.; Lin, Z. An artificial intelligence-assisted physiologically-based pharmacokinetic model to predict nanoparticle delivery to tumors in mice. *J. Control Release* **2023**, *361*, 53–63. [CrossRef]
54. Chou, W.-C.; Lin, Z. Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling. *Toxicol. Sci.* **2022**, *191*, 1–14. [CrossRef]
55. Li, Y.; Wang, Z.; Li, Y.; Du, J.; Gao, X.; Li, Y.; Lai, L. A Combination of Machine Learning and PBPK Modeling Approach for Pharmacokinetics Prediction of Small Molecules in Humans. *Pharm. Res.* **2024**, *41*, 1369–1379. [CrossRef]
56. Wu, K.; Li, X.; Zhou, Z.; Zhao, Y.; Su, M.; Cheng, Z.; Wu, X.; Huang, Z.; Jin, X.; Li, J.; et al. Predicting pharmacodynamic effects through early drug discovery with artificial intelligence-physiologically based pharmacokinetic (AI-PBPK) modelling. *Front. Pharmacol.* **2024**, *15*, 1330855. [CrossRef]
57. Umscheid, C.A.; Margolis, D.J.; Grossman, C.E. Key Concepts of Clinical Trials: A Narrative Review. *Postgrad. Med.* **2011**, *123*, 194–204. [CrossRef] [PubMed]
58. Fogel, D.B. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp. Clin. Trials Commun.* **2018**, *11*, 156–164. [CrossRef] [PubMed]
59. Cunningham, J.W.; Abraham, W.T.; Bhatt, A.S.; Dunn, J.; Felker, G.M.; Jain, S.S.; Lindsell, C.J.; Mace, M.; Martyn, T.; Shah, R.U.; et al. Artificial Intelligence in Cardiovascular Clinical Trials. *JACC* **2024**, *84*, 2051–2062. [CrossRef] [PubMed]
60. Chopra, H.; Annu; Kil Shin, D.; Munjal, K.; Priyanka; Dhama, K.; Bin Emran, T. Revolutionizing clinical trials: The role of AI in accelerating medical breakthroughs. *Int. J. Surg.* **2023**, *109*, 4211–4220. [CrossRef]
61. Zhang, Y.; Mastouri, M.; Zhang, Y. Accelerating drug discovery, development, and clinical trials by artificial intelligence. *Medicine* **2024**, *5*, 1050–1070. [CrossRef]
62. Goldberg, J.M.; Amin, N.P.; Zachariah, K.A.; Bhatt, A.B. The Introduction of AI Into Decentralized Clinical Trials. *JACC Adv.* **2024**, *3*, 101094. [CrossRef]
63. Lu, X.; Yang, C.; Liang, L.; Hu, G.; Zhong, Z.; Jiang, Z. Artificial intelligence for optimizing recruitment and retention in clinical trials: A scoping review. *J. Am. Med. Inform. Assoc.* **2024**, *31*, 2749–2759. [CrossRef] [PubMed]
64. Fountzilas, E.; Tsimberidou, A.M.; Vo, H.H.; Kurzrock, R. Clinical trial design in the era of precision medicine. *Genome Med.* **2022**, *14*, 101. [CrossRef]
65. Jayatunga, M.K.; Ayers, M.; Bruens, L.; Jayanth, D.; Meier, C. How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons. *Drug Discov. Today* **2024**, *29*, 104009. [CrossRef] [PubMed]
66. Askin, S.; Burkhalter, D.; Calado, G.; El Dakrouni, S. Artificial Intelligence Applied to clinical trials: Opportunities and challenges. *Health Technol.* **2023**, *13*, 203–213. [CrossRef]
67. Wang, J.; Zeng, Z.; Li, Z.; Liu, G.; Zhang, S.; Luo, C.; Hu, S.; Wan, S.; Zhao, L. The clinical application of artificial intelligence in cancer precision treatment. *J. Transl. Med.* **2025**, *23*, 120. [CrossRef]
68. Salas, M.; Petracek, J.; Yalamanchili, P.; Aimer, O.; Kasthuril, D.; Dhingra, S.; Junaid, T.; Bostic, T. The Use of Artificial Intelligence in Pharmacovigilance: A Systematic Review of the Literature. *Pharm. Med.* **2022**, *36*, 295–306. [CrossRef]
69. Joshi, P.; Masilamani, V.; Mukherjee, A. A knowledge graph embedding based approach to predict the adverse drug reactions using a deep neural network. *J. Biomed. Inform.* **2022**, *132*, 104122. [CrossRef] [PubMed]
70. Syrowatka, A.; Song, W.; Amato, M.G.; Foer, D.; Edrees, H.; Co, Z.; Kuznetsova, M.; Dulgarian, S.; Seger, D.L.; Simona, A.; et al. Key use cases for artificial intelligence to reduce the frequency of adverse drug events: A scoping review. *Lancet Digit. Health* **2022**, *4*, e137–e148. [CrossRef]
71. Widera, P.; Welsing, P.M.; Danso, S.O.; Peelen, S.; Kloppenburg, M.; Loef, M.; Marijnissen, A.C.; van Helvoort, E.M.; Blanco, F.J.; Magalhães, J.; et al. Development and validation of a machine learning-supported strategy of patient selection for osteoarthritis clinical trials: The IMI-APPROACH study. *Osteoarthr. Cartil. Open* **2023**, *5*, 100406. [CrossRef]
72. Cook, S.F.; Rhodes, T.; Schlusser, C.; Han, S.; Chen, C.; Zach, N.; Murthy, V.; Davé, S. A Descriptive Review of Global Real World Evidence Efforts to Advance Drug Discovery and Clinical Development in Amyotrophic Lateral Sclerosis. *Front. Neurol.* **2021**, *12*, 770001. [CrossRef]
73. Singh, G.; Schulthess, D.; Hughes, N.; Vannieuwenhuyse, B.; Kalra, D. Real world big data for clinical research and drug development. *Drug Discov. Today* **2018**, *23*, 652–660. [CrossRef] [PubMed]
74. Dou, L.; Xu, Z.; Xu, J.; Zang, C.; Su, C.; Pieper, A.A.; Leverenz, J.B.; Wang, F.; Zhu, X.; Cummings, J.; et al. A network-based systems genetics framework identifies pathobiology and drug repurposing in Parkinson's disease. *NPJ Park. Dis.* **2025**, *11*, 22. [CrossRef]
75. Zhang, W.; Kuang, Z.; Peissig, P.; Page, D. Adverse drug reaction discovery from electronic health records with deep neural networks. In Proceedings of the ACM Conference on Health, Inference, and Learning CHIL '20, Toronto, ON, Canada, 2–4 April 2020; pp. 30–39.

76. Denny, J.C.; Van Driest, S.L.; Wei, W.; Roden, D.M. The Influence of Big (Clinical) Data and Genomics on Precision Medicine and Drug Development. *Clin. Pharmacol. Ther.* **2018**, *103*, 409–418. [CrossRef] [PubMed]
77. Nagalakshmi, R.; Khan, S.B.; Kumar, A.V.; R, M.T.; Alojail, M.; Sangwan, S.R.; Saraee, M. Enhancing drug discovery and patient care through advanced analytics with the power of NLP and machine learning in pharmaceutical data interpretation. *SLAS Technol.* **2025**, *31*, 100238.
78. Liu, Z.; Roberts, R.A.; Lal-Nag, M.; Chen, X.; Huang, R.; Tong, W. AI-based language models powering drug discovery and development. *Drug Discov. Today* **2021**, *26*, 2593–2607. [CrossRef]
79. Liu, J.; Wang, X.; Ye, X.; Chen, D. Improved health outcomes of nasopharyngeal carcinoma patients 3 years after treatment by the AI-assisted home enteral nutrition management. *Front. Nutr.* **2025**, *11*, 1481073. [CrossRef]
80. Śledź, P.; Caflisch, A. Protein structure-based drug design: From docking to molecular dynamics. *Curr. Opin. Struct. Biol.* **2018**, *48*, 93–102. [CrossRef]
81. Farghali, H.; Kutinová Canová, N.; Arora, M. The potential applications of artificial intelligence in drug discovery and development. *Physiol. Res.* **2021**, *70* (Suppl. S4), S715–S722. [CrossRef] [PubMed]
82. Burki, T. A new paradigm for drug development. *Lancet Digit. Health* **2020**, *2*, e226–e227. [CrossRef]
83. Soni, K.; Hasija, Y. Artificial Intelligence Assisted Drug Research and Development. In Proceedings of the 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 11–13 February 2022.
84. Richardson, P.; Griffin, I.; Tucker, C.; Smith, D.; Oechsle, O.; Phelan, A.; Rawling, M.; Savory, E.; Stebbing, J. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet* **2020**, *395*, e30–e31. [CrossRef]
85. Cesaro, A.; de la Fuente-Nunez, C. Antibiotic identified by AI. *Nat. Chem. Biol.* **2023**, *19*, 1296–1298. [CrossRef]
86. Gupta, R.; Srivastava, D.; Sahu, M.; Tiwari, S.; Ambasta, R.K.; Kumar, P. Artificial intelligence to deep learning: Machine intelligence approach for drug discovery. *Mol. Divers.* **2021**, *25*, 1315–1360. [CrossRef]
87. Ren, F.; Aliper, A.; Chen, J.; Zhao, H.; Rao, S.; Kuppe, C.; Ozerov, I.V.; Zhang, M.; Witte, K.; Kruse, C.; et al. A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nat. Biotechnol.* **2025**, *43*, 63–75. [CrossRef]
88. Stokes, J.M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N.M.; Macnair, C.R.; French, S.; Carfrae, L.A.; Bloom-Ackermann, Z.; et al. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *180*, 688–702.e13. [CrossRef]
89. Nussinov, R.; Zhang, M.; Liu, Y.; Jang, H. AlphaFold, allosteric, and orthosteric drug discovery: Ways forward. *Drug Discov. Today* **2023**, *28*, 103551. [CrossRef]
90. Guo, S.-B.; Meng, Y.; Lin, L.; Zhou, Z.-Z.; Li, H.-L.; Tian, X.-P.; Huang, W.-J. Artificial intelligence alphafold model for molecular biology and drug discovery: A machine-learning-driven informatics investigation. *Mol. Cancer* **2024**, *23*, 223. [CrossRef]
91. Saber-Ayad, M.; Hammoudeh, S.; Abu-Gharbieh, E.; Hamoudi, R.; Tarazi, H.; Al-Tel, T.H.; Hamid, Q. Current Status of Baricitinib as a Repurposed Therapy for COVID-19. *Pharmaceuticals* **2021**, *14*, 680. [CrossRef]
92. Saito, R.; Yano, N.; Kojima, S.; Miyoshi, F. Data-driven drug discovery for drug repurposing. *Nihon Yakurigaku Zasshi* **2023**, *158*, 10–14. [CrossRef]
93. Program, T.A.A.; Wallach, I.; Bernard, D.; Nguyen, K.; Ho, G.; Morrison, A.; Stecula, A.; Rosnik, A.; O'sullivan, A.M.; Davtyan, A.; et al. AI is a viable alternative to high throughput screening: A 318-target study. *Sci. Rep.* **2024**, *14*, 7526.
94. Chen, J.; Bolhuis, D.L.; Laggner, C.; Kong, D.; Yu, L.; Wang, X.; Emanuele, M.J.; Brown, N.G.; Liu, P. AtomNet-Aided OTUD7B Inhibitor Discovery and Validation. *Cancers* **2023**, *15*, 517. [CrossRef]
95. Chen, Y.; Liu, Z.-W.; Qin, Y. inite-Time Topology Identification of Delayed Complex Dynamical Networks and Its Application. *Cyborg Bionic Syst.* **2024**, *5*, 0092. [CrossRef]
96. Zanca, F.; Brusasco, C.; Pesapane, F.; Kwade, Z.; Beckers, R.; Avanzo, M. Regulatory Aspects of the Use of Artificial Intelligence Medical Software. *Semin. Radiat. Oncol.* **2022**, *32*, 432–441. [CrossRef]
97. Mirakhori, F.; Niazi, S.K. Harnessing the AI/ML in Drug and Biological Products Discovery and Development: The Regulatory Perspective. *Pharmaceuticals* **2025**, *18*, 47. [CrossRef]
98. Da Silva, M.; Horsley, T.; Singh, D.; Da Silva, E.; Ly, V.; Thomas, B.; Daniel, R.C.; Chagal-Feferkorn, K.A.; Iantomasi, S.; White, K.; et al. Legal concerns in health-related artificial intelligence: A scoping review protocol. *Syst. Rev.* **2022**, *11*, 123. [CrossRef]
99. Krishnan Ganapathy, M.N. Artificial Intelligence and Healthcare Regulatory and Legal Concerns. *Telehealth Med. Today* **2021**, *6*, 252. [CrossRef]
100. Gerke, S.; Babic, B.; Evgeniou, T.; Cohen, I.G. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *npj Digit. Med.* **2020**, *3*, 53. [CrossRef]
101. Lee, T.T.; Kesselheim, A.S. U.S. Food and Drug Administration Precertification Pilot Program for Digital Health Software: Weighing the Benefits and Risks. *Ann. Intern. Med.* **2018**, *168*, 730–732. [CrossRef]

102. Khan, Z.; Gaidhane, A.M.; Singh, M.; Ganesan, S.; Kaur, M.; Sharma, G.C.; Rani, P.; Sharma, R.; Thapliyal, S.; Kushwaha, M.; et al. Diagnostic Accuracy of IDX-DR for Detecting Diabetic Retinopathy: A Systematic Review and Meta-Analysis. *Am. J. Ophthalmol.* **2025**, *273*, 192–204. [CrossRef]
103. Cheng, J.Y.; Abel, J.T.; Balis, U.G.; McClintock, D.S.; Pantanowitz, L. Challenges in the Development, Deployment, and Regulation of Artificial Intelligence in Anatomic Pathology. *Am. J. Pathol.* **2021**, *191*, 1684–1692. [CrossRef]
104. Lin, W.; Chen, Y.; Unadkat, J.D.; Zhang, X.; Wu, D.; Heimbach, T. Applications, Challenges, and Outlook for PBPK Modeling and Simulation: A Regulatory, Industrial and Academic Perspective. *Pharm. Res.* **2022**, *39*, 1701–1731. [CrossRef]
105. Pappalardo, F.; Russo, G.; Tshinanu, F.M.; Viceconti, M. In silico clinical trials: Concepts and early adoptions. *Brief. Bioinform.* **2018**, *20*, 1699–1708. [CrossRef]
106. Bordukova, M.; Makarov, N.; Rodriguez-Esteban, R.; Schmich, F.; Menden, M.P. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. *Expert Opin. Drug Discov.* **2024**, *19*, 33–42. [CrossRef]
107. Greenleaf, G.; Mowbray, A.; Chung, P. Building sustainable free legal advisory systems: Experiences from the history of AI & law. *Comput. Law Secur. Rev.* **2018**, *34*, 314–326.
108. Askr, H.; Elgeldawi, E.; Ella, H.A.; Elshaier, Y.A.M.M.; Gomaa, M.M.; Hassanien, A.E. Deep learning in drug discovery: An integrative review and future challenges. *Artif. Intell. Rev.* **2023**, *56*, 5975–6037. [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.

pharmaceuticals

MDPI

Review

Advances in Methods for Accurate Prediction of RNA–Small Molecule Binding Sites: From Isolated to AI-Integrated Strategies

Jiaming Gao, Chen Zhuo, Chengwei Zeng, Haoquan Liu and Yunjie Zhao *

Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan 430079, China
* Correspondence: yjzhaowh@ccnu.edu.cn

Abstract: Predicting RNA–small molecule binding sites is essential for developing RNA-targeted drugs. Identifying these sites experimentally is often costly, making computational methods essential for drug discovery. While traditional approaches rely on limited information, recent AI advancements allow the integration of diverse features, improving prediction accuracy. As methods for predicting RNA–small molecule binding sites continue to evolve, this review provides a timely overview of recent developments. It systematically traces the evolution from physics-based, isolated strategies to AI-integrated approaches, explains the fundamental principles behind different features, compares the tendencies of various features between binding and non-binding sites, evaluates the performance of approaches using different feature combinations on various test sets, and outlines remaining opportunities and challenges, offering guidance for researchers aiming for higher prediction accuracy.

Keywords: RNA binding sites prediction; artificial intelligence; machine learning; deep learning

1. Introduction

One of the central tasks in drug discovery is to identify bioactive compounds that selectively bind to disease-relevant targets [1]. Achieving such selective binding requires precise localization of binding sites on the target molecule that can bind to small molecules. Over the past two decades, the development of computer-aided drug design has led to revolutionary advances in protein-targeted drug discovery [2,3]. Benefiting from the progress of artificial intelligence (AI) algorithms, the accuracy of identifying protein-small molecule binding sites has been continuously improved. However, protein-coding genes constitute only ~1.5% of the human genome, and merely 10–15% of these are considered druggable [4,5]. In contrast, more than 95% of the human genome is transcribed to non-coding RNA, which participates in diverse biological processes and offers a vast pool of untapped drug targets [6].

The characteristics of RNA sequences and structures are closely related to the druggability of RNA. RNA is a biomolecule with high sequence specificity. Different RNA sequences exhibit variations in both structure and function, which makes RNA a highly selective target for drugs. Additionally, the unique structure of RNA enables small molecules to bind to and interact with specific regions of RNA, thereby influencing its function. This sequence and structural precision are relevant not only to human cellular RNA but also

to viral RNA. Numerous viruses possess RNA genomes, and small molecules that target RNA can suppress the transcription of RNA viruses, aiding in the prevention of pandemics caused by these viruses, including SARS-CoV-2 [7–10]. Therefore, RNA has attracted increasing research interest as a binding target for small molecules in recent years.

Accurate identification of small molecule binding sites in RNA facilitates the development of RNA-targeted therapeutics and advances the discovery of small-molecule drugs that selectively bind RNA. Small-molecule drugs targeting RNA have achieved breakthrough progress. For example, the synthetic small molecule Risdiplam functions as a splicing modifier of SMN2 pre-mRNA, regulating its alternative splicing and restoring the expression of functional SMN protein. Risdiplam has improved motor function in patients with spinal muscular atrophy, demonstrating that RNA can serve as a druggable target for small molecule [11–13]. Similarly, as a synthetic small molecule, Ribocil can selectively target the FMN riboswitch in bacteria. It mimics the function of the natural ligand FMN, blocking the transcription of downstream riboflavin biosynthesis genes. The success of Ribocil provided cellular proof that small molecule binding to a non-coding RNA structural element is sufficient to elicit antibacterial activity [14–16]. These examples collectively illustrate the druggable potential of RNA as a target for small-molecule therapeutics.

Nevertheless, experimentally screening small molecules that target RNA remains a challenging task. RNA molecules possess a highly charged backbone formed by phosphate groups and exhibit high flexibility and high polarity. The high flexibility of RNA structures can lead to unstable small molecule binding sites in high-throughput screening. In contrast, the high polarity affects binding specificity, increasing the likelihood of false-positive results. These structural characteristics make it challenging to experimentally determine the binding sites of RNA–small molecule complexes. In addition, experimentally determining the binding sites of RNA–small molecule is often time- and resource-consuming. Hence, developing highly accurate computational methods to recognize small molecule binding sites on RNA is vital for the discovery of RNA-targeted small-molecule drugs. With the increasing number of tertiary structures of RNA–small molecule complexes deposited in the Protein Data Bank (PDB), the development of high-precision computational algorithms has become increasingly feasible [17]. Hybrid strategies that combine experimental and predicted structures are also being employed to develop more accurate computational methods [18]. Moreover, RNA large language models (LLMs) effectively leverage RNA sequence data, which far exceeds the amount of RNA structural data. It can complement structure-based models, making it promising to improve the performance of computational methods further.

Numerous computational methods have been developed to predict small molecule binding sites in RNA structures, as shown in Figure 1C. Early methods primarily characterized RNA nucleotides by extracting isolated physics-based information from RNA structures, relying on both RNA tertiary or secondary structures [19–21]. By coarse-graining the structures, these methods distinguished individual nucleotides within the structures to predict potential binding sites. Recently, driven by the breakthroughs of AI algorithms in structure prediction, many researchers have developed machine learning (ML) and deep learning (DL)-based methods that integrate diverse information to identify potential binding patterns of RNA–small molecule interactions [22–30]. Different from previous physics-based methods, most AI-driven approaches leverage information from multiple RNA modalities as complementary features, such as sequence and geometry [31,32].

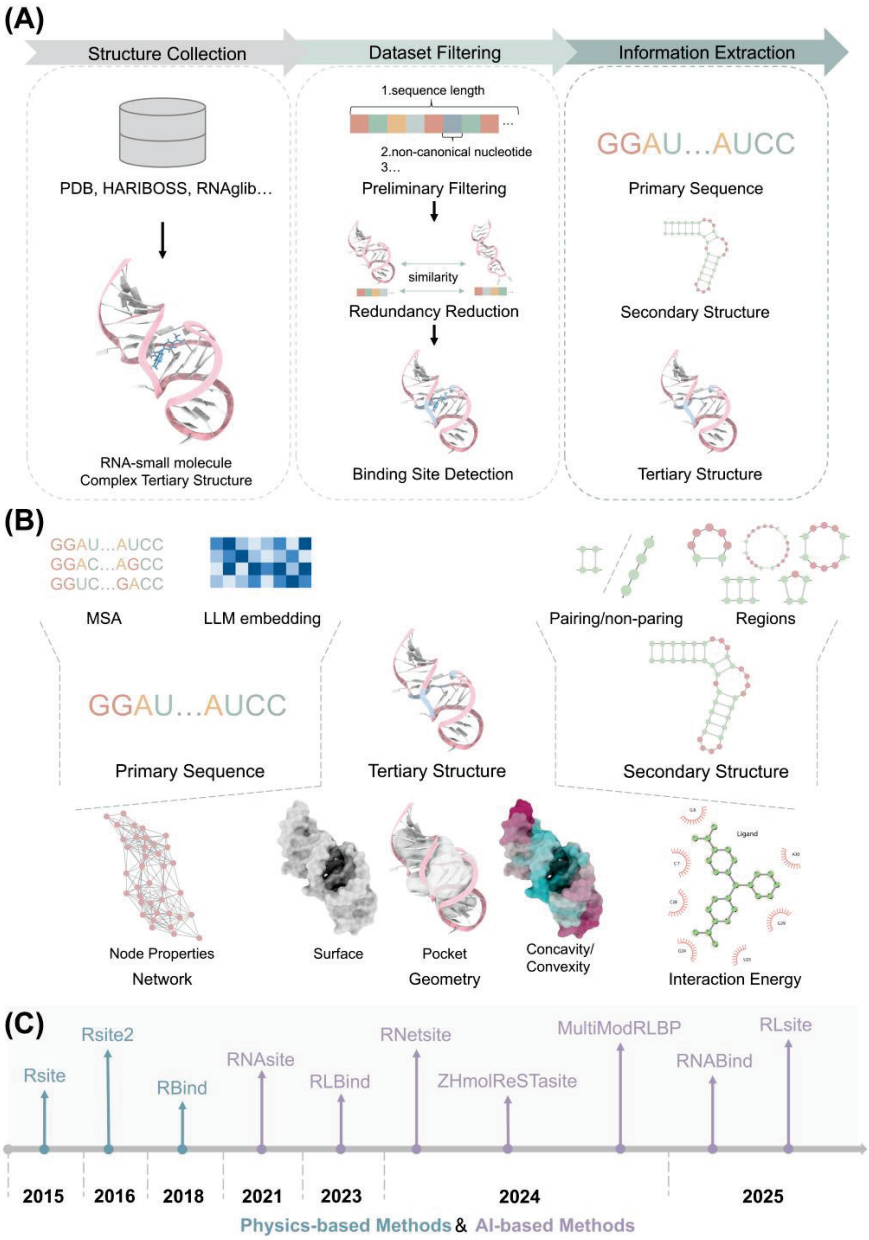


Figure 1. Overview of data preparation for RNA–small molecule binding site prediction. (**A**) Process of data collection and dataset construction including complex structure collection, dataset filtering and information extraction. (**B**) Feature extraction for predicting small molecule binding sites in RNA, incorporating features derived from RNA sequence, secondary structure and tertiary structure. (**C**) Timeline of methods for predicting RNA–small molecule binding sites, including both physics-based approaches and AI-based approaches.

This review offers a thorough and current overview of recent progress in predicting RNA–small molecule binding sites. We begin by outlining the main workflows of different methods and the feature types they utilize, then explain the fundamental principles behind these features. Next, we review how AI-based approaches construct datasets and assess their performance across various test sets, examining how the choice of features influences results. Finally, we discuss the main challenges and future opportunities in this field. Unlike earlier reviews mainly centered on RNA structure or physics-based methods, this review highlights AI-driven integrated strategies that utilize multimodal features, such as large language models, to predict RNA–small molecule binding sites.

2. Methods for Predicting RNA–Small Molecule Binding Sites

RNA–small molecule binding site prediction methods fall into two main categories: physics-based and AI-based. These approaches use different RNA data as inputs and combine various features. Table 1 details the input data, feature combinations, models, and available websites for each method. Below, the core workflows and features of these methods are explained.

Table 1. Methods for predicting RNA-small molecule binding sites. Input refers to the input information of the model, where seq represents sequence and 3D structure refers to tertiary structure. Feature combination refers to the feature types the method utilizes, MSA stands for multiple sequence alignment, SS stands for secondary structure. Model refers to the computing algorithm, where RF refers to Random Forest, CNN refers to Convolutional Neural Network, XGB refers to Extreme Gradient Boosting, LGBM refers to Light Gradient Boosting Machine, ResNet refers to Residual Convolutional Network, RGCN refers to Relational Graph Convolutional Network, EGNN refers to E(3)-Equivariant Graph Neural Network, 3DCNN refers to 3D Convolutional Neural Network, BiLSTM refers to Bidirectional Long Short-term Memory Network. Available records the available source or web server of method. Ref records the paper related to the method.

Name	Input	Feature Combination	Model	Available	Ref.
Rsite	3D structure	3D distance	Distance	http://www.cuilab.cn/rsite (accessed on 20 August 2025)	[19]
Rsite2	seq	2D distance	Distance	https://www.cuilab.cn/rsite2 (accessed on 20 August 2025)	[20]
RBind	3D structure	3D distance	Distance	http://zhaoserver.com.cn/RBinds/RBinds.html (accessed on 20 August 2025)	[21]
RNAsite	seq, 3D structure	MSA, Geometry, Network	RF	https://yanglab.qd.sdu.edu.cn/RNAsite/ (accessed on 20 August 2025)	[22]
RLBind	seq, 3D structure	MSA, Geometry, Network	CNN	https://github.com/KailiWang1/RLBind (accessed on 20 August 2025)	[23]
RNetsite	3D structure	Network	RF, XGB, LGBM	http://zhaoserver.com.cn/RNet/RNet.html (accessed on 20 August 2025)	[24]
ZHmolReSTasite	seq, 3D structure	MSA, SS, Geometry, Network	ResNet	http://zhaoserver.com.cn/ZHmol/ZHmolReSTasite/ZHmolReSTasite.html (accessed on 20 August 2025)	[25]
MultiModRLBP	seq, 3D structure	LLM, Geometry, Network	CNN, RGCN	https://github.com/lennylv/MultiModRLBP (accessed on 20 August 2025)	[26]
RNABind	seq, 3D structure	LLM	EGNN	https://github.com/jaminzzz/RNABind (accessed on 20 August 2025)	[27]
RLsite	seq, 3D structure	MSA, Geometry, Energy	3DCNN, BiLSTM	https://github.com/fine1231/RLsite (accessed on 20 August 2025)	[28]

2.1. Physics-Based Methods

Rsite. Rsite is a tertiary structure-based method to predict the functional sites of RNA [19]. Drawing inspiration from protein function prediction techniques, Rsite identified two key points: (1) Nucleotides with high closeness centrality in RNA tertiary structures tend to be functionally significant. (2) Surface-exposed nucleotides are more likely to engage in recognition and RNA interactions. Based on these insights, Rsite hypothesized that in non-coding RNA structures, both highly connected and less connected nucleotides are often functionally important, indicating potential functional sites.

Based on this hypothesis, Rsite recognize functional sites as following steps: (1) for a given RNA tertiary structure with n nucleotides, Rsite calculates the Euclidean distance between every pair of nucleotides, yielding $n \times (n-1)$ distance values. (2) for each nucleotide, Rsite calculates the sum distances to the other (n-1) nucleotides. For an RNA tertiary structure with n nucleotides, generate a distance curve comprising n data points. (3) a Gaussian filter is applied to smooth the distance values for each nucleotide. Specifically,

Rsite applies a sliding window of size w to each nucleotide, computing the average distance values within the window:

$$S_i = \frac{D_{i-w} + \cdots + D_i + \cdots + D_{i-w}}{2w + 1} \quad (1)$$

where D_i denotes the sum of distances between i-th nucleotide and all other nucleotides, w denotes the window size and S_i corresponds to the smoothed distance value for i-th nucleotide. Rsite implemented a window size of 2 in their algorithms. (4) on the smoothed distance curve, local minima and maxima are identified as putative RNA functional sites. In addition, both start and end points are considered functional sites if their distances deviate by more than 50% from the mean distance.

Rsite2. RNA's secondary structure is simpler to obtain than its tertiary structure, which makes analyzing RNA functional sites based on secondary structure more efficient. Building on this concept, Zeng et al. [20] examined RNA secondary structure and discovered that, in non-coding RNAs, the distance from a nucleotide to the centroid in the secondary structure is strongly positively correlated with its distance to the centroid in the tertiary structure. This indicates the potential for predicting functional sites using distance metrics derived from secondary structure. Based on this, Zeng et al. introduced Rsite2, a computational method for identifying RNA functional sites from secondary structure. Rsite2 takes an RNA sequence as input, predicts its secondary structure with RNAfold, and extracts the 2D coordinates of each nucleotide from the RNAfold output [33]. Rsite2 introduces two metrics to illustrate the position of each nucleotide in the RNA: the nucleotide-to-centroid distance, indicating the Euclidean distance between a nucleotide and the centroid, and the nucleotide distance sum, which totals the distances from that nucleotide to all others. Rsite2 computes these two metrics and uses the same algorithm as Rsite to predict RNA functional sites: for an RNA composed of n nucleotides, a distance curve of length n is plotted, where position i represents the i-th nucleotide. The curve is smoothed using a Gaussian filter, and nucleotides located at the extrema of the curve are identified as functional sites.

RBind. Wang et al. observed that while Rsite and Rsite2 can predict functional sites solely based on distances from RNA secondary and tertiary structures, they may generate false positives because they do not differentiate between strong and reliable connections. To improve this, Wang et al. created RBind, a computational approach that converts RNA tertiary structures into network representations [21]. Given an RNA tertiary structure, RBind treats nucleotides as nodes and their non-covalent interactions as edges to build an undirected network. Each nucleotide is characterized by network properties. An edge is added between two non-adjacent nucleotides if any pair of their heavy atoms is within 8 Å. RBind then calculates two network metrics, degree and closeness, to identify functional sites. A nucleotide is considered a functional site if both its degree and closeness are more than one standard deviation above the average values for all nodes.

2.2. AI-Based Methods

RNAsite. RNAsite is a machine learning method based on RNA sequences and tertiary structures, composed of two modules: sequence-based RNAsite_seq and tertiary structure-based RNAsite_str [22]. RNAsite_seq utilizes RNA sequences as input, performs BLASTN searches against the NCBI nucleotide database to construct multiple sequence alignments (MSA) [32]. The results of MSA are then employed to compute evolutionary conservation scores for each position in the RNA sequence. For each nucleotide, RNAsite_seq considers a window size of 12 to take its sequence context neighbors from a total of $(2 \times 12 + 1)$ nucleotides into account, then uses a Random Forest algorithm to implement training and

predicting processes. RNAsite_str utilizes RNA tertiary structure as input, calculates four structure-based features: accessible surface areas (ASA), Laplacian norm (LN), degree, and closeness [34,35]. Similarly to RNAsite_seq, RNAsite_str uses a window size of 14 to capture the context neighbors of target nucleotides, and then the structure-based features of each nucleotide are sent to a second Random Forest model. Additionally, using a window size of 16, the prediction results from RNAsite_seq were integrated with structure-based features, and the combined features were used to train the final Random Forest model in RNAsite. RNAsite combined RNA sequence and structural information, outperformed the sequence-only module RNAsite_seq and the structure-only module RNAsite_str.

RNetsite. Liu et al. proposed RNetsite, a machine learning method for predicting small molecule binding sites by modeling RNA tertiary structures as networks [24]. RNetsite converts RNA tertiary structures into networks, with nucleotides as nodes and non-covalent interactions as edges. An undirected edge connects two nucleotides if any pair of their heavy atoms are within 8 Å. RNetsite uses network analysis and machine learning to predict small molecule binding sites. The process involves calculating network properties for each nucleotide, including two local features (degree and neighborhood connectivity) and three global features (closeness, betweenness centrality, and eccentricity). To capture the influence of surrounding nucleotides, RNetsite employs a sliding window of size w, encoding each nucleotide's features by combining network properties from its 2w + 1 neighborhood. These features are then used to train three machine learning models: Random Forest (RF), Light Gradient Boosting Machine (LGBM), and Extreme Gradient Boosting (XGB). Predictions from the three models are combined using a voting strategy, where the final prediction for each nucleotide is made when at least two models agree.

RLBind. RLBind is a deep learning method that utilizes convolutional neural networks to extract both global and local information from RNA sequences and tertiary structures, capturing long-range and short-range interactions, respectively [23]. The method integrated nucleotide types and evolutionary conservations as sequence-based features, with the evolutionary conservations collected by multiple sequence alignment information of RNA sequences, acquired through BLASTN, into ConSurf-DB [32,36,37]. From RNA tertiary structures, two network features (degree, closeness), accessible surface areas and two biochemical properties (molecular mass, side-chain pka) were derived as structural features. RLBind employed a dual-channel architecture: a local channel that used an 11-nucleotide sliding window to capture the context of each nucleotide, and a global channel that processed fixed length (64-nt) RNA sequences. Features from the global channel were initially processed through a convolutional layer, then combined with those from the local channel, and finally passed into two dense layers. The RLBind prediction was produced by applying a sigmoid activation function to the output of these dense layers.

ZHmolReSTasite. Gao et al. developed a deep learning method, ZHmolReSTasite (ZeSTa for short), using RNA surface topography strategy to predict RNA–small molecule binding sites [25]. For an RNA tertiary structure, ZeSTa employed RNA surface topography as follows: the first step was to calculate the solvent-excluded surface and convert it into a cloud of vertices with 3D coordinates. Coordinates conversion was then applied to each nucleotide by rotating the RNA surface to align the geometric centers of nucleotides with the intersection of the equator and the prime meridian of the RNA surface. Three-dimensional coordinates of vertices were transferred to longitude and latitude coordinates, and vertices located within a 12Å geodesic radius from each nucleotide's geometric center were extracted to represent the nucleotide. An interpolation method is then utilized to obtain regular topography images of nucleotides. The topography images were characterized based on RNA sequences, secondary structures, and tertiary structures. ZeSTa obtained multiple

sequence alignment (MSA) for each RNA sequence by BLASTN searches against the NCBI nucleotide database. The results of MSA were utilized by the plmDCA method to analyze the coevolutionary relationship between nucleotides [32,38,39]. Secondary structure type information was derived from tertiary structures using RNAstat [40]. The tertiary structure features utilized by ZeSTa included three geometric-based features (SASA, LN, and pocket) and two network-based features (degree and closeness), as well as pKa values, which were incorporated based on nucleotide types. These features were added to the corresponding vertices of each nucleotide in the topography image, generating feature maps of nucleotides. ZeSTa then fed the nucleotide feature maps into a deep learning algorithm, ResNet18, for prediction in the RNA–small molecule binding site task [41].

MultiModRLBP. MultiModRLBP is a multimodal deep learning framework developed by Wang et al. [26]. The core concept involves encoding nucleotides through three perspectives: RNA structure relationship graph, nucleotide structural properties, and sequence semantics. Specifically, the method utilizes x3dna-dssr to extract two types of information from RNA tertiary structures: DSSR adjacency annotation and DSSR properties [42]. The DSSR adjacency annotation describes nucleotide neighborhoods based on the Leontis-Westhof system, where each nucleotide serves as a node, and interactions are categorized into 20 types of RNA base pairs. Network metrics like degree and closeness are derived from this annotation. DSSR also provides structural details such as torsion angles and key atomic coordinates. MultiModRLBP integrates these networks and DSSR properties with three types of accessible surface area and base encoding, resulting in 71-dimensional features per nucleotide. The process involves three components: (1) RNA graph. RGCN learns the RNA structural graph, with nodes representing nucleotides with 71D features and edges indicating Leontis-Westhof base interactions. (2) Nucleotide. The ELBFS module combines local features from a sliding window, global features from CNN, and individual 71D features to produce ELBFS features. (3) RNA semantic sequence. RNABERT, a pre-trained LLM for RNA sequences, is utilized to generate context-sensitive distributed word representations for each nucleotide [43]. Features from these three components are concatenated and fed into dense layers, producing prediction results of MultiModRLBP.

RNABind. Building upon the remarkable advances of LLMs in protein and DNA research, Zhu et al. [27] developed RNABind, a deep learning framework that integrates RNA LLMs with graph representations of RNA tertiary structures. Unlike the above methods, RNABind exclusively utilizes vector representations of nucleotides derived from RNA LLMs to characterize the nucleotide context. These context vector representations are used as node features in a network based on RNA tertiary structure. In this network, nodes are defined as the centroid of each nucleotide, with coordinates defined as the average of all atomic coordinates of the nucleotide. Nodes are connected to their 15 nearest neighbors via edges using a Euclidean distance-based k-NN strategy, with edge features consisting of distance encoded by 16 radial basis functions (RBFs). RNABind inputs node features, edge features, and nucleotide centroid coordinates into E(3)-equivariant graph neural networks (EGNNs) [44]. It processes the RNA graph representations through a stack of three equivariant graph convolutional layers (EGCLs), followed by a predictor that performs per-nucleotide classification to generate RNABind's predictions. RNABind employs eight single-nucleotide resolution RNA LLMs and compares their performance in predicting RNA–small molecule binding sites [43,45–51]. The LLMs ERNIE-RNA and RiNALMo demonstrated superior performance across multiple evaluation metrics and were therefore selected as the primary models for reporting in RNABind.

RLsite. Zou et al. suggested that interaction energies are crucial for identifying nucleotides involved in small-molecule binding, including π-π stacking interactions, electrostatic interac-

tions, and hydrogen-bond interactions [28]. To achieve this, they introduced RLSite, a deep learning method that integrates nucleotide interaction energy with structural and sequence information. RLSite uses a dual-channel architecture to process different feature types separately. The first channel converts the RNA's tertiary structure into a 3D voxel grid, performs probe scanning to identify potential binding regions, and calculates interaction energies between RNA and 18 types of atomic probes using the ITScore-NL pretrained statistical potential [52]. For each probe type, the top 10 grid points with the most favorable binding energies are selected. The distances from each nucleotide's six-membered ring center to these grid points are computed, and the energy values are mapped to nucleotides via inverse-distance weighting, forming a 10×18 energy feature map. A spatial sliding window is applied to stack the target nucleotide and its neighboring nucleotides, producing a $(2w+1) \times 10 \times 18$ (with $w = 5$) energy tensor. This tensor is processed by a 3D-CNN and a self-attention mechanism along the probe dimension to produce weighted spatial energy features [53].

The second channel integrates structural and sequence descriptors, including network properties (degree and closeness), Laplacian Norm, solvent-accessible surface area, and evolutionary conservation score, which is computed by plmDCA. These features are similarly aggregated with a sliding window of $w = 5$ and fed into a BiLSTM to capture bidirectional contextual dependencies among nucleotides. Finally, the outputs from both channels are concatenated and passed through a fully connected layer to generate the prediction result of RLSite.

3. Feature Construction

RNA information can be classified into three categories: sequence, secondary structure, and tertiary structure, as shown in Figure 1B. Various features can be derived from these three types of information. The features extracted from these categories for predicting RNA–small molecule binding sites are summarized in Table 2, including the RNA information type, brief principle, available websites, and related methods. The fundamental principles of each feature are explained below.

Table 2. Features utilized in methods for predicting RNA–small molecule binding sites. Name refers to the feature names, where MSA represents multiple sequence alignment, plmDCA represents pseudolikelihood maximization direct coupling analysis, LLM embedding represents large language model embedding, SASA represents solvent accessible surface area. Type refers to the categories of RNA information required for feature computation. Brief principle refers to concise explanations of the feature characteristics. Available refers to the available code or website in which the feature is obtained in the related methods. Related method refers to the method that utilizes the feature.

Name	Type	Brief Principle	Available	Related Method
MSA	seq	MSA reveals sequence similarity and evolutionary relationship	BLASTN: https://blast.ncbi.nlm.nih.gov/ (accessed on 20 August 2025)	RNAsite, RLBind, ZeSTa, RLsite
plmDCA	seq	plmDCA focuses on direct interactions within nucleotide mutual information	ZeSTa code: http://zhaoserver.com.cn/ZHmol/ZHmolReSTasite/ZHmolReSTasite.html (accessed on 20 August 2025) RLsite code: https://github.com/fine1231/RLsite (accessed on 20 August 2025) Consurf-DB webserver: http://consurf.tau.ac.il (accessed on 20 August 2025)	RLBind, ZeSTa, RLsite

Table 2. *Cont.*

Name	Type	Brief Principle	Available	Related Method
LLM embedding	seq	LLM embeddings capture the distribution patterns of RNA sequences and their contextual semantic information.	LucaOne: https://github.com/LucaOne/LucaOneApp (accessed on 20 August 2025)	MultiModRLBP, RNABind
			RNA-FM: https://github.com/ml4bio/RNA-FM (accessed on 20 August 2025)	
			ProtRNA: https://github.com/roxie-zhang/ProtRNA (accessed on 20 August 2025)	
			RiNALMo: https://github.com/lbcb-sci/RiNALMo (accessed on 20 August 2025)	
			ERNIE-RNA: https://github.com/Bruce-ywj/ERNIE-RNA (accessed on 20 August 2025)	
			RNAErnie: https://github.com/CatIIIIIIII/RNAErnie (accessed on 20 August 2025)	
			RNA-MSM: https://github.com/yikunpku/RNA-MSM (accessed on 20 August 2025)	
LLM embedding	seq	LLM embeddings capture the distribution patterns of RNA sequences and their contextual semantic information.	RNABERT: https://github.com/mana438/RNABERT (accessed on 20 August 2025)	MultiModRLBP, RNABind
Secondary Structure Region	secondary structure	Regional classification can distinguish the loop types of the secondary structure and the stem regions where nucleotides are located	RNAstat: https://github.com/RNA-folding-lab/RNAStat (accessed on 20 August 2025)	ZeSTa
SASA	3D structure	SASA describes the measure of surface exposure of nucleotides.	RNAsol: https://yanglab.qd.sdu.edu.cn/RNAsol/ (accessed on 20 August 2025)	RNAsite, RLBind, ZeSTa, MultiModRLBP, RLsite
			Freesasa, POPS	
Degree	3D structure	Degree reflects the direct interactions between a nucleotide node and its neighbors.	RLBind code: https://github.com/KailiWang1/ (accessed on 20 August 2025)	RBind, RNAsite, RNetsite, RLBind, ZeSTa, MultiModRLBP,RLsite
			RNetsite code: http://zhaoserver.com.cn/RNet/RNet.html (accessed on 20 August 2025)	
			ZeSTa code: http://zhaoserver.com.cn/ZHmol/ZHmolReSTasite/ZHmolReSTasite.html (accessed on 20 August 2025)	
			RLsite code: https://github.com/fine1231/RLsite (accessed on 20 August 2025)	

Table 2. *Cont.*

Name	Type	Brief Principle	Available	Related Method
Neighborhood Connectivity	3D structure	Neighborhood connec-tivity characterizes the average number of connections of the neighboring nodes of a nucleotide node.	RNetsite code: http://zhaoserver.com.cn/RNet/RNet.html (accessed on 20 August 2025)	RNetsite
Closeness	3D structure	Closeness measures the average distance from a nucleotide node to all other nodes.	RLBind code: https://github.com/KailiWang1/ (accessed on 20 August 2025) RNetsite code: http://zhaoserver.com.cn/RNet/RNet.html (accessed on 20 August 2025) ZeSTa code: http://zhaoserver.com.cn/ZHmol/ZHmolReSTasite/ZHmolReSTasite.html (accessed on 20 August 2025)	RBind, RNAsite, RNetsite, RLBind, ZeSTa, MultiModRLBP,RLsite
Closeness	3D structure	Closeness measures the average distance from a nucleotide node to all other nodes.	RLsite code: https://github.com/fine1231/RLsite (accessed on 20 August 2025)	RBind, RNAsite, RNetsite, RLBind, ZeSTa, MultiModRLBP,RLsite
Eccentricity	3D structure	Eccentricity measures the maximum distance from a nucleotide node to all other nodes in the network.	RNetsite code: http://zhaoserver.com.cn/RNet/RNet.html (accessed on 20 August 2025)	RNetsite
Betweenness	3D structure	Betweenness centrality measures how frequently a node appears on all shortest paths in the network.	RNetsite code: http://zhaoserver.com.cn/RNet/RNet.html (accessed on 20 August 2025)	RNetsite
Laplacian Norm	3D structure	The Laplacian Norm describes the concavity and convexity of nucleotides.	ZeSTa code: http://zhaoserver.com.cn/ZHmol/ZHmolReSTasite/ZHmolReSTasite.html (accessed on 20 August 2025) RLsite code: https://github.com/fine1231/RLsite (accessed on 20 August 2025)	RNAsite, ZeSTa RLsite
Pocket	3D structure	A pocket is an internal cavity within the tertiary structure of RNA.	Ghecom: https://pdbj.org/ghecom/README_ghecom.html (accessed on 20 August 2025)	ZeSTa
Interaction Energy	3D structure	The interaction between RNA and small molecules maintains the stability of the complex structure.	RLsite code: https://github.com/fine1231/RLsite (accessed on 20 August 2025)	RLsite

3.1. Sequence-Based Features

Multiple Sequence Alignment. Multiple sequence alignment (MSA) is a technique that searches for homologous sequences of a target sequence in a database and then aligns the homologous or functionally related segments [25]. The MSA of RNA sequences reveals simi-

larity and evolutionary relationships between sequences and has been successfully applied to RNA-related tasks. Evolutionary information extracted from MSA has been shown to improve the performance of RNA structure prediction significantly. Compared to relying solely on a single sequence, the incorporation of MSA can provide evolutionary information of nucleotides, such as evolutionary conservation. In predicting RNA–small molecule binding sites, the evolutionary information of nucleotides introduced by MSA can provide valuable additional information for characterizing small molecule binding sites. The MSA of RNA can be acquired from the BLASTN Field, which searches against the nucleic acid FASTA database provided by BLAST, and outputs the target RNA sequence along with its homologous sequences [32]. RNAsite, RLBind, ZeSTa, and RLsite all utilize MSA to derive evolutionary insights from RNA sequences [22,23,25,28]. RNAsite directly employs the MSA results to compute conservation scores, while the other three methods perform additional processing with direct coupling analysis.

Direct Coupling Analysis. The homologous sequences derived from the MSA contain information about RNA sequence conservation. Performing direct coupling analysis (DCA) on these sequences can uncover co-evolutionary relationships between nucleotides. Such relationships refer to interdependent changes happening at two nucleotides over evolutionary time, where a mutation in one often prompts a change in the other to maintain structural or functional stability. When analyzing these relationships in MSA, mutual information can be split into direct and indirect interactions. Direct interactions reflect actual correlations, while indirect interactions are mediated through other nucleotides. Relying only on mutual information makes it hard to distinguish between these types. DCA uses a statistical model to effectively differentiate direct from indirect interactions, emphasizing the direct nucleotide interactions usually linked to structural and functional roles. RLBind, ZeSTa, and RLsite use the plmDCA algorithm to derive evolutionary constraint information between nucleotides. Specifically, RLBind obtains plmDCA results via the online software ConSurf-DB, while ZeSTa and RLsite generate plmDCA outputs through local programs.

Large Language Model. Large language models (LLMs) were originally developed in the field of natural language processing and were later applied to the study of biomolecules. Since ESMfold achieved success in protein structure prediction with biological LLMs, RNA LLMs have also been rapidly developed and applied to many tasks, such as RNA structure prediction and small molecule binding site prediction [54]. RNA LLMs typically consists of two phases: pre-training and fine-tuning. During the pre-training phase, models are trained using RNA sequences from RNA sequence databases. The amount of RNA sequence data used for pre-training is usually enormous; for example, RNA-FM was pretrained on 23 million ncRNA sequences from the RNAcentral database [51,55]. Pre-training on extensive RNA sequence data allows the model to learn about sequence patterns and contextual meanings, helping it to grasp the fundamental functional and structural details. Afterward, for a specific downstream task, the model can be fine-tuned with task-specific datasets or used to produce a fixed-dimensional embedding for each nucleotide in an RNA sequence. These embeddings act as contextual-semantic features that highlight the sequential properties of nucleotides. MultiModRLBP and RNABind leverage RNA LLMs to encode nucleotides from RNA sequences. Specifically, MultiModRLBP uses RNABERT, while RNABind assesses and compares eight RNA LLMs, including RNABERT and RNA-FM, to determine their effectiveness in predicting RNA–small molecule binding sites.

3.2. Secondary Structure-Based Features

RNA folding largely depends on its base-paired structure, known as the RNA secondary structure. This secondary structure is essential for the proper function of many

RNA molecules. In interactions between RNA and small molecules, binding sites are frequently found near the loop regions of the secondary structure. This pattern indicates that using secondary structure information to analyze nucleotides could improve predictions of RNA–small molecule binding sites.

There are two methods for representing RNA secondary structure: dot-bracket notation and region notation. Dot-bracket notation is a concise way to depict RNA secondary structure, showing how nucleotides are paired within an RNA molecule. In this notation, dots and parentheses are used; a dot '.' indicates the nucleotide at that position is unpaired, while left/right parentheses indicate that the nucleotide is paired with a subsequent or preceding nucleotide. This method clearly illustrates which nucleotides are paired and which are not. The ZeSTa method, mentioned above, characterizes nucleotides based on the secondary structure regions they occupy, such as stems, internal loops, bulge loops, hairpin loops, and junction loops. Additionally, ZeSTa compared the performance of dot-bracket notation and region notation for characterizing RNA secondary structure and found that region notation, which offers a more detailed classification of the secondary structure regions of nucleotides, is more effective than simply classifying nucleotides as paired or unpaired using dot-bracket notation.

3.3. Network-Based Features

For an RNA consisting of n nucleotides, its structure can be represented as a nucleotide network, where nodes correspond to individual nucleotides and edges indicate non-covalent interactions between them. Network properties can then be used to characterize each node within this network. These properties, used in predicting RNA–small molecule binding sites, are classified into local and global categories based on their attributes. Local properties include degree and neighborhood connectivity, whereas global properties encompass closeness, betweenness centrality, and eccentricity. The following section details the definitions and equations for these network properties.

Degree. Degree centrality quantifies the number of edges connected to a node x, indicating its local connectivity. In nucleotide networks, degree reflects the direct interactions between a specific nucleotide node and its neighbors. Nodes with higher degree values show greater local connectivity and often act as binding sites for small molecules.

Neighborhood Connectivity. Neighborhood connectivity refers to the average degree of all the neighbors of a node x, characterizing the average number of connections of the node's neighbors. Nodes with high neighborhood connectivity are often located in densely connected core regions of the network, and are more likely to serve as binding sites of small molecules. The neighborhood connectivity (NC) of node x is defined as:

$$NC(x) = \frac{\sum DG(y)}{DG(x)} \tag{2}$$

where DG denotes the degree value and y is the neighbor of x.

Closeness centrality. A node's closeness centrality is the inverse of the sum of its shortest path lengths to all other nodes, measuring the average distance from the node to the others. A higher closeness value indicates that the node is closer to the other nodes and is located nearer to the geometric center of the network, making it more important within the network. In nucleotide networks, binding sites often have higher closeness values. The closeness centrality of a node x can be calculated as follows:

$$C(x) = \frac{n-1}{\sum_{x \neq y} d(x,y)} \tag{3}$$

where $d(x, y)$ is the shortest path distance between node x and any other node y, n is the number of nodes in the network.

Betweenness centrality. Betweenness centrality measures how often a node appears on all shortest paths in a network. Nodes with higher betweenness values lie on more shortest paths and play a more important role in the flow of information within the network. In interactions with small molecules, nucleotide nodes that are more crucial in information flow are more likely to serve as binding sites. Betweenness centrality (BC) of a node x can be calculated as follows:

$$BC(x) = \sum_{m \neq x \neq n} \frac{\sigma_{mn}(x)}{\sigma_{mn}} \tag{4}$$

where σ_{mn} is the number of shortest paths from node m to node n, and $\sigma_{mn}(x)$ is the number of shortest paths from node m to node n that pass through node x.

Eccentricity. Eccentricity is defined as the number of steps required from node x to reach any other node in the network. It measures the maximum distance from node x to all other nodes in the network. Nodes with high eccentricity are far from the most distant nodes and are located near the network's periphery, whereas nodes with low eccentricity are closer to the most distant nodes and are located near the network's center. In nucleotide networks, nucleotides that bind to small molecules often have lower eccentricity. The calculation of eccentricity (EC) for node x is given by:

$$EC(x) = \underset{x \neq y}{max}\, d(x, y) \tag{5}$$

where $d(x, y)$ denotes the distance between node x and y.

3.4. Geometry-Based Features

Solvent Accessible Surface Area. The Solvent Accessible Surface Area (SASA) is defined as the surface area traced by the center of a spherical probe, roughly the size of a water molecule, as it rolls over the RNA surface. It indicates how much of the RNA is accessible to solvent molecules, thus reflecting the exposure level of different RNA regions [56]. A nucleotide with a higher SASA is more exposed, while one with a lower SASA is more buried within the RNA structure. Among the methods listed, RNAsite, RLBind, ZeSTa, MultiModRLBP, and RLsite utilize SASA to describe nucleotide properties. Overall, SASA is a key feature for characterizing the structural properties of nucleotides.

Laplacian Norm. The surface features of biomolecules are vital for understanding their properties and function as a key geometric indicator [31]. Laplacian Norm (LN) quantifies the concavity and convexity of nucleotides. Nucleotides with higher LN values have a more convex surface, while those with lower values are more concave. LN reflects the geometric features of nucleotides and is derived by measuring the distance between a target nucleotide and the weighted centroid of neighboring nucleotides. The process involves two steps: first, applying the Laplacian operator, and second, calculating LN. The Laplacian operator is defined as:

$$\Omega_{ij}(\sigma) = \begin{cases} e^{-\frac{\|p_i - p_j\|^2}{\sigma^2}}, & if\ |i-j| > 1 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where p_i and p_j denote the C3 atom coordinates of nucleotide i and j, respectively. σ is a tunable parameter that acts as a scale factor. A smaller value of σ makes the Laplacian operator more sensitive to adjacent nucleotides, while a higher value of σ incorporates

contributions from more distant nucleotides. Once the Laplacian operator is obtained, LN can be computed as:

$$LN_i(\sigma) = \left\| p_i - \frac{\sum_j^{|i-j|>1} \left[p_i \times \Omega_{ij}(\sigma) \right]}{\sum_j^{|i-j|>1} \Omega_{ij}(\sigma)} \right\| \quad (7)$$

Several approaches, including RNAsite, ZeSTa and RLSite, have employed LN as a feature to characterize nucleotides. Specifically, these approaches adopt multiple σ values (0, 1/4, 1/2, 3/4, 1) to consider the influence of surrounding nucleotides across different scales, thereby characterizing nucleotide concavity and convexity across scales ranging from local to global.

Pocket. The RNA pocket region is the internal cavity formed by nucleotides in RNA tertiary structures. Small molecules tend to bind to pocket regions when interacting with RNA molecules [57–59]. Therefore, compared to non-pocket nucleotides, pocket nucleotides are more likely to be small molecule binding sites. By classifying a nucleotide as a pocket/non-pocket type, properties of the nucleotide can be distinguished from a geometric perspective. Many approaches for detecting RNA pockets have been developed, such as 3V, MSPocket, PocketFinder, CHUNNEL, Ghecom [60–66]. These approaches use different methodologies to detect RNA pockets, and the detected pockets show variations. In addition, different RNA pockets often have varying sizes and depths, and these differences can affect the pocket nucleotides' potential to bind small molecules. Using solely binary labels to mark nucleotides as pocket or non-pocket type, the geometric information of the pockets formed by nucleotides will be lost. The Ghecom method provides a measure, pocketness, to quantify the size and depth of the pocket formed by each nucleotide. This quantification of the geometric characteristics of pockets formed by different nucleotides can further subdivide pocket nucleotides. On this basis, ZeSTa utilizes probes of various sizes with the Ghecom algorithm to determine the pocketness of nucleotides when detecting pockets at different scales, thereby providing a comprehensive description of the geometric properties of nucleotides in forming pockets.

3.5. Energy-Based Features

The characterization of nucleotide energy properties explains the physical forces involved in RNA–small molecule interactions. These forces support the structural stability of RNA-ligand complexes and include long-range interactions, electrostatic forces, and four types of short-range interactions: stacking, hydrogen bonding, van der Waals forces, and hydrophobic interactions. Since RNA carries a strong negative charge, electrostatic interactions are especially crucial for RNA-ligand binding, as they facilitate long-distance recognition between RNAs and ligands [9,67]. In contrast, short-range interactions are associated with the optimization and stabilization of the complexes. For example, as one of the main features in nucleic acid structures, stackings are the non-covalent attractive interactions between aromatic rings and are crucial for stabilizing nucleic acid structures [28]. Hydrogen bonds are interactions that occur between a hydrogen atom and two electronegative atoms, like oxygen and nitrogen. The van der Waals interactions are weaker forces resulting from temporary uneven charge distributions between atoms [68,69]. Hydrophobic interactions are driven by the increase in entropy of surrounding water molecules when hydrophobic groups aggregate [70,71]. To consider these energy-related interactions, RLsite employed the potential function ITScore-NL to include nucleotide interaction energy features. ITScore-NL is a potential-based scoring system with three components: pairwise potential, stacking potential, and electrostatic potential.

4. Datasets for Predicting RNA–Small Molecule Binding Site: Training and Evaluating

Developing AI-based methods requires a training set, while evaluating their performance needs test sets that do not overlap with the training data [18]. Consequently, the process of constructing datasets is vital for the development of methods to predict RNA–small molecule binding sites. Figure 1A shows the three steps involved: structure collection, dataset filtering, and information extraction. Since known binding data for RNA–small molecule complexes is essential, data collection typically begins with gathering resolved structures of RNA–small molecule complexes from tertiary structure databases, such as PDB, HARIBOSS, and RNAglib [17,72,73]. Subsequently, dataset filtering is performed to eliminate redundancy. Initial filtering typically removes structures that do not meet criteria such as excessively long or short RNA sequences or the presence of non-canonical nucleotides. Redundancy is then further reduced based on sequence and structural similarity, aiming to exclude highly similar structures. Finally, nucleotides in RNA that interact with small molecules are designated as binding sites, using a 4 Å cutoff: if any atom of a nucleotide is within 4 Å of any atom of the small molecule, it is considered a binding site. These labeled structures are then used for feature extraction, including sequence, secondary, and tertiary structures, to facilitate training and testing.

4.1. Training Sets

4.1.1. TR60

The TR60 dataset, derived from the RNAsite methodology, has been employed to train models such as RNAsite, RNetsite, RLBind, ZeSTa, RNABind, and RLsite. Su et al. initially collected 712 RNA–small molecule complexes (excluding water molecules), with RNA lengths ranging from 20 to 1500 nucleotides [22]. Structural similarity between RNA structures was assessed using TM-scoreRNA, and a cutoff of 0.3 was applied to eliminate redundant structures: if the TM-scoreRNA between two structures exceeded 0.3, the structure with a higher proportion of binding sites was kept. After removing redundancies, 78 RNA structures remained. These were then clustered into 57 groups at a 30% sequence similarity cutoff. From these, 60 RNAs in 42 groups formed the training set TR60, while 18 RNAs in 15 groups constituted the testing set TE18.

4.1.2. TrainRLBP

Dataset TrainRLBP is the training set for the method MultiModRLBP. Wang et al. extracted 810 RNAs with small molecule binding information (comprising 1248 RNA chains) from datasets provided by the Python package RNAglib [26,73]. Due to the length limitations of the LLM RNABERT used in MultiModRLBP, RNA chains longer than 440 nt were excluded. RNA chains with fewer than 4 binding sites were also excluded, resulting in a final set of 653 RNA chains, which, along with 228 small molecules, formed 1012 RNA–small molecule complexes. Wang et al. combined these 653 RNA chains with TR60, resulting in a total of 666 RNA chains. MMseq2 was used to calculate the sequence similarity among the RNA chains, and the 666 chains were clustered into 180 groups at a similarity cutoff of 30%. After removing 20 groups included in the test sets, the remaining 160 groups contained 561 RNA chains. These chains formed 851 complexes with 196 small molecules and were designated as the TrainRLBP training set for MultiModRLBP.

4.1.3. RNABind Training Set

When compared to other methods, RNABind was trained on the TR60 dataset. However, Zhu et al. noted that TR60 includes only single-chain RNA structures. They aimed to focus on the entire RNA complex structures as a single unit, rather than on individual

chains [27]. To achieve this, they built a new dataset for predicting small molecule binding sites. The process involved first extracting 862 RNA–small molecule complexes from the HARIBOSS dataset (as of 18 September 2024). Zhu et al. then excluded structures with RNA sequences longer than 500 nucleotides, as such structures display a significant imbalance between binding and non-binding sites and are predominantly RNA-protein complexes. Structures containing modified nucleotides were also removed. In the end, 353 RNA structures remained for training and testing the RNABind model.

4.2. Evaluation

4.2.1. Test Sets

According to the ZeSTa method, predicting RNA–small molecule binding sites is categorized into easy and challenging tasks based on the presence of junction loops in the RNA secondary structure, as illustrated in Figure 2A. Easy tasks involve RNAs without junction loops, resulting in relatively simple structures. Conversely, RNAs with junction loops exhibit increased structural complexity, rendering the prediction of binding sites more challenging. Therefore, the evaluation of methods for predicting RNA–small molecule binding sites is divided into assessments of both easy and challenging tasks.

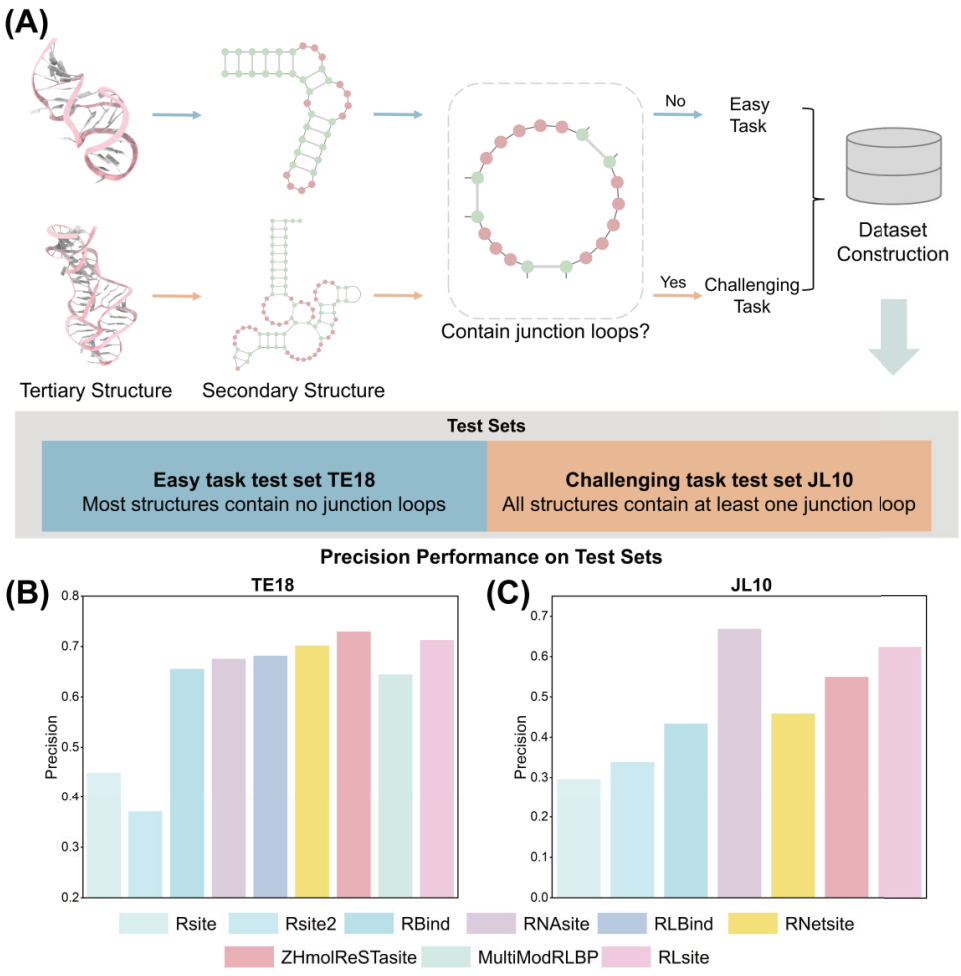


Figure 2. Division of test sets and precision performance of methods. (**A**) RNA structures are divided into easy and challenging prediction tasks based on whether junction loops are in their secondary structures. The easy task test set TE18 and challenging task test set JL10 are constructed under this criterion. (**B**) The precision performance of methods on test set TE18. (**C**) The precision performance of methods on test set JL10.

During the construction of the training set TR60, 60 of the 78 RNAs were selected for training, while the remaining 18 RNAs formed the test set TE18. Most structures in TE18 lack junction loops in their secondary structures, making it a simple test set for evaluating binding site prediction methods. Gao et al. constructed a more challenging, task-specific test set called JL10 [25]. Specifically, all RNA structures released after January 2021 from the PDB were included, provided they do not interact with proteins or DNA and have a resolution better than 4Å, resulting in 263 complexes. RNA chains that did not interact with small molecules and lacked junction loops in their secondary structures were then removed, leaving 76 RNA chains. Further, RNAs with more than two junction loops or binding sites making up less than 10% of the structure were excluded. To prevent redundancy, cd-hit was used for clustering, yielding 10 unique RNA structures [74]. All these RNAs contain at least one junction loop and constitute the challenging task test set JL10. We used JL10 to evaluate the performance of the methods on difficult cases.

4.2.2. Feature Tendency in Easy and Challenging Tasks

To investigate the influence of different features on binding site prediction, we computed nucleotide features in the TE18 and JL10 test sets and analyzed the tendency between binding and non-binding sites. Specifically, we calculated the feature value for each nucleotide. We performed normalization within each RNA chain (for SASA, normalization was conducted within nucleotides of the same type to obtain the relative SASA). Then, the nucleotides in the datasets were ranked according to their normalized feature values in descending order. The ranked nucleotides were divided into 10 groups. We calculated the ratio of binding sites to the total number of nucleotides in the two test sets, which was 0.32 for both TE18 (195/605) and JL10 (265/827). Therefore, for the 10 groups of nucleotides after ranking, if the proportion of binding sites within a group exceed 0.32, we regarded it as a binding-site tendency group and filled it in green; otherwise, it was considered a non-binding-site tendency group (5 nucleotides in TE18 and 7 nucleotides in JL10 were placed in additional groups due to the total number of nucleotides not being divisible by ten). Subsequently, we were able to obtain the feature tendencies of binding sites from the figures.

First, we calculated the sum of Euclidean distances of the tertiary structure from the C1′ atom of each nucleotide to all other nucleotides' C1′ atoms, similar to Rsite, with the resulting trends shown in Figure 3A,D. In TE18, four of the ten groups were identified as binding-site groups, ranking first to fourth in ascending order of these sums, indicating that binding sites tend to have smaller tertiary-structure distances. In JL10, four groups also appeared as binding-site groups, ranking second, fifth, sixth, and seventh, indicating a modest tendency toward shorter distances. For secondary structure distances, we extracted 2D coordinates from the RNAfold output files, as Rsite2 did. Figure 3B,E show that in TE18, five groups were recognized as binding sites, ranking first to fifth, while in JL10, four groups ranked first to fourth, indicating a tendency for binding sites in both to have smaller secondary-structure distance sums. Relative SASA was calculated using freesasa [75], with the tendencies shown in Figure 3C,F. In JL10, three binding-site groups ranked first to third, showing a clear tendency toward lower values, whereas no clear trend was observed in TE18. This suggests that relying solely on SASA is insufficient to distinguish binding from non-binding sites; combining multiple features is essential for accurately characterizing nucleotide properties.

As mentioned earlier, LN serves as a feature that describes how concave or convex nucleotides are. An LN value near 1 suggests the nucleotide has a more convex surface, while a value near 0 indicates a more concave surface. We computed LN using five different σ values, similarly to ZeSTa, to account for nucleotides at various distances from the target, offering a thorough overview of surface concavity and convexity. Figure 4A,F display the

LN1 results computed with σ = 0, which mainly reflect the influence of spatially neighboring nucleotides. In TE18, the four binding-site groups are ranked in increasing order of their values, from first to fourth, while in JL10, the five binding-site groups are ranked first, second, third, fifth, and ninth. The remaining panels of Figure 4 show the results of LN2, LN3, LN4, and LN5 as σ was gradually increased, thereby giving more weight to distant nucleotides. Binding-site groups tend to cluster within the first five in ascending order, indicating that whether the computation emphasizes nearby or distant nucleotides, binding sites generally have smaller LN values. Additionally, in both TE18 and JL10, binding sites often display lower LN values, suggesting they tend to have more concave surfaces. Both LN and SASA are features used to describe the surface properties of nucleotides. In JL10, binding sites tend toward smaller SASA values, and in both JL10 and TE18, they consistently have lower LN values, reinforcing the tendency toward concave surfaces. This is because nucleotides with convex surfaces are more exposed to solvent, resulting in larger SASA, while those with concave surfaces are less accessible to solvent probes, leading to lower SASA values.

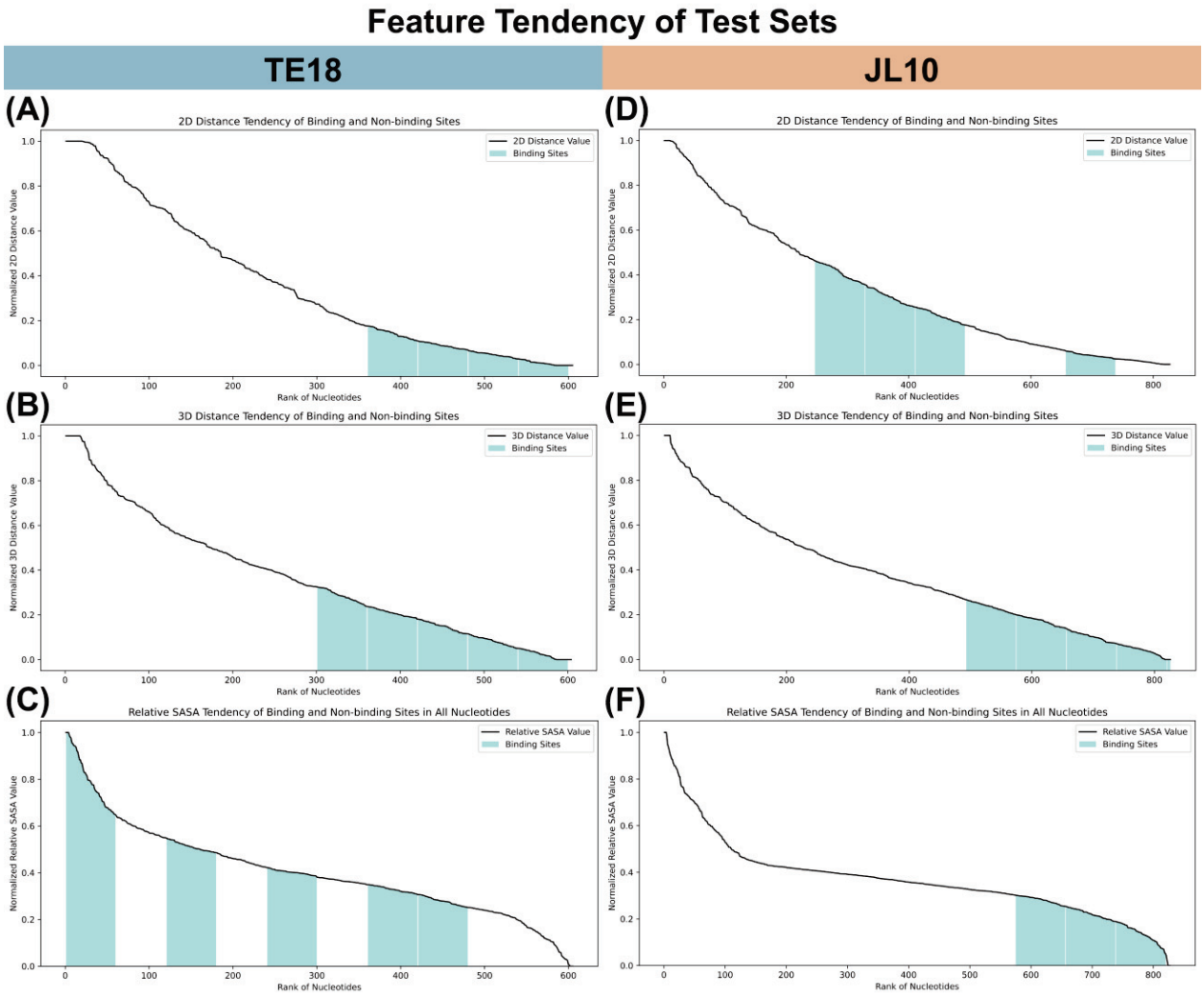


Figure 3. Feature tendency of binding and non-binding sites in TE18 and JL10 test sets. Binding site groups are labeled in green. (**A**) 2D distance tendency of nucleotides in TE18. (**B**) 3D distance tendency of nucleotides in TE18. (**C**) Relative SASA tendency of nucleotides in TE18. (**D**) 2D distance tendency of nucleotides in JL10. (**E**) 3D distance tendency of nucleotides in JL10. (**F**) Relative SASA tendency of nucleotides in JL10.

We also analyzed the tendencies of network and pocket features. Figures S1–S5 illustrates the five network features of nucleotides in TE18 and JL10. It shows that binding sites in both samples display similar tendencies. Figure S1A,B present the local feature degree, where the three binding-site groups in TE18 ranked first, second, and fifth, while the four groups in JL10 ranked within the top four in descending order. Results for neighborhood connectivity, closeness, and betweenness are shown in Figures S2–S4, indicating that binding sites tend to have higher values for these three features. Conversely, Figure S5 reveals that the five binding-site groups in TE18 ranked first, second, third, fifth, and seventh in ascending order, and those in JL10 ranked within the top five. Binding sites generally show higher degree and neighborhood connectivity, suggesting they are often situated in the core topological regions of nucleotide networks with strong local connectivity and dense links. For the global network features, binding sites tend to have higher closeness and betweenness scores and lower eccentricity, implying they are closer to the network's geometric center and participate more in shortest paths, thus playing a vital role in information flow. Additionally, their distances to the most distant nodes are shorter, indicating they are positioned nearer the network center rather than at the periphery.

4.2.3. Performance Analysis

Four metrics are used to evaluate the performances of methods, namely, precision, recall, Matthews correlation coefficient (MCC) and the area under the receiver operating characteristic curve (AUC). Precision, recall, MCC can be calculated as follows:

$$Precision = \frac{|TP|}{(|TP| + |FP|)} \tag{8}$$

$$Recall = \frac{|TP|}{(|TP| + |FN|)} \tag{9}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{10}$$

where TP is true positive and represents the number of true binding nucleotides within the predicted binding nucleotides. FP is false positive and represents the number of false binding nucleotides within the predicted binding nucleotides. TN is the true negative and represents the number of real non-binding nucleotides with the predicted non-binding nucleotides. FN is a false negative representing the number of binding nucleotides with the predicted non-binding nucleotides.

Figures S6–S10 depicts the trends in pocketness, a measure of the size and depth of nucleotide-formed pockets. We employed the same approach as ZeSTa for calculating pocketness, applying the Ghecom dual-probe method with a large default probe radius. We then measured the pocketness of each nucleotide at smaller radii of 1.5, 2.0, 2.5, 3.0, and 3.5 Å, labeled as PKT1 through PKT5. This multi-radius method captures pocket formation at different scales. Figures S6A–S10A show results for TE18, while Figures S6B–S10B display those for JL10. A pocketness close to 1 indicates larger, deeper pockets. In both test sets, binding-site nucleotides tend to cluster within the top five values, especially in PKT1–3, suggesting they form larger, deeper pockets suitable for small molecule binding. At these scales, binding sites are more likely to form sizable pockets compared to non-binding regions. Conversely, at PKT4 and PKT5, many nucleotides have pocketness values of 0, meaning the probes were too large for pocket detection at these scales. In TE18, this is evident from Figures S9A and S10A, and in JL10 from Figures S9B and S10B. The binding site tendency is less clear in PKT4 and PKT5, so focusing on PKT1-3 may be more effective.

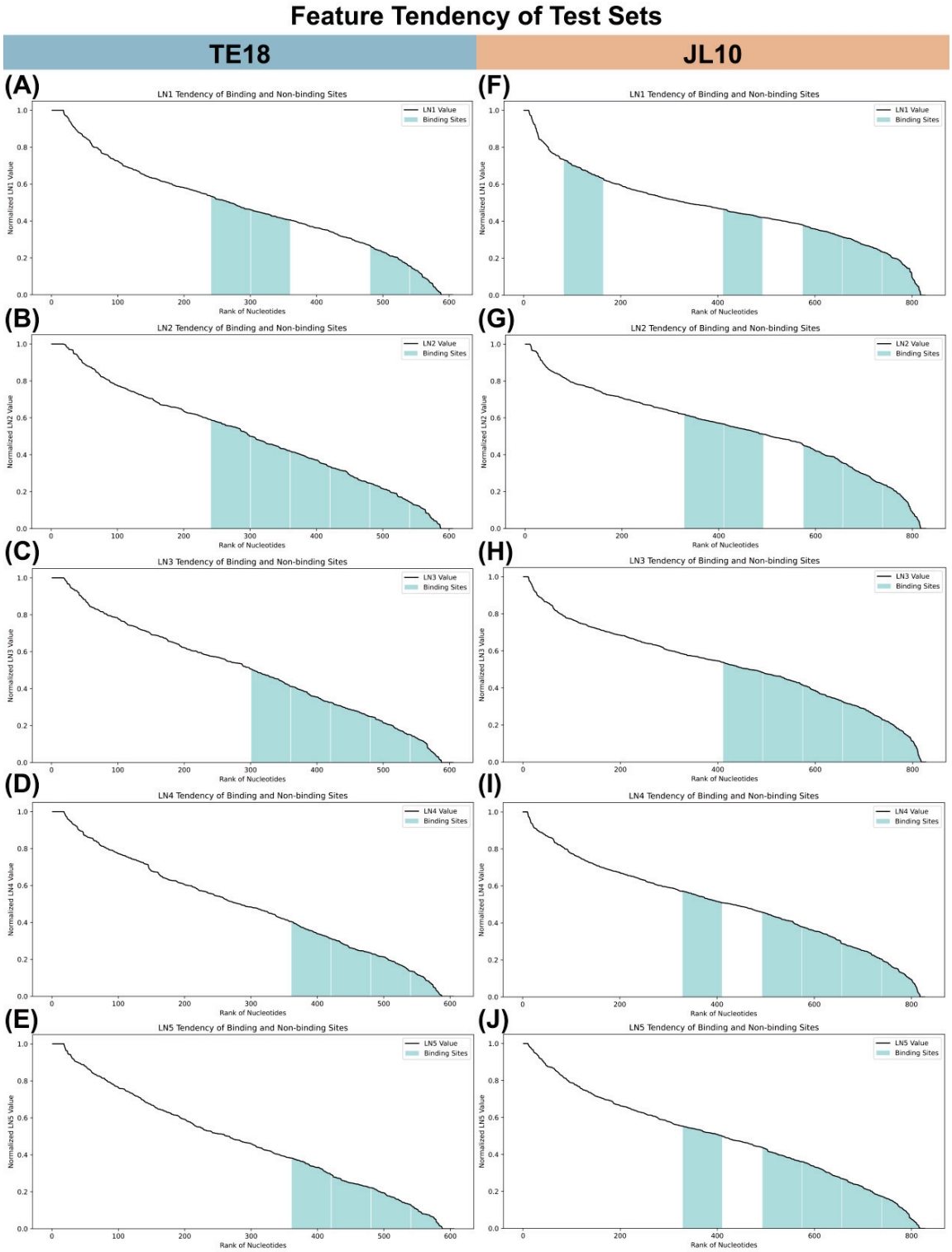


Figure 4. Feature tendency of binding and non-binding sites in TE18 and JL10 test sets. Binding site groups are labeled in green. Tendencies of nucleotides in TE18 across (**A**) LN1, (**B**) LN2, (**C**) LN3, (**D**) LN4, (**E**) LN5. Tendencies of nucleotides in JL10 across (**F**) LN1, (**G**) LN2, (**H**) LN3, (**I**) LN4, (**J**) LN5.

The precision performance of methods on TE18 is shown in Figure 2B. Results for Rsite, Rsite2, RBind, and RNAsite were obtained from the RNAsite publication, whereas results for the other methods were derived from their respective original papers. Among physics-

based methods, RBind (0.655), which incorporates two network features, outperforms the distance-based Rsite (0.449) and Rsite2 (0.370). As for the AI-based methods, the precision of MultiModRLBP (0.644) is approximately 1.7% lower than that of RBind (0.655), while RNAsite (0.675), RLBind (0.681), RNetsite (0.701), ZeSTa (0.729), and RLsite (0.712) all outperform the physics-based approaches. Figure 2C presents the precision performance of methods on JL10. The results for Rsite, Rsite2, RBind were from the ZeSTa publication, the RNAsite results were obtained from its web server in August 2025, and the RNetsite results were generated in August 2025 using the V2 package provided on its website. The remaining results were from their respective publications. The precision of all methods decreased in RNAs with more complex structures, but the overall trend remains consistent. RBind (0.433) still outperforms Rsite (0.295) and Rsite2 (0.338). Meanwhile, the AI-based methods RNAsite (0.668), RNetsite (0.458), ZeSTa (0.549), and RLsite (0.622) show precision improvements over RBind of 54.3%, 5.8%, 26.8%, and 43.6%, respectively. The other metrics for the various methods on the two test sets are presented in Tables 3 and 4.

Table 3. Model performance on easy-task test set TE18. The evaluation metrics include precision, recall, MCC and AUC, where MCC refers to Matthews Correlation Coefficient and AUC refers to Area Under the Receiver Operating Characteristic Curve. The results of Rsite, Rsite2, RBind, and RNAsite are derived from the report of RNAsite paper, while the results of the other methods are derived from their respective papers. The best performance is bolded, and the second-best is underlined.

Name	Precision	Recall	MCC	AUC
Rsite	0.449	0.288	0.071	0.509
Rsite2	0.370	0.214	0.010	0.474
RBind	0.655	0.173	0.187	0.559
RNAsite	0.675	0.263	0.253	<u>0.776</u>
RLBind	0.681	0.345	0.324	0.720
RNetsite	0.701	0.357	0.307	-
ZHmolReSTasite	**0.729**	0.379	0.327	0.709
MultiModRLBP	0.644	**0.523**	**0.378**	**0.780**
RNABind	-	-	-	0.737
RLsite	<u>0.712</u>	<u>0.392</u>	<u>0.335</u>	0.740

Table 4. Model performance on challenging-task test set JL10. The evaluation metrics include precision, recall, MCC and AUC, where MCC refers to Matthews Correlation Coefficient and AUC refers to Area Under the Receiver Operating Characteristic Curve. The results of Rsite, Rsite2, RBind, and ZHmolReSTasite are reported in the ZHmolReSTasite paper, the results of RNAsite are obtained from the RNAsite web server (August 2025), the results of RNetsite are obtained from the RNetsite V2 package on its website (August 2025), and the results of the other methods are derived from their respective papers. The best performance is bolded, and the second-best is underlined.

Name	Precision	Recall	MCC	AUC
Rsite	0.295	0.194	−0.046	0.477
Rsite2	0.338	0.131	0.007	0.504
RBind	0.433	0.142	0.083	0.532
RNAsite	**0.668**	**0.327**	**0.323**	0.637
RNetsite	0.458	0.136	0.115	0.540
ZHmolReSTasite	0.549	0.296	0.211	0.592
RNABind	-	-	-	<u>0.667</u>
RLsite	<u>0.622</u>	<u>0.320</u>	<u>0.286</u>	**0.700**

In the tendency figures shown in Figure 3A,B,D,E, binding sites consistently show smaller values in both tertiary-structure-based and secondary-structure-based distance

sums. This highlights the potential of 2D/3D distance sums to characterize binding site properties. Consequently, distance-sum-based methods, such as Rsite and Rsite2, effectively predict small molecule binding sites in RNA. However, calculating distance sums by giving equal weight to all nucleotides can introduce significant noise from irrelevant interactions. Additionally, relying only on distance sums cannot capture multi-body interactions between nucleotides or distinguish between long- and short-range interactions. RBind utilizes network modeling to integrate local and global features, enabling it to represent multi-body interactions and differentiate between different interaction types. Network topological features help identify key nodes and reduce noise in the interaction network. As a result, RBind outperforms Rsite and Rsite2. Still, these three physics-based methods are limited to a few features and cannot incorporate more diverse and complementary information, which restricts their predictive capabilities.

Machine learning and deep learning algorithms can effectively integrate multiple features. On both test sets, AI-based methods outperform physics-based methods by leveraging various features. Among methods relying solely on network features, RNetsite uses machine learning algorithms to incorporate more complementary network features. On the TE18/JL10 test sets, RNetsite (0.675/0.458) achieved higher precision than RBind (0.655/0.433). Although RNetsite considers complementary network features, its performance is lower than methods with a more diverse feature set, such as ZeSTa (0.729/0.549), due to its lack of other feature perspectives like sequence information. Especially on JL10, which has RNA with more complex structures, RNetsite's precision is only 5.8% higher than RBind's, compared to ZeSTa's 26.8% improvement over RBind. ZeSTa achieved the highest precision on TE18 but underperformed on JL10 compared to RNAsite (0.675/0.668) and RLsite (0.712/0.622), likely due to the high structural sensitivity of its geometric features. Like ZeSTa, RLsite combines sequence-based, geometric-based, and network-based features, and additionally introduces energy features to analyze nucleotides through interaction energy, which ZeSTa does not include. These findings suggest that effective RNA–small molecule binding site prediction requires not only complementary features but also a diverse combination of features.

Since features derived from RNA sequences are independent of RNA structures, incorporating sequence-based features offers a more comprehensive characterization of nucleotides. Therefore, most AI methods rely on sequence-based features. RNA sequence information can be extracted via two approaches: using MSA-DCA (as seen in RNAsite, RLBind, ZeSTa, and RLsite) and utilizing RNA LLMs (as in MultiModRLBP and RNABind). MSA-DCA examines relationships between nucleotides from an evolutionary standpoint. However, a limitation of MSA-DCA is that many RNA sequences lack sufficient homologous sequences, which restricts the information it can provide. Conversely, RNA LLMs are trained on large-scale sequence datasets and do not require searching for homologs during application, making them more efficient. Notably, RNABind trained separate models using embeddings from eight different LLMs and observed that their predictive performances varied. Therefore, choosing the right RNA LLM to represent sequence information is crucial for accurate predictions.

Besides sequence-based features, methods for predicting RNA–small molecule binding sites utilize various structure-based features, which are classified into four groups: network, geometry, energy, and secondary structure. Combining features from the same group can enhance the description of nucleotide properties, while integrating features from different groups provides a more complete characterization. For instance, network features derived from tertiary structure models describe binding sites via network topology. These sites are often located in the dense core of the local nucleotide network and are closer to the network

center, thereby facilitating their role in communication. While network features alone offer valuable insights, they have limitations when it comes to describing the intrinsic properties of individual nucleotides. To address this, many methods incorporate geometric features, such as SASA, which indicates solvent exposure, showing a lower tendency, mainly on JL10, with no clear pattern on TE18. Methods such as RNAsite, ZeSTa, and RLsite combine SASA with LN or pocketness to better describe the geometric traits of nucleotides. This combined approach helps identify binding sites in pocket regions with more concave surfaces, leading to lower solvent exposure. Some methods further improve prediction accuracy by adding energy features (RLsite) and secondary structure features (ZeSTa), demonstrating the benefit of combining features across different categories. Ultimately, effective nucleotide structure characterization depends on integrating diverse features within and across categories for a thorough understanding.

5. Future Directions

Existing approaches for predicting RNA–small molecule binding sites employ various features and models. Overall, AI-based methods, which integrate diverse features to characterize nucleotides, tend to outperform physics-based techniques that rely on isolated strategies. The effectiveness of AI models largely depends on the quality and quantity of the training data used. Therefore, increasing the amount of RNA data that does not overlap with test sets for training can enhance predictive accuracy. By analyzing the structural and sequence similarities within the current training datasets, we observed many RNAs that are structurally alike, and a few RNA chains that share numerous homologous sequences across different datasets [76,77]. To expand the amount of structural data available for training AI models with improved accuracy, tools like AlphaFold can be utilized to perform enhanced sampling, such as introducing mutations and creating conformational ensembles [78–80].

Additionally, selecting the right combination of features is essential for effectively characterizing nucleotides. However, adding more feature types increases computational costs. It is essential to identify a feature set that is both diverse and complementary, yet computationally manageable. Previous studies highlight that including geometric features like LN, pocket, and interaction-based energy features is essential for improving the model's predictive accuracy, as these features display clear distribution differences between binding and non-binding sites [22,25,28]. Nonetheless, while structure-based refined features effectively capture RNA's structural properties, their high sensitivity to structural variations often causes a notable drop in predictive performance when input structures deviate from the native conformation. Conversely, sequence-based features, which do not depend on detailed structural information, or coarse-grained network features less sensitive to structural changes respond more consistently to structural perturbations and tend to be more robust under variations [24].

Furthermore, choosing an appropriate AI algorithm can maximize the use of limited data to identify underlying patterns. For simple or short structures, machine learning models with fewer parameters or lightweight deep learning models, combined with a single feature type, can deliver satisfactory predictive results. However, for more complex structures, advanced model architectures are necessary, involving multiple modules and multimodal features to fully represent RNA from sequence to structure, thus improving prediction accuracy. Addressing these factors can lead to more accurate predictions of RNA–small molecule binding sites.

The flexibility of RNA structure is a critical aspect in the development of RNA–small molecule drugs [69]. This flexibility manifests not only in the molecule's dynamic conformational changes but also in the structural shifts that occur before and after ligand

binding, reflecting differences between the unbound and bound states. The secondary and tertiary structures can vary significantly between these states, influencing small-molecule binding. Therefore, considering RNA's structural flexibility is essential in practical drug development. However, most current binding-site prediction methods mainly depend on RNA conformations derived from known RNA–small molecule complexes (the bound state), often overlooking the unbound state. This restriction makes applying predicted sites to guide drug development more challenging. Future research should emphasize RNA's structural flexibility to better mirror real-world scenarios. For example, by incorporating conformational ensembles of RNA structures obtained from molecular dynamics simulations or NMR data, it is possible to include information on the dynamic structural changes of RNA and account for its structural flexibility [81,82]. As the accuracy of RNA–small molecule binding site prediction improves, it will lay a robust foundation and create new opportunities for designing drugs that target RNA.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ph18101593/s1, Figure S1: Degree tendencies of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding site groups are labeled in green; Figure S2: Neighborhood connectivity tendencies of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding site groups are labeled in green; Figure S3: Closeness tendencies of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding site groups are labeled in green; Figure S4: Betweenness tendencies of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding site groups are labeled in green; Figure S5: Eccentricity tendencies of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding site groups are labeled in green; Figure S6: PKT1 tendencies of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding site groups are labeled in green; Figure S7: PKT2 tendencies of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding site groups are labeled in green; Figure S8: PKT3 tendencies of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding site groups are labeled in green; Figure S9: PKT4 tendencies of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding site groups are labeled in green; Figure S10: PKT5 tendencies of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding site groups are labeled in green; Figure S11: Feature distribution in 2D distance, 3D distance, and RSASA of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding sites are labeled in purple, and non-binding sites are labeled in green; Figure S12: Feature distribution in Laplacian Norm (LN) of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding sites are labeled in purple, and non-binding sites are labeled in green; Figure S13: Feature distribution in network features (degree, neighborhood connectivity, closeness, betweenness, and eccentricity) of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding sites are labeled in purple, and non-binding sites are labeled in green; Figure S14: Feature distribution in pocketness of binding and non-binding sites in (A) TE18 and (B) JL10 test sets. Binding sites are labeled in purple, and non-binding sites are labeled in green; Figure S15: Sequence and structure similarity of three training sets (TR60, trainRLBP, and training set of RNABind). RLBP represents trainRLBP, and RNAB stands for the training set of RNABind. Each dot depicts an RNA in the dataset; the x-coordinate shows the maximum TM-score between the query RNA and any structure in the other dataset, while the y-coordinate indicates the number of RNA chains in the other dataset that fall into the same cluster as the query RNA.

Author Contributions: J.G. collected the information on RNA–small molecule binding sites and wrote the manuscript; C.Z. (Chen Zhuo) and C.Z. (Chengwei Zeng) assisted with computational model information collection; H.L. assisted with feature computation; Y.Z. designed the project and supervised the overall study. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (grant no. 12175081); Hubei Science Fund for Distinguished Young Scholars (grant no. 2024AFA077); Fundamental Research Funds for the Central Universities (grant no. CCNU25ai010).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable.

Conflicts of Interest: The authors declare no competing financial interest.

Abbreviations

LLM: Large language model; MSA, Multiple sequence alignment; DCA, Direct coupling analysis; ResNet, Residual network; PKT, Pocket; SASA, Solvent accessible surface area; LN, Laplacian norm; AI, Artificial intelligence; EGNN, E(3)-equivariant graph neural network.

References

1. Yang, C.; Chen, E.A.; Zhang, Y. Protein–Ligand Docking in the Machine-Learning Era. *Molecules* **2022**, *27*, 4568. [CrossRef]
2. Kognole, A.A.; Hazel, A.; MacKerell, A.D. SILCS-RNA: Toward a Structure-Based Drug Design Approach for Targeting RNAs with Small Molecules. *J. Chem. Theory Comput.* **2022**, *18*, 5672–5691. [CrossRef]
3. Liu, H.; Jian, Y.; Zeng, C.; Zhao, Y. RNA-protein interaction prediction using network-guided deep learning. *Commun. Biol.* **2025**, *8*, 247. [CrossRef] [PubMed]
4. Cheng, A.C.; Coleman, R.G.; Smyth, K.T.; Cao, Q.; Soulard, P.; Caffrey, D.R.; Salzberg, A.C.; Huang, E.S. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **2007**, *25*, 71–75. [CrossRef]
5. Fan, C.; Wang, X.; Ling, T.Z.; Yang, Y.D.; Zhao, H.Y. Characterizing RNA-binding ligands on structures, chemical information, binding affinity and drug-likeness. *RNA Biol.* **2023**, *20*, 431–443. [CrossRef]
6. Esteller, M. Non-coding RNAs in human disease. *Nat. Rev. Genet.* **2011**, *12*, 861–874. [CrossRef] [PubMed]
7. Meyer, C.; Garzia, A.; Miller, M.W.; Huggins, D.J.; Myers, R.W.; Hoffmann, H.-H.; Ashbrook, A.W.; Jannath, S.Y.; Liverton, N.; Kargman, S.; et al. Small-molecule inhibition of SARS-CoV-2 NSP14 RNA cap methyltransferase. *Nature* **2024**, *637*, 1178–1185. [CrossRef] [PubMed]
8. Childs-Disney, J.L.; Yang, X.Y.; Gibaut, Q.M.R.; Tong, Y.Q.; Batey, R.T.; Disney, M.D. Targeting RNA structures with small molecules. *Nat. Rev. Drug Discov.* **2022**, *21*, 736–762. [CrossRef]
9. Chen, S.Y.; Mao, Q.; Cheng, H.; Tai, W.Y. RNA-Binding Small Molecules in Drug Discovery and Delivery: An Overview from Fundamentals. *J. Med. Chem.* **2024**, *67*, 16002–16017. [CrossRef]
10. Kronenberger, T.; Laufer, S.A.; Pillaiyar, T. COVID-19 therapeutics: Small-molecule drug development targeting SARS-CoV-2 main protease. *Drug Discov. Today* **2023**, *28*, 103579. [CrossRef]
11. Baranello, G.; Darras, B.T.; Day, J.W.; Deconinck, N.; Klein, A.; Masson, R.; Mercuri, E.; Rose, K.; El-Khairi, M.; Gerber, M.; et al. Risdiplam in Type 1 Spinal Muscular Atrophy. *N. Engl. J. Med.* **2021**, *384*, 915–923. [CrossRef] [PubMed]
12. Mercuri, E.; Deconinck, N.; Mazzone, E.S.; Grp, S.S. Safety and efficacy of once-daily risdiplam in type 2 and non-ambulant type 3 spinal muscular atrophy (SUNFISH part 2): A phase 3, double-blind, randomised, placebo-controlled trial. *Lancet Neurol.* **2022**, *21*, E5. [CrossRef] [PubMed]
13. Ratni, H.; Ebeling, M.; Baird, J.; Bendels, S.; Bylund, J.; Chen, K.S.; Denk, N.; Feng, Z.H.; Green, L.; Guerard, M.; et al. Discovery of Risdiplam, a Selective Survival of Motor Neuron-2 (SMN2) Gene Splicing Modifier for the Treatment of Spinal Muscular Atrophy (SMA). *J. Med. Chem.* **2018**, *61*, 6501–6517. [CrossRef]
14. Howe, J.A.; Wang, H.; Fischmann, T.O.; Balibar, C.J.; Xiao, L.; Galgoci, A.M.; Malinverni, J.C.; Mayhood, T.; Villafania, A.; Nahvi, A.; et al. Selective small-molecule inhibition of an RNA structural element. *Nature* **2015**, *526*, 672–677. [CrossRef]
15. Blount, K.F.; Breaker, R.R. Riboswitches as antibacterial drug targets. *Nat. Biotechnol.* **2006**, *24*, 1558–1564. [CrossRef]
16. Disney, M.D. Targeting RNA with Small Molecules to Capture Opportunities at the Intersection of Chemistry, Biology, and Medicine. *J. Am. Chem. Soc.* **2019**, *141*, 6776–6790. [CrossRef] [PubMed]
17. 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. [CrossRef]
18. Alexandrovich, D.; Kagan, S.; Mandel-Gutfreund, Y. BindUP-Alpha: A Webserver for Predicting DNA-and RNA-binding Proteins based on Experimental and Computational Structural Models⋆. *J. Mol. Biol.* **2025**, *437*, 169240. [CrossRef]
19. Zeng, P.; Li, J.; Ma, W.; Cui, Q. Rsite: A computational method to identify the functional sites of noncoding RNAs. *Sci. Rep.* **2015**, *5*, 9179. [CrossRef]

20. Zeng, P.; Cui, Q. Rsite2: An efficient computational method to predict the functional sites of noncoding RNAs. *Sci. Rep.* **2016**, *6*, 19016. [CrossRef]
21. Wang, K.; Jian, Y.; Wang, H.; Zeng, C.; Zhao, Y.; Valencia, A. RBind: Computational network method to predict RNA binding sites. *Bioinformatics* **2018**, *34*, 3131–3136. [CrossRef]
22. Su, H.; Peng, Z.; Yang, J.; Ponty, Y. Recognition of small molecule–RNA binding sites using RNA sequence and structure. *Bioinformatics* **2021**, *37*, 36–42. [CrossRef] [PubMed]
23. Wang, K.; Zhou, R.; Wu, Y.; Li, M. RLBind: A deep learning method to predict RNA–ligand binding sites. *Brief. Bioinform.* **2023**, *24*, bbac486. [CrossRef]
24. Liu, H.; Jian, Y.; Hou, J.; Zeng, C.; Zhao, Y. RNet: A network strategy to predict RNA binding preferences. *Brief. Bioinform.* **2024**, *25*, bbad482. [CrossRef] [PubMed]
25. Gao, J.; Liu, H.; Zhuo, C.; Zeng, C.; Zhao, Y. Predicting Small Molecule Binding Nucleotides in RNA Structures Using RNA Surface Topography. *J. Chem. Inf. Model.* **2024**, *64*, 6979–6992. [CrossRef]
26. Wang, J.; Quan, L.; Jin, Z.; Wu, H.; Ma, X.; Wang, X.; Xie, J.; Pan, D.; Chen, T.; Wu, T.; et al. MultiModRLBP: A Deep Learning Approach for Multi-Modal RNA-Small Molecule Ligand Binding Sites Prediction. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 4995–5006. [CrossRef] [PubMed]
27. Zhu, W.; Ding, X.; Shen, H.-B.; Pan, X. Identifying RNA-small Molecule Binding Sites Using Geometric Deep Learning with Language Models. *J. Mol. Biol.* **2025**, *437*, 169010. [CrossRef]
28. Zou, Y.; Yang, L.; Liu, Y.; Feng, Y. RLsite: Integrating 3D-CNN and BiLSTM for RNA-ligand binding site prediction. *Chin. Phys. B* **2025**, *34*, 088709. [CrossRef]
29. Wang, X.; Yu, S.; Lou, E.; Tan, Y.L.; Tan, Z.J. RNA 3D Structure Prediction: Progress and Perspective. *Molecules* **2023**, *28*, 5532. [CrossRef]
30. Zhao, Y.; Huang, Y.; Gong, Z.; Wang, Y.; Man, J.; Xiao, Y. Automated and fast building of three-dimensional RNA structures. *Sci. Rep.* **2012**, *2*, 734. [CrossRef]
31. Shazman, S.; Elber, G.; Mandel-Gutfreund, Y. From face to interface recognition: A differential geometric approach to distinguish DNA from RNA binding surfaces. *Nucleic Acids Res.* **2011**, *39*, 7390–7399. [CrossRef] [PubMed]
32. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef] [PubMed]
33. Hofacker, I.L.; Fontana, W.; Stadler, P.F.; Bonhoeffer, L.S.; Tacker, M.; Schuster, P. Fast folding and comparison of RNA secondary structures. *Monatshefte Für Chem. Chem. Mon.* **1994**, *125*, 167–188. [CrossRef]
34. Cavallo, L.; Kleinjung, J.; Fraternali, F. POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Res.* **2003**, *31*, 3364–3366. [CrossRef] [PubMed]
35. Liu, H.-F.; Liu, R. Structure-based prediction of post-translational modification cross-talk within proteins using complementary residue- and residue pair-based features. *Brief. Bioinform.* **2019**, *21*, 609–620. [CrossRef]
36. Goldenberg, O.; Erez, E.; Nimrod, G.; Ben-Tal, N. The ConSurf-DB: Pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res.* **2009**, *37*, D323–D327. [CrossRef]
37. Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **2016**, *44*, W344–W350. [CrossRef]
38. Ekeberg, M.; Lövkvist, C.; Lan, Y.; Weigt, M.; Aurell, E. Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. *Phys. Rev. E* **2013**, *87*, 012707. [CrossRef]
39. Jian, Y.; Wang, X.; Qiu, J.; Wang, H.; Liu, Z.; Zhao, Y.; Zeng, C. DIRECT: RNA contact predictions by integrating structural patterns. *BMC Bioinform.* **2019**, *20*, 497. [CrossRef]
40. Guo, Z.-H.; Yuan, L.; Tan, Y.-L.; Zhang, B.-G.; Shi, Y.-Z. RNAStat: An Integrated Tool for Statistical Analysis of RNA 3D Structures. *Front. Bioinform.* **2022**, *1*, 809082. [CrossRef]
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
42. Lu, X.-J.; Bussemaker, H.J.; Olson, W.K. DSSR: An integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.* **2015**, *43*, e142. [CrossRef]
43. Akiyama, M.; Sakakibara, Y. Informative RNA base embedding for RNA structural alignment and clustering by deep representation learning. *NAR Genom. Bioinform.* **2022**, *4*, lqac012. [CrossRef] [PubMed]
44. Garcia Satorras, V.; Hoogeboom, E.; Welling, M. E(n) Equivariant Graph Neural Networks. *arXiv* **2021**, arXiv:2102.09844. [CrossRef]
45. He, Y.; Fang, P.; Shan, Y.; Pan, Y.; Wei, Y.; Chen, Y.; Chen, Y.; Liu, Y.; Zeng, Z.; Zhou, Z.; et al. Generalized biological foundation model with unified nucleic acid and protein language. *Nat. Mach. Intell.* **2025**, *7*, 942–953. [CrossRef]

46. Zhang, R.; Ma, B.; Xu, G.; Ma, J. ProtRNA: A Protein-derived RNA Language Model by Cross-Modality Transfer Learning. *Cell Syst.* **2025**, *16*, 101371. [CrossRef]
47. Penić, R.J.; Vlašić, T.; Huber, R.G.; Wan, Y.; Šikić, M. RiNALMo: General-purpose RNA language models can generalize well on structure prediction tasks. *Nat. Commun.* **2025**, *16*, 5671. [CrossRef] [PubMed]
48. Yin, W.; Zhang, Z.; He, L.; Jiang, R.; Zhang, S.; Liu, G.; Zhang, X.; Qin, T.; Xie, Z. ERNIE-RNA: An RNA Language Model with Structure-enhanced Representations. *bioRxiv* **2024**. bioRxiv:2024.2003.2017.585376. [CrossRef]
49. Wang, N.; Bian, J.; Li, Y.; Li, X.; Mumtaz, S.; Kong, L.; Xiong, H. Multi-purpose RNA language modelling with motif-aware pretraining and type-guided fine-tuning. *Nat. Mach. Intell.* **2024**, *6*, 548–557. [CrossRef]
50. Zhang, Y.; Lang, M.; Jiang, J.; Gao, Z.; Xu, F.; Litfin, T.; Chen, K.; Singh, J.; Huang, X.; Song, G.; et al. Multiple sequence alignment-based RNA language model and its application to structural inference. *Nucleic Acids Res.* **2023**, *52*, e3. [CrossRef]
51. Chen, J.; Hu, Z.; Sun, S.; Tan, Q.; Wang, Y.; Yu, Q.; Zong, L.; Hong, L.; Xiao, J.; Shen, T.; et al. Interpretable RNA Foundation Model from Unannotated Data for Highly Accurate RNA Structure and Function Predictions. *arXiv* **2022**, arXiv:2204.00300. [CrossRef]
52. Feng, Y.; Huang, S.-Y. ITScore-NL: An Iterative Knowledge-Based Scoring Function for Nucleic Acid–Ligand Interactions. *J. Chem. Inf. Model.* **2020**, *60*, 6698–6708. [CrossRef]
53. Li, J.; Zhu, W.; Wang, J.; Li, W.; Gong, S.; Zhang, J.; Wang, W. RNA3DCNN: Local and global quality assessments of RNA 3D structures using 3D deep convolutional neural networks. *PLoS Comput. Biol.* **2018**, *14*, e1006514. [CrossRef]
54. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379*, 1123–1130. [CrossRef] [PubMed]
55. Consortium, R. RNAcentral 2021: Secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.* **2020**, *49*, D212–D220. [CrossRef]
56. Sun, S.; Wu, Q.; Peng, Z.; Yang, J. Enhanced prediction of RNA solvent accessibility with long short-term memory neural networks and improved sequence profiles. *Bioinformatics* **2019**, *35*, 1686–1691. [CrossRef] [PubMed]
57. Zhuo, C.; Zeng, C.; Yang, R.; Liu, H.; Zhao, Y. RPflex: A Coarse-Grained Network Model for RNA Pocket Flexibility Study. *Int. J. Mol. Sci.* **2023**, *24*, 5497. [CrossRef]
58. Yang, R.; Liu, H.; Yang, L.; Zhou, T.; Li, X.; Zhao, Y. RPpocket: An RNA-Protein Intuitive Database with RNA Pocket Topology Resources. *Int. J. Mol. Sci.* **2022**, *23*, 6903. [CrossRef]
59. Zhou, T.; Wang, H.; Zeng, C.; Zhao, Y. RPocket: An intuitive database of RNA pocket topology information with RNA-ligand data resources. *BMC Bioinform.* **2021**, *22*, 428. [CrossRef]
60. Coleman, R.G.; Sharp, K.A. Finding and Characterizing Tunnels in Macromolecules with Application to Ion Channels and Pores. *Biophys. J.* **2009**, *96*, 632–645. [CrossRef]
61. Kawabata, T.; Go, N. Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins* **2007**, *68*, 516–529. [CrossRef] [PubMed]
62. Kawabata, T. Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins* **2010**, *78*, 1195–1211. [CrossRef]
63. Kawabata, T. Detection of cave pockets in large molecules: Spaces into which internal probes can enter, but external probes from outside cannot. *Biophys. Physicobiol* **2019**, *16*, 391–406. [CrossRef]
64. Voss, N.R.; Gerstein, M. 3V: Cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **2010**, *38*, W555–W562. [CrossRef]
65. Zhu, H.; Pisabarro, M.T. MSPocket: An orientation-independent algorithm for the detection of ligand binding pockets. *Bioinformatics* **2011**, *27*, 351–358. [CrossRef] [PubMed]
66. Trosset, J.Y.; Vodovar, N. Structure-based target druggability assessment. *Methods Mol. Biol.* **2013**, *986*, 141–164. [CrossRef]
67. Zhuo, C.; Zeng, C.; Liu, H.; Wang, H.; Peng, Y.; Zhao, Y. Advances and Mechanisms of RNA-Ligand Interaction Predictions. *Life* **2025**, *15*, 104. [CrossRef]
68. McDonald, I.K.; Thornton, J.M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, *238*, 777–793. [CrossRef] [PubMed]
69. Schwerdtfeger, P.; Wales, D.J. 100 Years of the Lennard-Jones Potential. *J. Chem. Theory Comput.* **2024**, *20*, 3379–3405. [CrossRef] [PubMed]
70. Nandy, A.; Shekhar, S.; Paul, B.K.; Mukherjee, S. Exploring the Nucleobase-Specific Hydrophobic Interaction of Cryptolepine Hydrate with RNA and Its Subsequent Sequestration. *Langmuir* **2021**, *37*, 11176–11187. [CrossRef]
71. Dezanet, C.; Kempf, J.; Mingeot-Leclercq, M.P.; Décout, J.L. Amphiphilic Aminoglycosides as Medicinal Agents. *Int. J. Mol. Sci.* **2020**, *21*, 7411. [CrossRef]
72. Panei, F.P.; Torchet, R.; Ménager, H.; Gkeka, P.; Bonomi, M. HARIBOSS: A curated database of RNA-small molecules structures to aid rational drug design. *Bioinformatics* **2022**, *38*, 4185–4193. [CrossRef]

73. Mallet, V.; Oliver, C.; Broadbent, J.; Hamilton, W.L.; Waldispühl, J. RNAglib: A python package for RNA 2.5 D graphs. *Bioinformatics* **2022**, *38*, 1458–1459. [CrossRef] [PubMed]
74. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]
75. Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Research* **2016**, *5*, 189. [CrossRef]
76. Zhang, Y.; Skolnick, J. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **2005**, *33*, 2302–2309. [CrossRef]
77. Zhang, C.; Shine, M.; Pyle, A.M.; Zhang, Y. US-align: Universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **2022**, *19*, 1109–1115. [CrossRef] [PubMed]
78. Guan, X.; Tang, Q.Y.; Ren, W.; Chen, M.; Wang, W.; Wolynes, P.G.; Li, W. Predicting protein conformational motions using energetic frustration analysis and AlphaFold2. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2410662121. [CrossRef]
79. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]
80. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. [CrossRef]
81. Oxenfarth, A.; Kümmerer, F.; Bottaro, S.; Schnieders, R.; Pinter, G.; Jonker, H.R.A.; Fürtig, B.; Richter, C.; Blackledge, M.; Lindorff-Larsen, K.; et al. Integrated NMR/Molecular Dynamics Determination of the Ensemble Conformation of a Thermodynamically Stable CUUG RNA Tetraloop. *J. Am. Chem. Soc.* **2023**, *145*, 16557–16572. [CrossRef]
82. Reißer, S.; Zucchelli, S.; Gustincich, S.; Bussi, G. Conformational ensembles of an RNA hairpin using molecular dynamics and sparse NMR data. *Nucleic Acids Res.* **2020**, *48*, 1164–1174. [CrossRef] [PubMed]

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.

pharmaceuticals

MDPI

Article

Pharmacokinetic Equations Applied to Obtain New Topological Models in the Search of Antibacterial Compounds

Jose I. Bueso-Bordils [1,*], Gerardo M. Antón-Fos [2], Rafael Martín-Algarra [1] and Pedro A. Alemán-López [1,*]

[1] Pharmacy Department, Universidad Cardenal Herrera-CEU, CEU Universities C/Ramón y Cajal s/n, Alfara del Patriarca, 46115 Valencia, Spain
[2] Department of Chemistry and Biochemistry, Faculty of Pharmacy, Universidad San Pablo-CEU, CEU Universities, Urbanización Montepríncipe, 28003 Madrid, Spain
* Correspondence: jose.bueso@uchceu.es (J.I.B.-B.); paleman@uchceu.es (P.A.A.-L.)

Abstract: Background: QSAR (Quantitative Structure–Activity Relationships) methods have been the basis for the design of new molecules with a certain activity. The great advantage of QSAR methods is that they can predict the pharmacological activity of compounds without the need to obtain or synthesize them previously. Currently, the development of antibiotic resistance by microorganisms is the most important issue in the treatment of infectious diseases. This elevated resistance is associated with expanded morbidity and mortality, as well as an increase in healthcare costs. The development of new molecules with antibacterial activity is therefore urgently needed. **Methods:** By means of molecular topology, we developed discriminant functions (DF1 and DF2) capable of predicting antibacterial activity. When applied to a database with 6373 chemicals, they selected 266 molecules as candidates, from which 41% have this activity, according to the bibliography. Regression equations determining pharmacokinetic properties such as mean residence time (MRT), volume of distribution (V_D), and clearance (CL) were applied to the selected molecules. **Results:** We have observed that most antibacterial compounds have pharmacokinetic theoretical values in the intervals $20 > MRT > 0$, $3 > V_D > 0$, and $500 > CL > 0$. We have applied these intervals to our antibacterial model with the objective of finding new antibacterials with a good pharmacokinetic profile. We show that they are an effective tool for discriminating antibacterial compounds, increasing the bibliographic success rate to 50.8, 59, and 61.5%, respectively. When drug-like filters are applied to these new models, the vast majority (89.9–100%) of the selected molecules present antibacterial activity. **Conclusions:** Considering these results, these new models could avoid the application of drug-likeness filters when searching for new potential antibacterials. All of this proves the usefulness of these mathematical–topological models.

Keywords: antibacterial; computational chemistry; pharmacokinetics; topological indices; drug-like; QSAR

1. Introduction

The introduction of antibacterial drugs into clinical practice in the 1940s was one of the most important interventions for its control and increased the life expectancy of the population by several years. Like vaccines, they saved millions of lives, but they also brought about a revolution in medicine, significantly contributing to progress in fields such as solid organ and hematopoietic stem cell transplantation, survival of premature

and immunosuppressed infants, and surgery of prosthetic material and vascular catheters, where infections are especially prevalent and important [1].

Since the mid-20th century, the pharmaceutical industry has intensified efforts to discover new antibiotic molecules from different microorganisms, preferably from soil samples or semisynthetic derivatives thereof. A great variety of these compounds belonging to very diverse families (beta-lactams, glycopeptides, macrolides, aminoglycosides, tetracyclines, etc.) were discovered. It was the golden age for these drugs, and it was believed that the battle against infectious diseases had already been won. In fact, infant mortality and morbidity fell considerably. Research was also conducted on the development of synthetic antimicrobials that were used in human and animal therapeutics, such as quinolones obtained from nalidixic acid, another synthetic agent. Quinolones constitute a widely used class of antibiotics due to their pharmacokinetic qualities, their bactericidal action against many pathogens, and their low toxicity [2].

However, soon after the discovery of antibiotics, the ability of bacteria to mutate and generate resistance to them was reported. Despite early warnings about the perils of their overuse, widespread over-prescription of antibiotics and their use in agriculture resulted in a rise in the growing resistance of some bacteria, such as pneumococcus and staphylococcus, to conventional antibiotics [3]. The extreme versatility and adaptability of microorganisms have prevented a decrease in the prevalence of these diseases since the number of antibiotic-resistant bacteria has multiplied exponentially in the last decades [4]. So much so that the World Health Organization (WHO) already considers antibiotic resistance developed by bacteria as one of the main threats to global health and points out that if the emergence of resistance continues at this rate, infections by antibiotic-resistant bacteria will be the leading cause of death worldwide, surpassing deaths caused by cancer, diabetes, and cardiovascular diseases [5].

The alarming increase in these resistances is, without a doubt, one of the biggest current public health problems since these compounds constitute one of the main tools for controlling and treating bacterial infections, both in human and veterinary medicine. This issue spans multiple sectors and presents several complex challenges. Aside from the medical components, it has economic, ecological, sociological, and developmental dimensions. The loss of effectiveness of these drugs would lead us to a silent pandemic that could make the world and the modern medicine built upon it collapse [6].

In the current context, rapid and inexpensive methods are necessary for the immediate expansion of the therapeutic arsenal that we currently possess. Within this framework, the drug repurposing strategy has emerged, which consists of using drugs that are already on the market and that have other therapeutic purposes to treat infections caused by multi-resistant bacteria [7–9]. This approach using known drugs is a common practice in the design and development of new drugs, saving time and resources in clinical and preclinical phases since, for example, toxicological and/or pharmacokinetic assays would have already been carried out previously [10].

In this sense, it is worth highlighting the utility of mathematical–topological models for the prediction of antibacterial activity, which allows obtaining new drugs with antibacterial capacity and which are alternative tools, given the reduction in investment by the pharmaceutical industry in this scope. Thus, within these methods, one can find the Quantitative Structure–Activity Relationship (QSAR) methods, which play an important role since they provide useful information for rational drug design and development with minimal cost. The great advantage of these QSAR methods is that they can predict the pharmacological activity of a compound without the need to previously obtain or synthesize it. This has caused a great rise in the number of studies regarding Computational Chemistry and Virtual Combinatorial Chemistry in the last years [11]. Specifically, in our research

group, we use molecular topology, which consists of obtaining a series of topological indices (TIs) that are calculated from the structure of a molecule. In this way, we can classify a molecular structure as active or not active against a certain pharmacological activity using pattern recognition techniques such as linear discriminant analysis, multilinear regression, or random decision forests [12,13].

This field offers an alternative and inexpensive method for the rapid identification of new antibacterial agents. Of special interest are antibiotics with optimum pharmacokinetic properties. Besides modeling biological activity endpoints, QSAR methods can also involve modeling the pharmacokinetic properties of drug candidates prior to their synthesis, ensuring the efficacy of the designed compounds inside biological systems [14]. The inclusion of the prediction of pharmacokinetic and/or pharmacodynamic properties, such as the ability to cross the blood–brain barrier, prediction of cytochrome P450-mediated metabolism, albumin binding, or water solubility, attempt to simultaneously optimize potency and pharmacokinetics [15].

We therefore reiterate QSAR methods as an effective tool to perform rapid virtual screening for potential antibacterials, allowing the prediction of other pharmacokinetic and toxicological properties directly related to activity to find safer and more effective drugs. Thus, one promising possibility appears to be the combination of multiple QSAR models that will generate the most suitable antibacterial compounds, considering not only their activity but also their properties. We want to apply pharmacokinetic prediction equations to generate new antibacterial models with the objective of finding antibacterials with good pharmacokinetic profiles. Furthermore, by introducing pharmacokinetic values, these models select drug-like molecules. Also, the clear definition of a model's applicability domain (AD) is crucial since it enhances confidence in the predictions and enables the practical application of QSAR models [16]. Indeed, the Organization for Economic Cooperation and Development (OECD) requires assessing the AD of QSAR models as a prerequisite for using QSAR in regulatory contexts [17].

As pharmaceutical companies increasingly step back from antibiotic research—largely due to high costs and limited financial incentives—there is an urgent need for smarter, more affordable ways to discover new treatments. This is where QSAR modeling comes in. Unlike traditional methods that require physically creating and testing compounds, QSAR can predict a molecule's potential just by analyzing its structure. It is a practical and efficient alternative to high-throughput screening, which demands extensive lab resources, or AI and machine learning approaches that often need massive, well-curated datasets. QSAR strikes a balance: it is accessible, cost-effective, and scientifically robust. Especially when combined with molecular topology, it becomes a powerful tool for quickly identifying promising antibacterial candidates—something we desperately need in the fight against resistant bacteria.

2. Results and Discussion

After extending thorough bibliographic research on antibacterial activity from a previous study [18], a total of 266 molecules from our dataset were selected by the AB model as potential antibacterials. Of these, 109 have proven antibacterial activity bibliographically. That is, considering only the scientific literature, the AB model correctly classified as antibacterial in 41% of the selected compounds.

To build the discriminant functions, we used four groups: two training groups with 38 active and 38 inactive compounds and two test groups with 12 active and 11 inactive compounds, which allowed for assessing the quality of the selected discriminant functions.

The chosen functions were as follows:

$$DF1 = -27.13819 - 4.7164S_{=C<} + 1.27264S_{\text{-}NH\text{-}} - 0.30296S_{=O} + 43.39831^{0}C - 13.57036^{3}C_{c} \quad (1)$$
$$N = 76 \quad \lambda = 0.2653522 \quad F = 38.76$$

$$DF2 = -3.27384 - 3.14402S_{=C<} + 0.66311S_{\text{-}NH2} + 1.44735S_{\text{-}NH\text{-}} + 1.15103S_{=N\text{-}} - 0.47139S_{\text{-}OH} - 4.41362S_{aSa} \quad (2)$$
$$N = 76 \quad \lambda = 0.3138095 \quad F = 25.146$$

where the quality of the discriminant functions is evaluated by the Wilks' λ parameter (also known as U-statistic), which is obtained by a multivariate analysis of variance statistics that tests the equality of group mean values for the variables in the DFs.

We plotted the corresponding PDDs for every function to visualize the values of the function in which the probability of classifying a compound as active or inactive is maximum. In other words, to find areas where the overlap between the two groups of molecules is minimal. The maximum expectancy zone for new antibacterials is the range of 0–12 for DF1 and the range of 2–12 for DF2 in both the training and the test sets [18].

With the purpose of finding three mathematical models able to identify antibacterials with good pharmacokinetic profiles, RFs built based on the bibliographic information of a limited number of antibacterial quinolones were applied to the same chemical graphs used for DF1 and DF2, only that this time there were two, not four, large sets of molecules: one with 50 antibacterial quinolones and another one comprising 49 inactive quinolones and quinolone-related structures.

We aimed to build three models by combining DF1 and DF2 with a pharmacokinetic equation, which related, respectively, the mean residence time (MRT) in hours, the volume of distribution (V_D) in L/kg, and the clearance (CL) in mL/min [19,20]:

$$MRT = -79.3319 - 4534.37^{10}\chi^{v}_{ch} + 13.6939^{2}\kappa_{\alpha} - 2.21175S_{\text{-}O\text{-}} \quad (3)$$
$$N = 14 \quad r^2 = 0.85378 \quad r^2_{cv} = 0.74026 \quad SEE = 6.306649 \quad F = 19.46$$

$$V_D = 10.1212 + 0.704187S_{\text{-}NH2} - 0.612674S_{\text{-}Cl} - 0.615152G_2 \quad (4)$$
$$N = 16 \quad r^2 = 0.85816 \quad r^2_{cv} = 0.75868 \quad SEE = 0.694465 \quad F = 24.20$$

$$CL = 8981.08 + 1596.68^{6}\chi_{ch} - 109.481S_{>N\text{-}} - 2637.68NI_2 \quad (5)$$
$$N = 12 \quad r^2 = 0.92043 \quad r^2_{cv} = 0.70282 \quad SEE = 65.42953 \quad F = 30.85$$

where r^2_{cv} is obtained by cross-validation analysis using the leave-one-out method [21].

Since the equations have been obtained from antibacterial quinolones, it is expected that the molecules selected by the models maintain the same mechanism of action (inhibition of DNA gyrase and/or topoisomerase IV) and similar pharmacokinetic properties (MRT, V_D, or CL).

After the selection of the three equations constituting the pharmacokinetic property in the topological model, we built the corresponding PDDs. Application of these equations to the two large sets of drugs (one comprising compounds exhibiting antibacterial activity and the other not possessing it) reveals that most of the actives are gathered within a specific interval, while the inactives are generally scattered or fall outside this interval. The PDDs were obtained by using Equations (3)–(5) (MRT, V_D, and CL, respectively) along with the highest property intervals for each function, which are shown in Figures 1–3. Thus, the intervals derived from these PDDs establish the AD for each model.

The derived discriminant functions are applicable not only to novel quinolone derivatives outside the scope of the original dataset but also to structurally unrelated compounds, provided they exhibit analogous mathematical–topological descriptors. The robustness of the topological approach has been substantiated through its successful implementation across a broad spectrum of pharmacological classes [13].

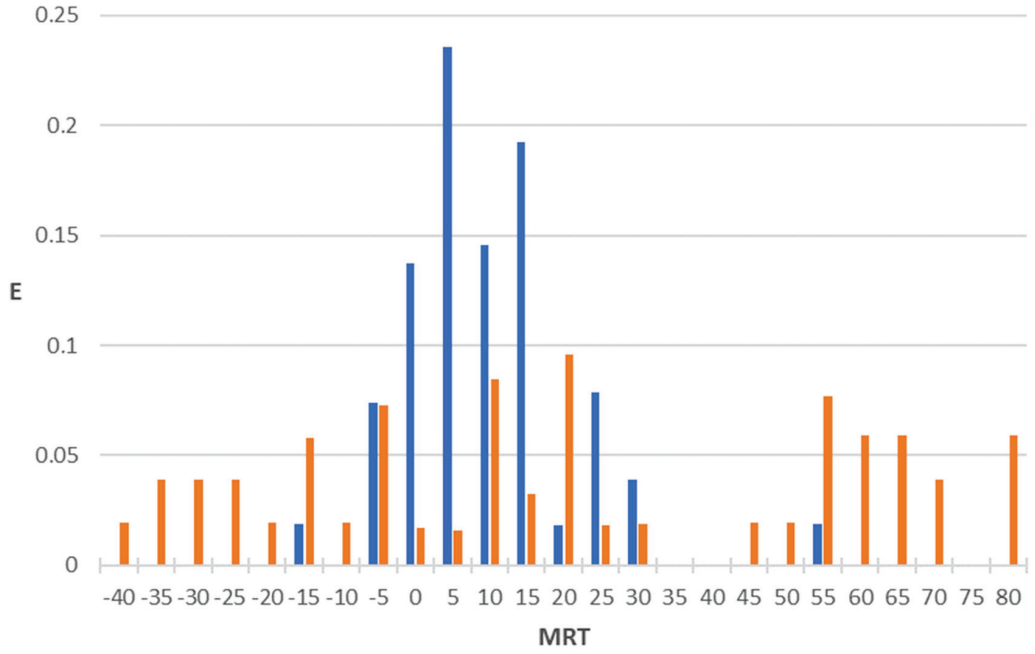


Figure 1. PDD for mean residence time (hours) property as obtained by using the regression function MRT: blue bars, antibacterial compounds; orange bars, non-antibacterial compounds. Highest property expectancy (E): 0 < MRT < 20.

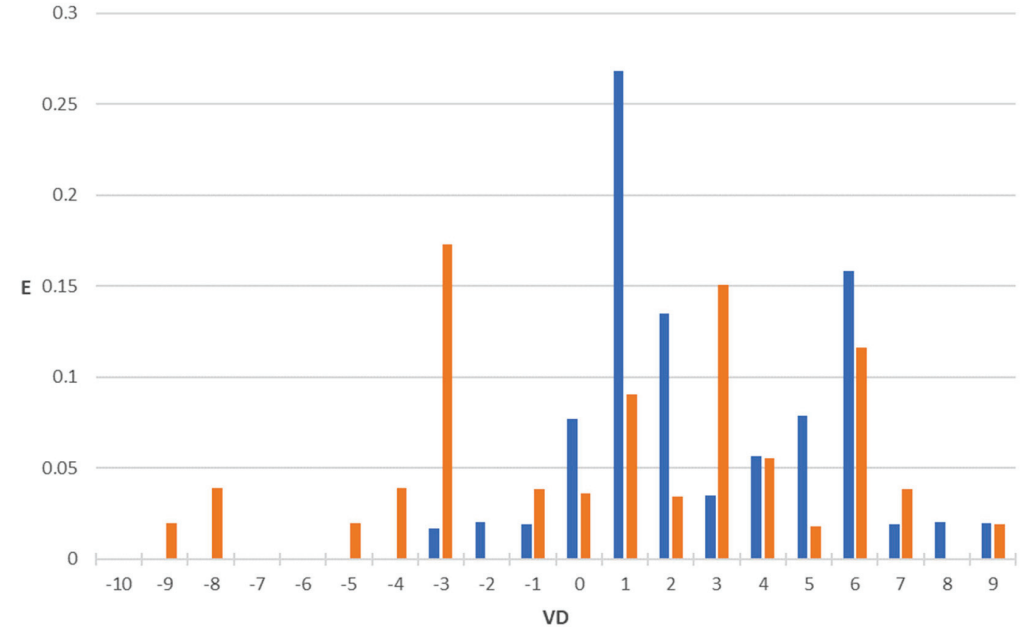


Figure 2. PDD for volume of distribution (L/kg) property as obtained by using the regression function V_D: blue bars, antibacterial compounds; orange bars, non-antibacterial compounds. Highest property expectancy (E): $0 < V_D < 3$.

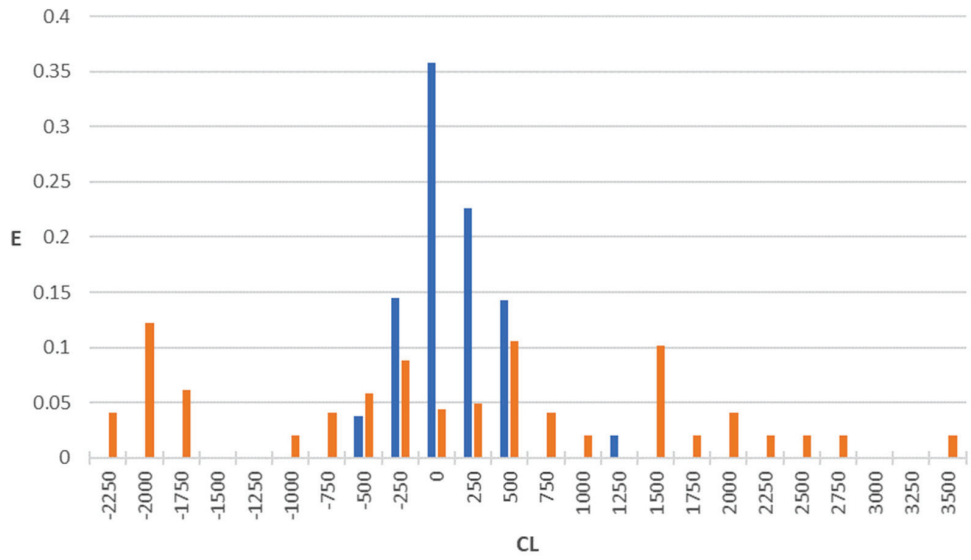


Figure 3. PDD for clearance (mL/min) property as obtained by using the regression function CL: blue bars, antibacterial compounds; orange bars, non-antibacterial compounds. Highest property expectancy (E): 0 < CL < 500.

We observe that some calculated values are negative (Figures 1–3). The explanation for this fact is that the function assigns these negative values to very small values of the property. Such is the case of antibacterial quinolones with observed pharmacokinetic values within the first positive interval (MRT: 0–5 h; VD: 0–1 L/kg; CL: 0–250 mL/min), whose calculated pharmacokinetic values are negative.

2.1. Antibacterial + Mean Residence Time Model (AB+MRT Model)

Considering the obtained results, an optimized topological MRT model for antibacterial activity can be formulated through DF1, DF2, and MRT (see Equations (1)–(3) and Figure 1), where DF1 and DF2 address antibacterial activity and MRT addresses mean residence time property. A molecule would be classified as active if it fulfilled the following requirements:

$$12 > DF1 > 0$$

$$12 > DF2 > 2$$

$$20 > MRT > 0$$

After applying all three functions (Equations (1)–(3)) to our dataset, we identified 59 compounds as theoretically active because their DF1, DF2, and MRT values (AB+MRT model) fell within the predefined intervals. Of these, 30 compounds (50.8%) had demonstrated antibacterial activity, according to the bibliography (see Supplementary Section S2).

Figure 1 indicates that the vast majority of antibacterials fall within the MRT equation's selected interval of calculated values (74% of the antibacterials in the 20 > MRT > 0 interval), while non-antibacterial compounds' calculated values are mostly dispersed (only 18.4% in the same interval). It must be noted that the scientific literature generally refers to half-life ($t_{1/2}$) or beta-phase half-life ($bpt_{1/2}$) rather than MRT, where MRT = $t_{1/2}$/ln2 in the case of one-compartment models and MRT = $bpt_{1/2}$/ln2 in the case of two-compartment models [22]. This consistency MRT experimental values, where first- and second-generation quinolones such as enoxacin, norfloxacin or ciprofloxacin with low to medium elimination rate (2 h $\leq t_{1/2} \leq$ 4 h or 2.89 h $\leq$ MRT $\leq$ 5.77 h) while later quinolones such as gatifloxacin, moxifloxacin, or trovafloxacin, whose structures' optimization usually led to

longer elimination rates ($t_{1/2} \leq 13$ h or MRT ≤ 18.76 h) [23–32], supports the validity of this model.

The encouraging performance of the AB+MRT model suggests that integrating pharmacokinetic parameters could refine our screening process. With this in mind, we next turn our attention to the volume of distribution.

2.2. Antibacterial + Volume of Distribution Model (AB+VD Model)

The volume of distribution (V_D), which is the volume of body fluid into which a drug's dose is dissolved, is an important determinant of drug concentration. An optimized topological V_D model for antibacterial activity can be formulated through DF1, DF2, and V_D (see Equations (1), (2), and (4) and Figure 2), where DF1 and DF2 address antibacterial activity and V_D provides additional insight into the compounds' volume of distribution. A molecule would be classified as active if it fulfilled the following requirements:

$$12 > DF1 > 0$$

$$12 > DF2 > 2$$

$$3 > V_D > 0$$

By applying the proposed topological model (AB+V_D model) to our dataset, 39 compounds were selected as theoretically active, of which 22 (59%) were accurately classified as active according to the bibliography (see Supplementary Section S2).

In the case of quinolones, their V_D is generally greater than that of total body water, which suggests a wide distribution of these compounds throughout the body due to sequestration in the intracellular fluid of certain tissues due to their moderately lipophilic nature [22,33,34]. However, key structural modifications to improve the pharmacokinetics of third- and fourth-generation quinolones, such as the addition of an amino or methyl group in R_5 (e.g., sparfloxacin and grepafloxacin, respectively) or the presence of piperazine or azabiclylic ring in R_7 (e.g., gatifloxacin and garenoxacin, respectively) increased the quinolones' lipophilicity and, thus, their experimental values of V_D, resulting in a high variability among the group [24–32,35,36].

Figure 2 indicates that antibacterials tend to cluster within the V_D equation's selected interval of calculated values (52% of antibacterials in the $3 > V_D > 0$ interval), while non-antibacterial compounds' calculated values are more scattered (only 20.4% fall in this interval). Therefore, these results align well with known distribution patterns observed in quinolones, where modifications such as the addition of amino or methyl groups have led to significant variations in V_D.

Building upon the insights gained from the AB+MRT model, the AB+V_D model adds another pharmacokinetic characterization. In our continuous effort to enhance predictive accuracy, we now extend our approach to include clearance.

2.3. Antibacterial + Clearance Model (AB+CL Model)

Clearance (CL) is a key parameter that indicates the volume of body fluid from which the drug is completely removed per unit time and is critical for understanding a drug's elimination and duration of action. An optimized topological CL model for antibacterial activity can be formulated by combining DF1, DF2, and CL (see Equations (1), (2), and (5) and Figure 3), where DF1 and DF2 assess antibacterial activity while CL further refines the

profile based on elimination characteristics. A molecule would be classified as active if it fulfilled the following requirements:

$$12 > DF1 > 0$$

$$12 > DF2 > 2$$

$$500 > CL > 0$$

When applying the AB+CL model to our dataset, 39 compounds were selected as potential antibacterials, of which 24 (61.5%) were confirmed as active, according to the bibliography (see Supplementary Section S2).

Regarding clearance (CL), changes in quinolones' elimination were observed over time: early quinolones (e.g., enoxacin, levofloxacin and lomefloxacin) were eliminated primarily by renal clearance, with high observed CL values, whereas later quinolones (e.g., grepafloxacin, moxifloxacin and trovafloxacin) were modified to be more lipophilic, and hence eliminated primarily by hepatic clearance, which tended to show lower observed CL values [2,24–32].

Figure 3 illustrates that antibacterial compounds' calculated values are strongly clustered in the CL equation's selected interval (62% of antibacterials in the 500 > CL > 0 interval), in contrast to non-antibacterial agents, of which only 12.2% fall within this range. Thus, these findings are consistent with the changes observed over time in quinolone clearance, where shifting from primarily renal to hepatic elimination results in differing CL values.

Hence, by incorporating additional pharmacokinetic parameters, our antibacterial activity models refine the selection of compounds with optimal pharmacokinetic profiles. This approach not only minimizes false positives but also enhances the overall drug-likeness of the selected candidates.

2.4. Study of Drug-like Properties in Compounds Selected by Models

To demonstrate in vivo activity of a substance, in addition to having the correct stereo-electronic properties to bind with the receptor, it must be safe and have the appropriate pharmacokinetic properties to reach it. This means that most of the designed compounds do not successfully succeed in phase I clinical trials [37].

The search for novel antibacterial agents is carried out based on a two DF classification model and their PDDs applied to antibacterials and non-antibacterials achieved in previous work [38]. After the application of such a model (as described under Materials and Methods) to a large dataset comprising 6373 chemicals (including either drugs or chemical reactants), those 266 molecules passing the model requirements were selected (4.2% of the total) of which, considering only the bibliographic data, the AB model correctly classified as antibacterial at least 41% of the selected compounds.

Later, the same group of antibacterials and non-antibacterials is screened for three different pharmacokinetic properties (functions MRT, $V_{D,}$ and CL and their respective PDDs [19,20]), which leads to 3 different models to search for potential antibacterial with optimum pharmacokinetic properties.

When three functions along with their PDDs are used, the accuracy for active compounds decreases, whereas that for inactive ones increases, so the probability of a false active compound being selected decreases. In these cases, the proportion of correctly classified active compounds rose to 50.8, 59, and 61.5% (AB+TRM, AB+$V_{D,}$ and AB+CL, respectively). This suggests that the values obtained from Equations (3)–(5) are related to

TIs that are somehow measures of antibacterial activity, so these equations can also be used as discriminant functions for this type of activity.

Since Lipinski's rules, many drug-like (DL) properties have been developed to predict the appropriate pharmacokinetic behavior of substances. Thus, an extra evaluation of drug-likeness by application of constitutional indices (number of atoms, nAT [39]), molecular properties (Ghose–Crippen molar refractivity, AMR [40]; topological polar surface area, TPSA [41]; rotatable bond number, RBN [42]) and DL indices (Lipinski's rule of five, Ro5 [43]), key for oral absorption, is performed using alvaMolecule [38] to all discussed models, including the previously published antibacterial model [18] (see Supplementary Section S3). Table 1 summarizes the percentage of correct classification of the different models obtained for the DL filters selected (Table 1).

Table 1. Percentage of molecules from the starting dataset and different models that fulfill the drug-like filters.

DL Filter [a]	IM Dataset [b]	AB Model [c]	AB+MRT Model [d]	AB+V_D Model [e]	AB+CL Model [f]
Ro5 $\geq$ 0.75	75.8	91.6	100	100	100
20 $\leq$ nAT $\leq$ 70	77.5	91.6	100	97.4	100
40 < AMR < 130	82.6	93.4	100	94.6	100
TPSA < 140	78.9	82.8	89.9	97.4	94.9
1 $\leq$ RBN $\leq$ 9	81.7	85.3	100	92.1	94.9

[a] Ro5: Lipinski's rule of five; nAT: number of atoms; AMR: molar refractivity; TPSA: topological polar surface area; RBN: rotatable bond number. [b] Percentage of molecules from the IM dataset that fulfill each of the DL filters. [c] Percentage of molecules from the IM dataset that are theoretically active according to the AB model and fulfill each of the DL filters. [d] Percentage of molecules from the IM dataset that are theoretically active according to the AB+MRT model and fulfill each of the DL filters. [e] Percentage of molecules from the IM dataset that are theoretically active according to the AB+V_D model and fulfill each of the DL filters. [f] Percentage of molecules from the IM dataset that are theoretically active according to the AB+MRT model and fulfill each of the DL filters.

As can be seen in Table 1, all the pharmacokinetic equations increased the percentage of molecules with DL properties with respect to the initial database and the molecules selected by the AB model, which increases the probability of finding active molecules with optimum pharmacokinetic properties.

While each of the three models—AB+MRT, AB+V_D, and AB+CL—demonstrated improved predictive power over the base AB model, their performance varied in terms of accuracy and drug-likeness enrichment. The AB+CL model achieved the highest bibliographic success rate, correctly identifying 61.5% of compounds as antibacterial, followed closely by AB+V_D (59%) and AB+MRT (50.8%). In terms of drug-likeness, all three models significantly outperformed the original dataset, with the AB+MRT and AB+CL models achieving 100% compliance with key filters such as Lipinski's Ro5 and molar refractivity. These findings suggest that while all models are effective, the AB+CL model offers the best overall balance between predictive accuracy and pharmacokinetic relevance, making it a particularly promising tool for the virtual screening of antibacterial candidates.

Finally, the compounds showing a previously known antibacterial activity according to an exhaustive literature search were taken out of the pool of selected candidates so that we finally obtained a group of novel and (potentially) active candidates. We believe that the remaining compounds that, up to our knowledge, have not been previously described as antibacterials (see Supplementary Section S4) could be strong candidates to increase the currently available antimicrobial arsenal. We think that only these data justify and demonstrate the great utility of these new models.

3. Materials and Methods

3.1. Compound Selection

A mathematical–topological antibacterial prediction model (AB model) previously described by our research group [18] is used to find new antibacterial compounds from a database retrieved from the Index Merck (IM) [44], which gathers 10250 marketed compounds, including pharmaceutical drugs, biological products, and chemical reactants. A series of modifications explained below had to be carried out to transform them into chemical graphs that our index calculation software programs, DESMOL13 [45] and MOLCONN-Z [46], could use. The molecular descriptors used were computed from the adjacency topological matrix obtained from the hydrogen-depleted graphs using the drawing program ChemDraw 12.0 Professional from the ChemOffice 2010 software package (these descriptors are described in Supplementary Section S1 along with their definitions and references).

Firstly, stereochemistry is removed from all molecules since the software only supports 2D graphs. This had to be performed one molecule at a time, changing the bonds that indicated chirality for flat bonds. Furthermore, this software does not accept molecules with more than 99 bonds from end to end of the molecule. Thus, molecular entities exceeding 99 bonds had to be removed. Moreover, we were only able to calculate the indexes of molecules containing C, F, Cl, Br, I, O, N, S, P, B, Si, Ge, Sn, Pb, and H. Therefore, each molecule had to be analyzed to detect any other atoms and delete them. Moreover, an automatized data curation of the molecular structures (duplicate analysis, removal of undesired characteristics such as mixture, salts, polymers, organometallic, and inorganic substances, fixing erroneous molecular representation, and checking for anomalies) is performed using alvaMolecule 2.0.6 [38]. The curation of molecular data can be one of the most time-consuming phases of the model-building process; it often requires human expertise to check molecular structures, even manually, to identify potential problems, and it certainly is a mandatory task when working with software tools that are not able to calculate molecular descriptors for such types of chemical structures. Having taken all these considerations, we were able to build our own dataset with 6373 compounds of approved and experimental drugs (62.2% of the IM database), whose chemical graphs were drawn and saved in "mol" format.

3.2. Linear Discriminant Analysis and Multilinear Regression

Linear Discriminant Analysis (LDA) and Multiple Linear Regression (MLR) are commonly employed techniques in the development of QSAR models. However, a significant limitation of these methods is their inability to handle variables with discrete values, as such data cannot be effectively analyzed using these statistical approaches since this type of index would lead to overfitting [47].

The aim of LDA, a heuristic algorithm capable of distinguishing between two or more categories of entities, is to find linear functions that allow one to discriminate between those categories. In one of our previous studies [18], two large sets comprising quinolones with proven antibacterial activity and quinolones and quinolone-related structures without such activity, respectively, were considered. Quinolones represent a well-characterized and diverse class of compounds, offering a rich dataset that enhances the accuracy and reliability of the predictive equations [47].

Chemical graphs used to calculate the AB model formed by two discriminant functions (DF1 and DF2) were randomly split into training (molecules used to build the DFs) and test (molecules used to ensure the quality of the DFs) groups for each set using the statistical package called BMDP module 7D [48]. The discriminant ability of antibacterial activity is

assessed in terms of the proportion of correct classification in each group, and a search of potential antibacterial compounds is performed in a big dataset with a vast majority of non-related structures [18], which identified substances with antibacterial activity previously described in the bibliography (41% of the selected compounds). Moreover, we also demonstrated that it is possible to obtain a high degree of molecular characterization of certain pharmacokinetic properties (mean residence time after oral administration, MRT; volume of distribution, V_D; clearance, CL) by an adequate choice of TIs in the form of regression functions (RF) [19,20] using MLR, performed by a BMDP 7.0 software module 9R [48].

Experimental procedures used to obtain property values were carefully reviewed. Specifically, activities against various microorganisms expressed as different MIC values—derived from in vitro assays conducted in accordance with CLSI guidelines [49]. Additionally, data obtained from studies involving healthy human subjects were used to determine pharmacokinetic properties.

Randomness tests were conducted to detect potential spurious correlations in the regression models. This involved randomly permuting the property values of each compound and performing linear regression with the selected independent variables. The outcomes of these tests were visualized by plotting the correlation coefficient (r^2) against the cross-validated prediction coefficient (r^2_{cv}), providing a graphical assessment of model robustness [21]. It is also recommended that the number of indices chosen is well below the number of molecules used to build the functions to avoid possible fortuitous correlations so that the optimal relationship between them is the one that achieves a better prediction with the minimum number of descriptors possible, in order to avoid random variation as much as possible [50].

3.3. Pharmacological Distribution Diagrams

In recent work [19,20], we found that connectivity functions can be used not only to predict properties or classify chemical structures according to activity classification but also to predict the activity of a given compound in a specific pharmacological context. This is accomplished by using pharmacological distribution diagrams (PDD), which allow us to estimate the probability of a given molecule being active. These diagrams have been proven useful with various therapeutic categories (e.g., antihistamines, bronchodilators, and anti-tubercular drugs [51–53]).

After selecting the DFs and RFs, we built the corresponding pharmacological distribution diagrams (PDD). These plots are useful to determine the intervals of the discriminant function in which the expectancy, E, to find active molecules is maximum. PDDs are histogram-like plots of connectivity functions in which the expectancies appear on the ordinate axis [54]. For an arbitrary interval of values of a given function, we can define the expectancy of activity as $E_a = a/(i + 1)$, where "a" is the number of active compounds in the interval divided by the total number of active compounds, and "i" is the number of inactive compounds. The expectancy of inactivity is defined in a symmetrical way, as $E_i = i/(a + 1)$. This representation provides good visualization of the regions of minimum overlap and selects regions in which the probability of finding improved molecules is maximum.

3.4. Topological Models

By means of MLR and LDA techniques, three topological mathematical models comprising three functions per model were achieved for antibacterial activity. The different topological models were formed by combining both DFs (AB model) with an RF. Furthermore, PDDs allowed us to carry out the assignment of thresholds useful to discriminate active from inactive compounds with the highest probability of success. Only the molecules

predicted as active by all three functions comprised in each of the models within the predetermined thresholds were identified as potential antibacterial drugs with an adequate pharmacokinetic profile. This way, the number of false actives will be minimized.

4. Conclusions

Currently, the emergence of antibiotic resistance in microorganisms is one of the biggest issues that have appeared in recent years in the treatment of infectious diseases. MT has proven to be a powerful tool in the design and development of new antibacterials, considering the increasing appearance of multi-resistant bacteria and the lack of investment from the pharmaceutical industry. Furthermore, RFs aiming at predicting pharmacokinetic properties have been proven to have the ability to group active compounds effectively, which suggests a close relationship between them and the antibacterial activity of compounds. By combining them with DFs, we developed three QSAR models with clear-cut intervals in which the probability of finding active molecules reaches its maximum according to the corresponding PDDs.

In this study, we applied these three models to a dataset derived from the IM database, identifying molecules with antibacterial activity previously described in the bibliography with outstanding discriminant ability according to the success rates. The model has also selected compounds whose antibacterial activity, up to our knowledge, has not been proven yet, which could be interesting candidates to be studied, providing new possible action mechanisms and thus increasing contribution to tackle antibiotic resistance.

These models provide useful information for a better understanding of how the molecular structure may affect the antibacterial activity of a molecule. With this information, it is simpler to understand how topological characterization of chemical structures may affect the antibacterial activity of a compound and therefore optimize existing drugs, which is a cost-effective approach to obtain new treatments against resistant bacteria.

We can conclude that the results obtained in this study confirm QSAR and molecular topology as extremely useful and cost- and time-effective tools, where DL filters could be avoided in the virtual screening of new antibacterials. The new models generated could be of application in drug repurposing, where we find thousands of safe and efficient candidates to be tested for new pharmacological activities.

Thus, having demonstrated again the efficacy of this predictive tool, our present and future work is focused towards their validation by means of experimental in vitro testing.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/ph18060865/s1. Supplementary Section S1: Symbols and Definitions of Topological Indices used with DESMOL13 and MOLCONN-Z programs; Supplementary Section S2: Compounds selected by the models as candidates with proven antibacterial activity; Supplementary Section S3: SMILES file with the list of molecules and properties calculated by alvaMolecule; Supplementary Section S4: Compounds selected by the models as candidates without proven antibacterial activity. References [55–78] have been cited to Supplementary Materials.

Author Contributions: Conceptualization, P.A.A.-L.; validation, P.A.A.-L. and G.M.A.-F.; formal analysis, G.M.A.-F. and R.M.-A.; investigation, R.M.-A.; data curation, J.I.B.-B. and P.A.A.-L.; writing—original draft preparation, J.I.B.-B.; writing—review and editing, P.A.A.-L.; supervision, G.M.A.-F.; project administration, G.M.A.-F.; funding acquisition, G.M.A.-F. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by University CEU Cardenal Herrera (GIR23/08 and INDI23/08).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained in the paper.

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

References

1. Pagalilauan, G.L.; Limaye, A.P. Infections in transplant patients. *Med. Clin. N. Am.* **2013**, *97*, 581–600. [CrossRef] [PubMed]
2. Pham, T.D.M.; Ziora, Z.M.; Blaskovich, M.A.T. Quinolone antibiotics. *MedChemComm* **2019**, *10*, 1719–1739. [CrossRef] [PubMed]
3. Levy, S.B.; Marshall, B. Antibacterial resistance worldwide: Causes, challenges and responses. *Nat. Med.* **2004**, *10*, S122–S129. [CrossRef] [PubMed]
4. Vila, J.; Moreno-Morales, J.; Ballesté-Delpierre, C. Current landscape in the discovery of novel antibacterial agents. *Clin. Microbiol. Infect.* **2020**, *26*, 596–603. [CrossRef]
5. World Health Organization. The WHO AWaRe (Access, Watch, Reserve) Antibiotic Book. 2022. Available online: https://apps.who.int/iris/bitstream/handle/10665/365237/9789240062382-eng.pdf (accessed on 20 April 2023).
6. Alós, J.I. Antibiotic resistance: A global crisis. *Enferm. Infecc. Microbiol. Clin.* **2015**, *33*, 692–699. [CrossRef]
7. Pushpakom, S.; Iorio, F.; Eyers, P.A.; Escott, K.J.; Hopper, S.; Wells, A.; Doig, A.; Guilliams, T.; Latimer, J.; McNamee, C.; et al. Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **2019**, *18*, 41–58. [CrossRef]
8. Pinzi, L.; Bisi, N.; Rastelli, G. How drug repurposing can advance drug discovery: Challenges and opportunities. *Front. Drug Discov.* **2024**, *4*, 1460100. [CrossRef]
9. Suay-Garcia, B.; Bueso-Bordils, J.I.; Falco, A.; Perez-Gracia, M.T.; Anton-Fos, G.; Aleman-Lopez, P. Quantitative structure-activity relationship methods in the discovery and development of antibacterials. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1472. [CrossRef]
10. Hopkins, A.L. Network pharmacology: The next paradigm in drug discovery. *Nat. Chem. Biol.* **2008**, *4*, 682–690. [CrossRef]
11. Suay-García, B.; Bueso-Bordils, J.I.; Falcó, A.; Antón-Fos, G.M.; Alemán-López, P.A. Virtual combinatorial chemistry and pharmacological screening: A short guide to drug design. *Int. J. Mol. Sci.* **2022**, *23*, 1620. [CrossRef]
12. Suay-Garcia, B.; Falcó, A.; Bueso-Bordils, J.I.; Anton-Fos, G.M.; Pérez-Gracia, M.T.; Alemán-López, P.A. Tree-based QSAR model for drug repurposing in the discovery of new antibacterial compounds against *Escherichia coli*. *Pharm.* **2020**, *13*, 431. [CrossRef] [PubMed]
13. Bueso-Bordils, J.I.; Alemán-López, P.A.; Martín-Algarra, R.; Duart, M.J.; Falcó, A.; Antón-Fos, G.M. Molecular topology for the search of new anti-MRSA compounds. *Int. J. Mol. Sci.* **2021**, *22*, 5823. [CrossRef] [PubMed]
14. Muratov, E.N.; Bajorath, J.; Sheridan, R.P.; Tetko, I.V.; Filimonov, D.; Poroikov, V.; Oprea, T.I.; Baskin, I.I.; Varnek, A.; Roitberg, A.; et al. QSAR without borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564. [CrossRef]
15. Farha, M.A.; Brown, E.D. Drug repurposing for antimicrobial discovery. *Nat. Microbiol.* **2019**, *4*, 565–577. [CrossRef]
16. Hanser, T.; Barber, C.; Marchaland, J.F.; Werner, S. Applicability domain: Towards a more formal definition. *SAR QSAR Environ. Res.* **2016**, *27*, 865–881. [CrossRef]
17. Organization for Economic Cooperation and Development. OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure–Activity Relationship Models. Available online: https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf (accessed on 26 May 2025).
18. Bueso-Bordils, J.I.; Aleman, P.A.; Zamora, L.L.; Martin-Algarra, R.; Duart, M.J.; Antón-Fos, G.M. Topological model for the search of new antibacterial drugs. 158 theoretical candidates. *Curr. Comput. Aided Drug Des.* **2015**, *11*, 336–345. [CrossRef] [PubMed]
19. Bueso-Bordils, J.I.; Aleman-López, P.A.; Costa-Piles, S.; Duart, M.J.; Lahuerta-Zamora, L.; Martin-Algarra, R.; Anton-Fos, G.M. Obtaining microbiological and pharmacokinetic highly predictive equations. *Curr. Top Med. Chem.* **2018**, *18*, 908–916. [CrossRef]
20. Bueso-Bordils, J.I.; Antón-Fos, G.M.; Falcó, A.; Duart, M.J.; Martín-Algarra, R.; Alemán-López, P.A. New pharmacokinetic and microbiological prediction equations to be used as models for the search of antibacterial drugs. *Pharmaceuticals* **2022**, *15*, 122. [CrossRef]
21. Allen, D.M. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* **1974**, *16*, 125–127. [CrossRef]
22. Levison, M.E.; Levison, J.H. Pharmacokinetics and pharmacodynamics of antibacterial agents. *Infect. Dis. Clin. N. Am.* **2009**, *23*, 791–815. [CrossRef]
23. Adjei, M.D.; Deck, J.; Heinze, T.M.; Freeman, J.P.; Williams, A.J.; Sutherland, J.B. Identification of metabolites produced from *N*-phenylpiperazine by *Mycobacterium* spp. *J. Ind. Microbiol. Biotechnol.* **2007**, *34*, 219–224. [CrossRef] [PubMed]
24. Dong, Y.; Xu, C.; Zhao, X.; Domagala, J.; Drlica, K. Fluoroquinolone action against mycobacteria: Effects of C-8 substituents on growth, survival, and resistance. *Antimicrob. Agents Chemother.* **1998**, *42*, 2978–2984. [CrossRef] [PubMed]

25. Scoper, S.V. Review of third-and fourth-generation fluoroquinolones in ophthalmology: In-vitro and in-vivo efficacy. *Adv. Ther.* **2008**, *25*, 979–994. [CrossRef]
26. Bolon, M.K. The newer fluoroquinolones. *Infect. Dis. Clin. N. Am.* **2009**, *23*, 1027–1051. [CrossRef]
27. Rusu, A.; Lungu, I.A.; Moldovan, O.L.; Tanase, C.; Hancu, G. Structural characterization of the millennial antibacterial (fluoro) quinolones-shaping the fifth generation. *Pharmaceutics* **2021**, *13*, 1289. [CrossRef]
28. Cojutti, P.G.; Ramos-Martin, V.; Schiavon, I.; Rossi, P.; Baraldo, M.; Hope, W.; Pea, F. Population pharmacokinetics and pharmacodynamics of levofloxacin in acutely hospitalized older patients with various degrees of renal function. *Antimicrob. Agents Chemother.* **2017**, *61*, e02134-16. [CrossRef]
29. Ball, P. Moxifloxacin (Avelox): An 8-methoxyquinolone antibacterial with enhanced potency. *Int. J. Clin. Pract.* **2000**, *54*, 329–332. [CrossRef]
30. Brighty, K.E.; Gootz, T.D. The chemistry and biological profile of trovafloxacin. *J. Antimicrob. Chemother.* **1997**, *39*, 1–14. [CrossRef]
31. Gai, X.; Shen, N.; He, B.; Zhou, Q.; Bo, S.; Li, X.; Zhai, S.; Yin, A.; Lu, W. Population pharmacokinetics of ciprofloxacin in Chinese elderly patients with lower respiratory tract infection. *Zhonghua Yi Xue Za Zhi* **2015**, *95*, 1581–1585.
32. Arjona, A. Nemonoxacin quinolone antibiotic. *Drug Future* **2009**, *34*, 196–203.
33. Butranova, O.I.; Ushkalova, E.A.; Zyryanov, S.K.; Chenkurov, M.S.; Baybulatova, E.A. Pharmacokinetics of antibacterial agents in the elderly: The body of evidence. *Biomedicines* **2023**, *11*, 1633. [CrossRef]
34. Póvoa, P.; Moniz, P.; Pereira, J.G.; Coelho, L. Optimizing antimicrobial drug dosing in critically ill patients. *Microorganisms* **2021**, *9*, 1401. [CrossRef]
35. Domagala, J.M. Structure-activity and structure-side-effect relationships for the quinolone antibacterials. *J. Antimicrob. Chemother.* **1994**, *33*, 685–706. [CrossRef]
36. Gootz, T.D.; Brighty, K.E. Fluoroquinolone antibacterials: SAR, mechanism of action, resistance, and clinical aspects. *Med. Res. Rev.* **1996**, *16*, 433–486. [CrossRef]
37. Di, L.; Kerns, E.H. *Drug-like Properties Concepts, Structure Design and Methods from ADME to Toxicity Optimization*, 2nd ed.; Academic Press: Cambridge, MA, USA, 2016.
38. Mauri, A.; Bertola, M. Alvascience: A new software suite for the QSAR workflow applied to the blood-brain barrier permeability. *Int. J. Mol. Sci.* **2022**, *23*, 12882. [CrossRef]
39. Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623. [CrossRef]
40. Viswanadhan, V.N.; Ghose, A.K.; Revankar, G.R.; Robins, R.K. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
41. Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43*, 3714–3717. [CrossRef]
42. Oprea, T.I.; Gottfries, J.; Sherbukhin, V.; Svensson, P.; Kühler, T.C. Chemical information management in drug discovery: Optimizing the computational and combinatorial chemistry interfaces. *J. Mol. Graph. Model.* **2000**, *18*, 512–524. [CrossRef]
43. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2012**, *64*, 4–17. [CrossRef]
44. O'Neil, M.J. *The Merck Index*, 13th ed.; Merck & Co. Inc.: Rahway, NJ, USA, 2001.
45. *DESMOL Software*, version 13; Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Facultad de Farmacia, Universitat de València: Valencia, Spain, 2000.
46. Hall, L.H. *MOLCONN-Z Software*, version 3.0; Eastern Nazarene College: Quincy, MA, USA, 1995.
47. Roy, K.; Kar, S.; Das, R.N. *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2015.
48. Dixon, W.J. *BMDP Statistical Software*, version 7.0; University of California: Berkeley, CA, USA, 1990.
49. CLSI. *CLSI Performance Standards for Antimicrobial Susceptibility Testing*; Twenty-Second Informational Supplement; CLSI Document M100-S22; Clinical and Laboratory Standards Institute: Wayne, PA, USA, 2012; Volume 32.
50. Hawkins, D.M. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12. [CrossRef]
51. Duart, M.J.; García-Domenech, R.; Gálvez, J.; Alemán, P.A.; Martín-Algarra, R.V.; Antón-Fos, G.M. Application of a mathematical topological pattern of antihistaminic activity for the selection of new drug candidates and pharmacology assays. *J. Med. Chem.* **2006**, *49*, 3667–3673. [CrossRef]
52. Balasubramaniyan, D.; Chidambaramb, N. On some neighbourhood degree-based topological indices with QSPR analysis of asthma drugs. *Eur. Phys. J. Plus* **2023**, *138*, 823. [CrossRef]

53. García-García, Á.; Julián-Ortiz, J.V.; Gálvez, J.; Font, D.; Ayats, C.; Guna Serrano, M.D.R.; Muñoz-Collado, C.; Borrás, R.; Villalgordo, J.M. Similarity-Based Virtual Screening to Find Antituberculosis Agents Based on Novel Scaffolds: Design, Syntheses and Pharmacological Assays. *Int. J. Mol. Sci.* **2022**, *23*, 15057. [CrossRef]
54. Gálvez, J.; García-Domenech, R.; de Gregorio Alapont, C.; de Julián-Ortiz, J.V.; Popa, L. Pharmacological distribution diagrams: A tool for de novo drug design. *J. Mol. Graph.* **1996**, *14*, 272–276. [CrossRef]
55. Kier, L.B.; Hall, L.H. General definition of valence delta-values for molecular connectivity. *J. Pharm. Sci.* **1983**, *72*, 1170–1173. [CrossRef] [PubMed]
56. Gálvez, J.; García-Domenech, R.; Salabert, M.T.; Soler, R. Charge indexes. New topological descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520–525. [CrossRef]
57. Kier, L.B.; Hall, L.H. The E-state as an extended free valence. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 548–552. [CrossRef]
58. Basak, S.C.; Mills, D. Quantitative structure-property relationships (QSPRs) for the estimation of vapor pressure: A hierarchical approach using mathematical structural descriptors. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 692–701. [CrossRef]
59. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Champaign, IL, USA, 1949.
60. Moliner, R.; Garcia, F.; Galvez, J.; Garcia-Domenech, R.; Serrano, C. Nuevos índices topológicos en conectividad molecular. Su aplicación a algunas propiedades fisicoquímicas de un grupo de hidrocarburos alifáticos. *An. Real Acad. Farm.* **1991**, *57*, 287–298.
61. Cummins, D.J.; Andrews, C.W.; Bentley, J.A.; Cory, M. Molecular diversity in chemical databases: Comparison of medicinal chemistry knowledge bases and databases of commercially available compounds. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 750–763. [CrossRef] [PubMed]
62. Bonchev, D.; Trinajstic, N. On topological characterization of molecular branching. *Int. J. Quantum Chem.* **1978**, *14*, 293–303. [CrossRef]
63. Kier, L.B. A shape index from molecular graphs. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109–116. [CrossRef]
64. Kier, L.B. Distinguishing atom differences in a molecular graph shape index. *Quant. Struct.-Act. Relat.* **1986**, *5*, 7–12. [CrossRef]
65. Kier, L.B. An index of molecular flexibility from Kappa shape attributes. *Quant. Struct.-Act. Relat.* **1989**, *8*, 221–224. [CrossRef]
66. Suay-García, B.; Alemán-López, P.; Bueso-Bordils, J.I.; Falcó, A.; Pérez-Gracia, M.T.; Antón-Fos, G.M. Opological index Nclass as a factor determining the antibacterial activity of quinolones against *Escherichia coli*. *Future Med. Chem.* **2019**, *11*, 2255–2262. [CrossRef] [PubMed]
67. Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20. [CrossRef]
68. Wiener, H. Relation of the physical properties of the isomeric alkanes to molecular structure. Surface tension, specific dispersion, and critical solution temperature in aniline. *J. Phys. Chem.* **1948**, *52*, 1082–1089. [CrossRef]
69. Randic, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. Wiener matrix invariants. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 361–367. [CrossRef]
70. Niederfellner, J.; Lenoir, D.; Matuschek, G.; Rehfeldt, F.; Utschick, H.; Brügemann, R. Description of vapor pressures of polycyclic aromatic compounds by graph theoretical indices. *Quant. Struct.-Act. Relat.* **1997**, *16*, 38–48. [CrossRef]
71. Platt, J.R. Influence of neighbor bonds on additive bond properties in paraffins. *J. Chem. Phys.* **1947**, *15*, 419–420. [CrossRef]
72. Harras, M.F.; Sabour, R.; Farghaly, T.A.; Ibrahim, M.H. Drug repurposing approach in developing new furosemide analogs as antimicrobial candidates and anti-PBP: Design, synthesis, and molecular docking. *Bioorg. Chem.* **2023**, *137*, 106585. [CrossRef] [PubMed]
73. Wien, R.; Harrison, J.; Freeman, W.A. Diamidines as antibacterial compounds. *Brit. J. Pharmacol.* **1948**, *3*, 211–218. [CrossRef]
74. Szatlóczky, E. Cause, diagnosis, and chemotherapy of lactose intolerance. *Br. Med. J.* **1982**, *284*, 1405. [CrossRef]
75. Wasi, N.; Singh, H.B. In vitro evaluation of bacteriostatic activity of metal complexes of amodiaquine and primaquine. *Inorg. Chim. Acta* **1988**, *151*, 287–289. [CrossRef]
76. Bichowsky-Slomnitzki, L. The effect of aromatic diamidines on bacterial growth. *J. Bacteriol.* **1948**, *55*, 27–31. [CrossRef]
77. Piotrowski, W.; Sas-Piotrowska, B.; Slizak, W. Response of *Rhizobacterium leguminosarum* (strain B-73B) to some fungicides and antibiotics. *Prog. Plant Prot.* **1999**, *39*, 845–847.
78. Kamath, A.; Ojima, I. Advances in the chemistry of β-lactam and its medicinal applications. *Tetrahedron* **2012**, *68*, 10640–10664. [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.

pharmaceuticals

MDPI

Article

The Allosteric Regulation of the DNA-Binding Domain of p53 by the Intrinsically Disordered C-Terminal Domain

Shangbo Ning [1,*,†], Chengwei Zeng [2,†], Huiwen Wang [3], Junfeng Zhang [1], Yun Xue [1] and Yunjie Zhao [2,*]

[1] School of Medical Technology and Engineering, Henan University of Science and Technology, Luoyang 471023, China; jfzhang018@163.com (J.Z.); xueyun6688@163.com (Y.X.)
[2] Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan 430079, China; cwzengwuhan@mails.ccnu.edu.cn
[3] School of Physics and Engineering, Henan University of Science and Technology, Luoyang 471023, China; huiwenwang@haust.edu.cn
* Correspondence: sning@haust.edu.cn (S.N.); yjzhaowh@mail.ccnu.edu.cn (Y.Z.)
† These authors contributed equally to this work.

Abstract

Background: Intrinsically disordered regions (IDRs) within proteins often act as pivotal linkage units for the interaction of functional domains. The p53 tumor suppressor protein contains intrinsically disordered N-terminal and C-terminal domains (NTD and CTD), playing crucial regulatory roles in cellular processes. Furthermore, experimental approaches have encountered challenges in elucidating the structural regulation by the IDRs. **Methods**: In this work, we employed microsecond-scale molecular dynamics simulations to explore the allosteric regulation mechanism of the p53 DNA binding domain (DBD) induced by the CTD and the DNA binding. Subsequently, we integrated dynamic cross-correlation analysis with binding free energy calculations to evaluate the interaction between the CTD and DNA. **Results**: The free energy landscapes (FELs) were utilized to identify the conformational ensemble of the p53 DBD. The FELs revealed that the CTD enhances the allosteric regulatory mechanisms. **Conclusions**: Firstly, the conformation of DBD changes on the S6-S7 loop and L1 upon DNA binding. Then the CTD directly interacts with DNA and further regulates the allosteric network (involving the S6-S7 loop, L1 loop, S4, S10, H1, and H3) to promote the binding of DBD to DNA. The allosteric mechanisms presented in this work will provide new insights into the functional mechanisms of the p53 CTD and inform the rational design of p53-targeted drugs.

Keywords: intrinsically disordered proteins; allosteric mechanism; free energy landscapes; molecular dynamics; MM-PBSA

1. Introduction

Intrinsically disordered regions (IDRs) in proteins are segments of at least 30 amino acids in length that lack stable three-dimensional structures under physiological conditions [1,2]. IDRs play crucial roles in various biological functions, including molecular recognition, protein–protein interactions, and regulation of cellular processes [3–5]. Based on previous studies, approximately 44% of human proteins contain disordered regions [6]. These regions are characterized by their dynamic nature, which allows them to sample a wide range of conformations and interact with multiple partners, often through short linear motifs [7,8]. IDRs are enriched in polar and charged amino acids, such as serine, threonine, and glutamic acid [9,10].

These characteristics contribute to the disordered state due to reduced hydrophobic interactions and a higher net charge.

The p53 tumor suppressor protein is a transcription factor that maintains cellular stability and prevents cancer [11]. The p53 modulates fundamental cellular processes through protein–protein interactions [12]. Structurally, the p53 comprises four subunits of 393 amino acids and contains IDRs in both N-terminal and C-terminal domains. As shown in Figure 1, the N-terminal domain (NTD, residues: 1–90) includes two transactivation domains (TAD1, residues: 1–40, TAD2, residues: 40–60) and a proline-rich region (PRR, residues: 60–90) [13]. TAD1 contains the primary binding sites for MDM2 and MDMX. TAD2 is involved in transcriptional activation and protein interactions, such as with the RPA protein [14]. Mutations in TAD1 and TAD2 are required to eliminate tumor suppressor function [15]. TAD1 and TAD2 also contain multiple phosphorylation sites that regulate the degradation and activity of p53 in response to stress [16]. The DNA binding domain (DBD: 94–312) is essential for sequence-specific DNA binding and is the target of most point mutations in cancer [17]. The C-terminal domain (CTD, residues: 312–393) contains an oligomerization domain (residues: 323–355) and a lysine-rich C-terminal tail (residues: 364–393), which includes sites for acetylation, methylation, and phosphorylation [18].

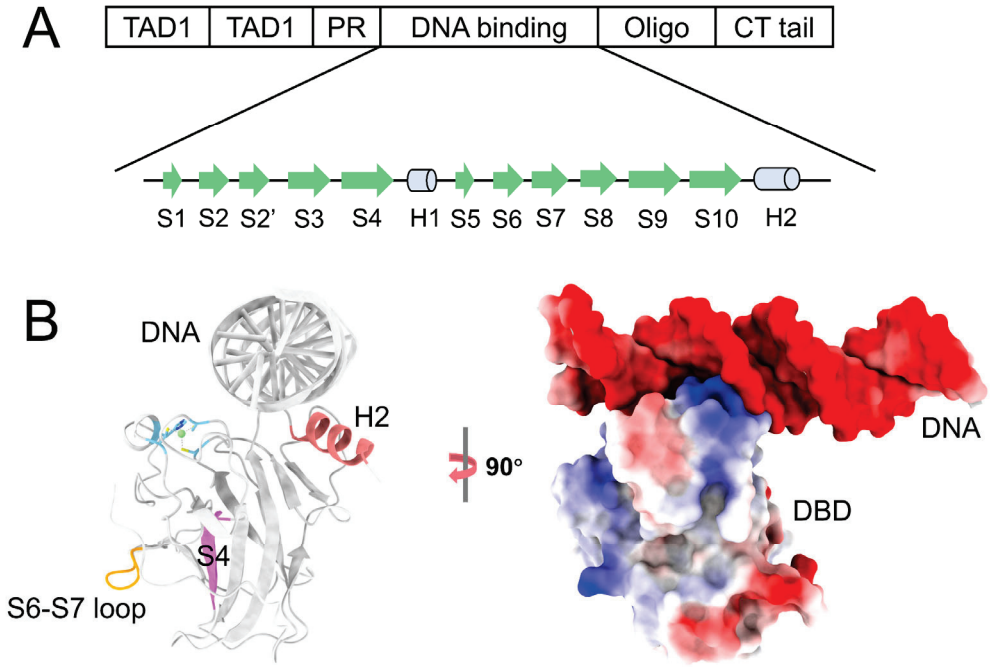


Figure 1. Structural domains of p53 and the conformation of its DNA-binding domain (DBD). (**A**) Schematic representation of the domain architecture of p53. The N-terminal domain (NTD) comprises two transactivation domains (TAD1: residues 1–40; TAD2: 40–60) and a proline-rich domain (PRD: 60–90). The central core corresponds to the DNA-binding domain (DBD: 94–312). The C-terminal domain (CTD) includes an oligomerization domain (OD: 323–355) and a basic, lysine-rich regulatory tail (364–393). (**B**) Structure of the p53 DBD in complex with double-stranded DNA. The DBD adopts a β-sandwich fold, with the DNA-binding interface primarily formed by helix H2. This region contains negatively charged residues critical for DNA interaction.

Functionally, the DBD is the main functional region of p53 that is particularly crucial for the cytoplasmic and nuclear functions of the protein [19]. In addition, the intrinsically disordered N-terminal and C-terminal domains also have important regulatory functions. For instance, the N-terminal transactivation domain of p53 may be involved in multiple functions. TAD1 and TAD2 interact with transcription factors such as TBP, TFIID, and coac-

https://doi.org/10.3390/ph19010124

tivators like p300/CBP, regulating signaling pathways during DNA damage [20]. The NTD is phosphorylated at multiple sites, exerting diverse functions through the involvement of various protein kinases [21]. The CTD may modulate DNA binding, and the resulting stimulation plays a key role in subsequent signaling pathways [22,23]. The acetylation of the CTD by p300/CBP can stimulate specific DNA binding in vitro. The CT tail interacts with the DBD in subunit-subunit interactions, promoting p53 oligomerization [24]. The CTD can bind to non-specific DNA and enable p53 to slide along the DNA, which is an important step in finding specific binding sites [25]. Therefore, a series of molecular experiments implies the key roles of these IDRs in DNA binding and regulation.

Previous studies for the p53 structure primarily focus on the DBD and DNA binding [26–28]. Lambruschini et al.'s study adopted classical molecular dynamics (MD) simulations and Metadynamics to unveil the allosteric mechanisms of the p53 DBD during DNA binding [29]. The local perturbation within the p53 structural network can lead to the transmission of a protein structure. On the other hand, a recent study suggests that the p53 NTD can inhibit binding to DNA and increase p53 DNA binding specificity [13]. The observations for CTD lead to a model for allosteric regulation of p53 and modulate the conversion between latent and activated forms of p53 [21]. However, the mechanism by which the CTD regulates the structure and function of the DBD remains unclear. To understand the regulatory mechanism of p53 functional activity, obtaining the structural details and dynamic characteristics between CTD and DBD upon DNA binding is urgent.

In this work, the conformational dynamics of the p53 DBD and the DNA binding affinity upon CTD-DNA binding were studied by all-atom molecular dynamics (MD) simulations. We observe a significant correlation between the structural characteristics of the S6-S7 loop and the global allosteric network of p53. These conformational shifts are pivotal to the functionality of p53, potentially impacting both its DNA binding affinity and its capacity for transcriptional regulation. Furthermore, the CTD engages in direct interactions with DNA, thereby amplifying the DNA-binding affinity. The CTD further promotes allosteric regulation and conformational transition by stabilizing DBD binding to DNA. The global allosteric network of DBD comprises residues F212, R280, Q167, V122, R156, and R267. Our results indicate that the CTD regulates the global allosteric network of the DBD, including the S6-S7 loop, etc. The structural details provided in this article highly coincide with the experimental observations. These discoveries provide novel insights into the allosteric regulatory mechanisms of p53 and the disease pathology.

2. Results

The structural ensemble of p53 DBD and dynamic changes is crucial for understanding the mechanisms of transcriptional regulation and other functions. Previous studies suggested that the conformation of the DBD changes after binding to DNA [29]. p53_DBD as the core DNA-binding domain alone, and full-length p53 as the DBD flanked by both the N-terminal and C-terminal intrinsically disordered regions (IDRs). Therefore, these IDRs, particularly the CTD, may not merely be passive tethers but actively modulate the structure, dynamics, and allosteric regulation of the DBD upon DNA binding. To elucidate the regulatory role of the CTD region for the p53 and DNA binding, we integrated microsecond-scale unbiased dynamics, dynamic correlation networks, and free-energy landscapes to analyze the conformational dynamics and thermodynamic changes.

To observe the structural property of the core domain upon DNA binding, we characterized the dynamic changes in DBD in both $p53_{DBD\text{-}DNA}$ and $p53_{ALL\text{-}DNA}$ models. As shown in Figure 2A, we calculated the root-mean-square deviations (RMSDs) of the $p53_{DBD}$ and $p53_{ALL}$ structures upon DNA binding. After a 500 ns equilibration, the RMSD values

https://doi.org/10.3390/ph19010124

for all four systems fluctuated within a certain range, suggesting that the system had generally reached a stable state. Notably, the RMSDs of the DBD region in both experimental and theoretical models show significant changes upon DNA binding. The DNA binding induces larger fluctuations in the overall structure of DBD. The R_g value decreases from 16.6 Å to 16.4 Å ($p53_{DBD\text{-}DNA}$) and the $p53_{ALL\text{-}DNA}$ decreases from 16.8 Å to 16.6 Å upon DNA binding (Figure 2B). It suggests the protein structures are more compact upon binding to DNA, aligning with a transition to a more stable conformation. In our analysis of the root-mean-square fluctuations (RMSFs) of amino acid residues within the p53 protein, significant fluctuations are observed in the heavy atoms of residues V143-M160, H179-H193, V218-T230, T253-N268, and A276-K291 (Figure 2C,D). These fluctuations indicate that the DBD adopts different conformations upon DNA binding.

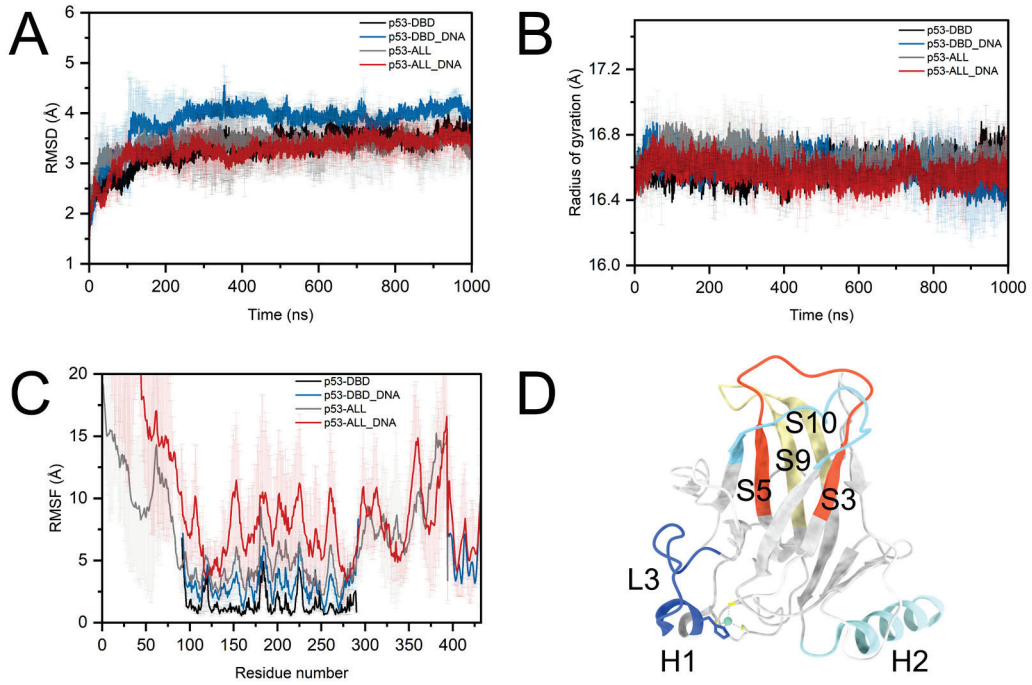


Figure 2. Conformational dynamics of $p53_{DBD}$ and $p53_{ALL}$ upon DNA binding. (**A**) Root-mean-square deviation (RMSD) of the DNA-binding domain (DBD) in the four simulated systems: $p53_{DBD}$, $p53_{ALL}$, $p53_{DBD\text{-}DNA}$, and $p53_{ALL\text{-}DNA}$. (**B**) Radius of gyration (Rg) of the DBD across the four systems. (**C**) Root-mean-square fluctuation (RMSF) of backbone heavy atoms in the four systems. (**D**) Structural mapping of regions exhibiting significant conformational flexibility upon DNA binding, with residues color-coded according to their RMSF values.

2.1. The Internal Contact Within the p53 DBD Changes upon DNA Binding

The residue contact maps are employed to delineate the interaction changes within the p53 DBD upon DNA binding (Figures 3, S2 and S3). As shown in Figure 3A, the contact probability of the DBD ($p53_{DBD\text{-}DNA}$) increases by 60.0–80.0%. The dynamic regions are mapped onto the three-dimensional structure of the DBD (Figure 3B), mainly located at the zinc finger binding region and the DNA binding interface (A276-K291). In the contact map of the $p53_{ALL\text{-}DNA}$, the contact probability of the DBD increases by over 60.0% upon DNA binding compared to the $p53_{DBD\text{-}DNA}$ simulations (Figure 3C,D). Consequently, the interaction network within the DBD also undergoes significant changes upon DNA binding, with a notable increase in internal contacts. The internal correlations of the DBD are further enhanced in the $p53_{ALL\text{-}DNA}$ simulation.

To further elucidate the binding mode of the DBD with DNA and the conformational changes in the DBD upon DNA binding, the free energy landscapes were employed to

https://doi.org/10.3390/ph19010124

analyze the structural ensemble of the DBD. The x-axis represents the RMSDs of the p53 DBD, and the y-axis represents the R_g of the DBD (Figure S4), describing the structural fluctuation and stability. Three major minima of the DBD conformation (states 1, 2, and 3) are characterized by the free energy landscapes upon DNA binding, and the representative structures of the minima are given in Figure S4. The population of structures in state 1 is 57.5% (Figure S4A). The population of state 2 is 68.2% upon DNA binding (Figure S4B). Moreover, a new sub-conformation (state 3) appears after DNA binding, and the population is about 19.2%. The FEL of the $p53_{DBD\text{-}DNA}$ suggests that the DNA binding promotes the population of structures 2 and 3 (the closed state of the S6-S7 loop), and the S6-S7 loop is closer to the flanking S4 β-sheet. Previous studies have also pointed out similar observations regarding the structural heterogeneity of the S6-S7 loop [26,29]. In addition, the conformations of the L1 loop in states 2 and 3 (Figure S4C) are closer to the DNA binding site (H2 helix), indicating that L1 also plays a role in DNA binding. We further selected the distances between two sets of amino acid residues as collective variables to probe the correlation between the conformational changes in the S6-S7 loop and allosteric regulation. As shown in Figure S5A,B, the distance between the two groups of amino acid residues is approximately 13.0 Å. The FEL shows subtle changes in the two sets of distances. Therefore, the DNA binding alters the free energy landscape of the p53 DBD and promotes population shifts in the conformation (S6-S7 loop).

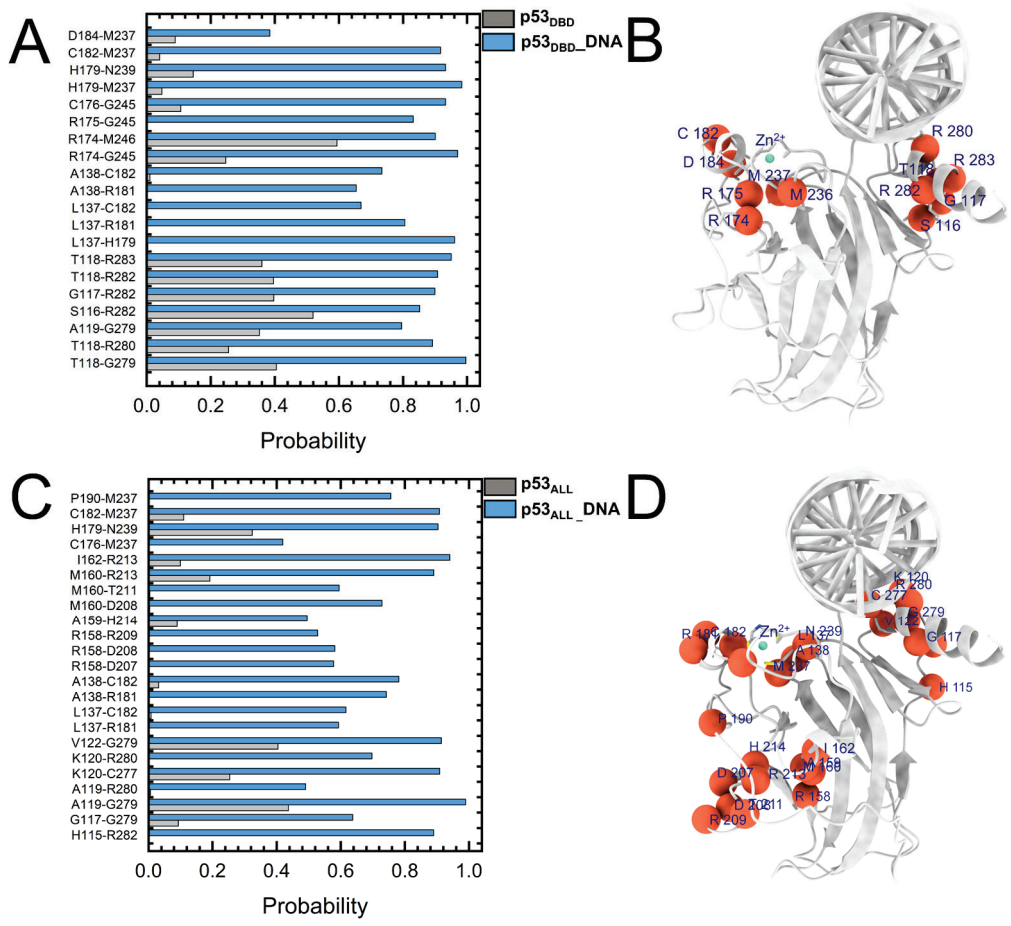


Figure 3. Changes in internal residue contacts within the p53 DBD upon DNA binding. (**A**) Residue pairs exhibiting high contact probabilities in the $p53_{DBD\text{-}DNA}$. (**B**) Residues contributing to altered internal interactions within the DBD in the $p53_{DBD\text{-}DNA}$ system. (**C**) Residue pairs with high contact probabilities in the $p53_{ALL\text{-}DNA}$. (**D**) Residues associated with changes in DBD internal interactions in the $p53_{ALL\text{-}DNA}$ system.

https://doi.org/10.3390/ph19010124

2.2. The CTD Stabilizes the Specific Interactions Between p53 and DNA

To analyze the regulation of CTD for the internal network of the DBD, we calculated the dynamic cross-correlation coefficients (DCCs) of DBD and the difference in the DCCs between the $p53_{DBD\text{-}DNA}$ and $p53_{ALL\text{-}DNA}$ systems upon DNA binding (Figures S6 and 4). Positive values (red) indicate an increase in the magnitude of correlation, while negative values (blue) indicate a decrease. In the DCCs of the $p53_{DBD\text{-}DNA}$ system, the correlations of multiple regions increase by more than 50.0% upon DNA binding (Figures 4A,B and S7). The position of these regions is consistent with the residue contact map. Compared to the residue contact map, the DCCs are more acute for the motion of the internal networks. For the $p53_{ALL\text{-}DNA}$, the network of the DCCs within DBD is amplified upon DNA binding and involves the global DBD (Figures 4D,E and S8). Meanwhile, Figure S2 shows the interactions between the CTD and DNA, suggesting that the CTD may regulate the allosteric networks by binding to DNA. We also compared changes in the DCC network of DBD following the incorporation of the CTD upon DNA binding. As shown in Figure 4C,F, the incorporation of CTD resulted in significantly enhanced positive correlations among four regions within DBD (including S127, S166, G187, F212, E285, etc.). This result likely corresponds to a more active conformational state that facilitates DNA binding.

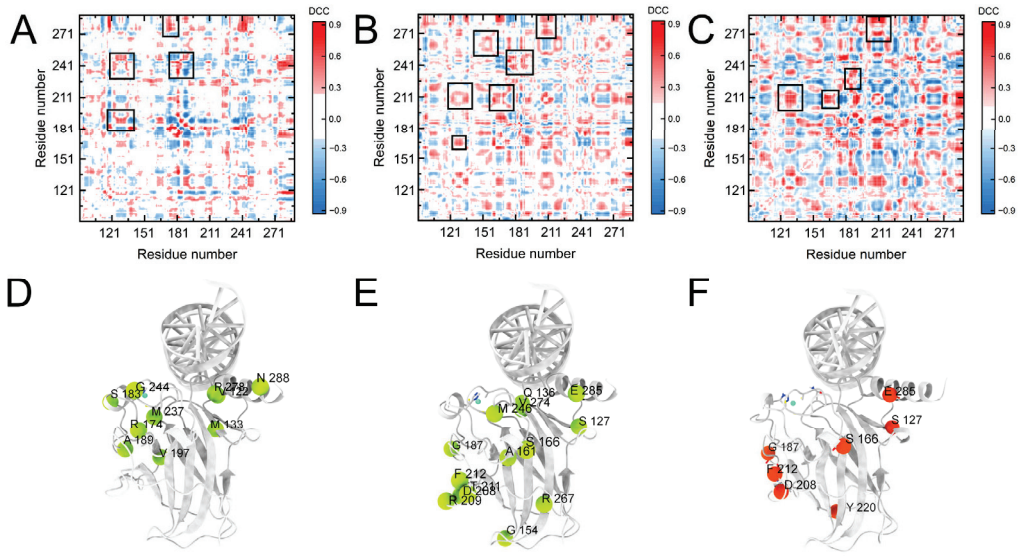


Figure 4. Dynamic correlation changes within the DBD upon DNA binding. (**A**–**C**) Difference maps of residue–residue dynamic correlation coefficients. Blue and red grids indicate decreased and increased correlations, respectively. The black boxes highlight regions of significantly enhanced positive correlation within the two systems. (**A**) $p53_{DBD\text{-}DNA}$ vs. $p53_{DBD}$. (**B**) $p53_{ALL\text{-}DNA}$ vs. $p53_{ALL}$. (**C**) $p53_{ALL\text{-}DNA}$ vs. $p53_{DBD\text{-}DNA}$. (**D**–**F**) Structural mapping of residues involved in dynamic network reorganization. (**D**) $p53_{DBD\text{-}DNA}$ vs. $p53_{DBD}$. (**E**) $p53_{ALL\text{-}DNA}$ vs. $p53_{ALL}$. (**F**) $p53_{ALL\text{-}DNA}$ vs. $p53_{DBD\text{-}DNA}$. Regions with altered dynamic correlations in $p53_{DBD\text{-}DNA}$ are localized near the DNA-binding interface, whereas changes in $p53_{ALL\text{-}DNA}$ involve the entire DBD.

As shown in Figure S2B, the contact probability of the DNA binding interface (H2 helix) with DNA is only around 0.5. In contrast, the contact probability of amino acid residues near the H2 helix with DNA in the $p53_{ALL\text{-}DNA}$ system has increased to 0.8–0.9 (corresponding to the black grid). To further examine the impact of the CTD on the stability of p53 binding to DNA, we compared the residue contact coordination numbers between proteins and DNA in both the $p53_{DBD\text{-}DNA}$ and $p53_{ALL\text{-}DNA}$ systems. In Figure 5A, the result shows that the coordination number in the $p53_{ALL\text{-}DNA}$ system increases from around 25 ($p53_{DBD\text{-}DNA}$) to 80 ($p53_{ALL\text{-}DNA}$). In the full-length p53-DNA system, the interaction

https://doi.org/10.3390/ph19010124

between the DBD and DNA is enhanced, and the CTD provides a significant contribution to the binding free energy with DNA (Figure 5B). These results suggest that the CTD enhances the binding interaction between p53 and DNA (Figure 5C,D) and also confirm the role of the CTD in regulating DNA binding, as mentioned in experiments [22].

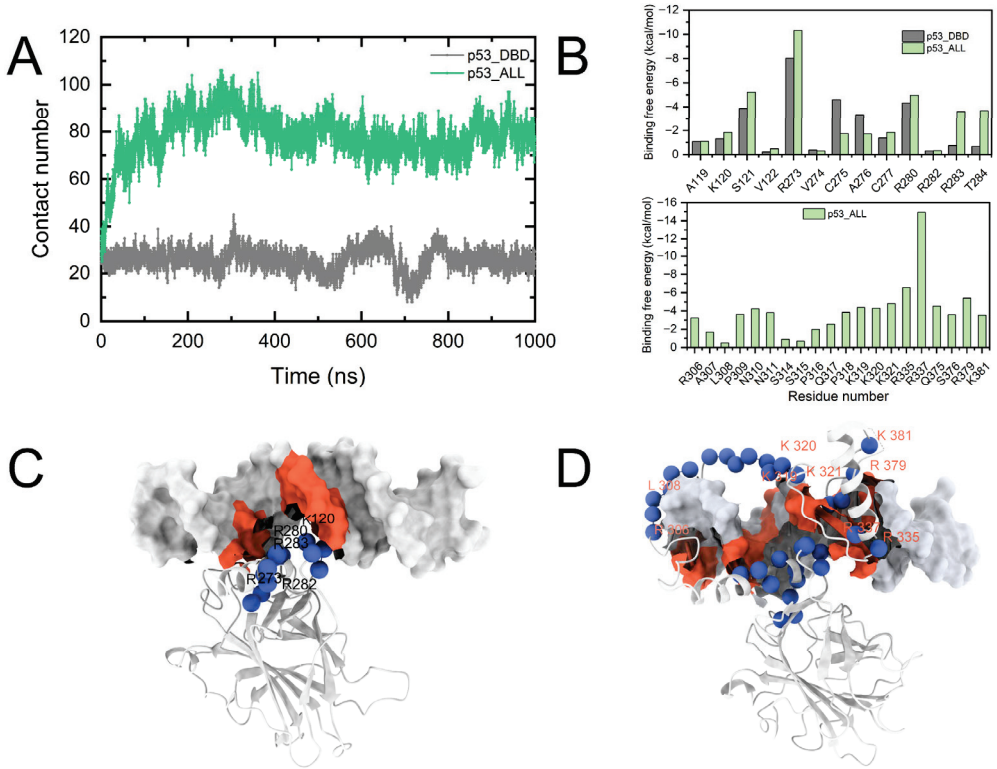


Figure 5. Molecular interactions between p53 and DNA. (**A**) Coordination number of residue–DNA contacts in the $p53_{DBD\text{-}DNA}$ and $p53_{ALL\text{-}DNA}$ systems. (**B**) Per-residue decomposition of binding free energy for the $p53_{DBD\text{-}DNA}$ and $p53_{ALL\text{-}DNA}$ interactions. (**C**) Structural view of the interaction interface between the DBD and DNA. (**D**) Structural view of the interaction interface between the C-terminal domain (CTD) and DNA.

We also calculated the binding free energy of both systems and performed an energy decomposition of the contributions from amino acid residues during the binding. The binding free energy represents the binding affinity between two molecules. The negative value of the binding free energy for receptors and ligands in the gas or aqueous phase indicates a favorable binding, while a positive value suggests an unfavorable interaction. We first calculated the binding free energy between the $p53_{DBD}$ and DNA. As shown in Table S1, the binding free energy between $p53_{DBD}$ and DNA is -41.3 kcal mol^{-1}. The main contribution of the binding energy between $p53_{DBD}$ and DNA is derived from the electrostatic interaction energy, which is favorable in the gas phase. In the case of the $p53_{ALL\text{-}DNA}$, the result shows that the binding free energy significantly increases to -161.1 kcal mol^{-1} (Table S2). To study the binding mechanism between the CTD and DNA, we performed per-residue energy decomposition of the binding free energy between the $p53_{ALL}$ ($p53_{DBD}$) and DNA. Figure 5B shows the significantly enhanced binding free energy contributions in the $p53_{ALL\text{-}DNA}$ compared to the $p53_{DBD\text{-}DNA}$. The result shows that the enhanced energy contributions are mainly derived from the CTD, primarily located in residues R306-K321, R335, and R379 of the CTD (Figure 5D). The simulation of $p53_{ALL\text{-}DNA}$ shows a more favorable contribution (p53 and DNA) to the $p53_{DBD\text{-}DNA}$. These results support the experimental proposition that the p53 CTD plays a substantial role in DNA binding [24]. As shown

https://doi.org/10.3390/ph19010124

in Figure 6A, the CTD contributes a more significant favorable energy (−54.32 kcal/mol), specific residues in the NTD provide an insignificant contribution (0.35 kcal/mol), consistent with the increased correlations observed in Figure 5. The p53 DBD specifically binds to DNA through the H2 helix and L1 loop, while the CTD further promotes the interaction between p53 and DNA.

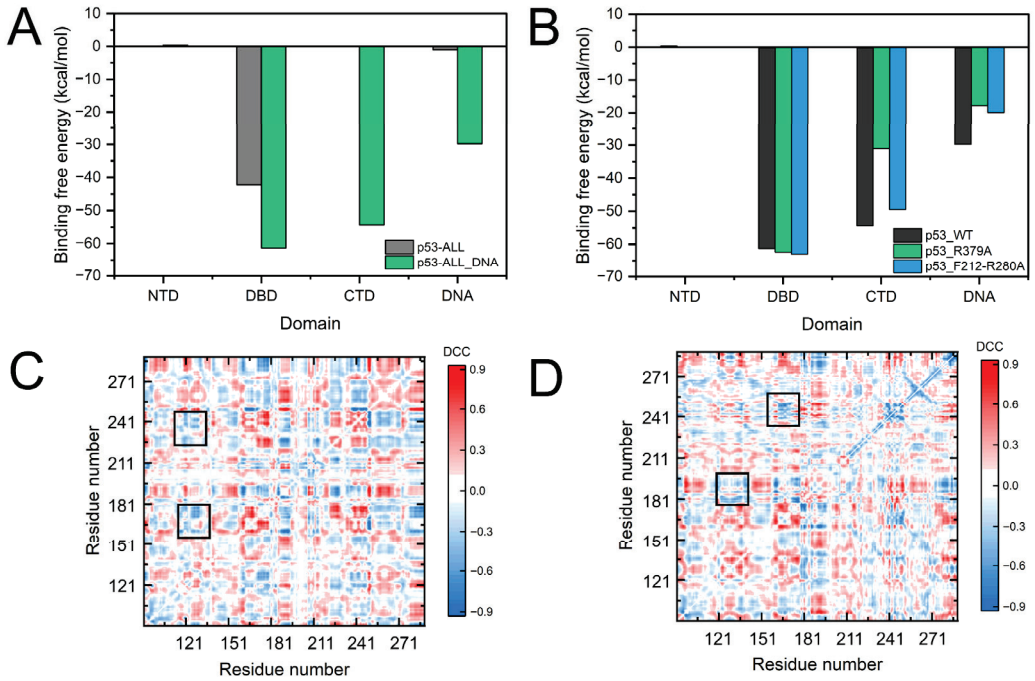


Figure 6. Binding free energy and DCC analysis of p53–DNA interactions in different structural contexts and mutants. (**A**) Domain-wise contributions to the total binding free energy between p53 and DNA. Values are shown for the full-length protein ($p53_{ALL}$) and its DNA-bound state ($p53_{ALL\text{-}DNA}$). Negative values indicate favorable binding contributions. (**B**) The plots compare wild-type (p53_WT) with two mutants: p53_R379A and p53_F212–R280A. (**C**,**D**) Difference maps of residue–residue dynamic correlation coefficients. (**C**) $p53_{ALL\text{-}DNA}$ (R379A) vs. $p53_{ALL\text{-}DNA}$. (**D**) $p53_{ALL\text{-}DNA}$ (F212A/R280A) vs. $p53_{ALL\text{-}DNA}$.

2.3. The CTD Induces Allosteric Regulation of DBD by Promoting DNA Binding

To understand the regulation mechanism of the conformational changes in DBD and DNA binding induced by the CTD, we also constructed the free energy landscape for the $p53_{ALL}$/$p53_{ALL\text{-}DNA}$. In Figure S9, the RMSD of the DBD increased to 3.6 Å, and the R_g decreased to 16.5 Å upon DNA binding. Simulations by Pradhan et al. suggested that the conformational state of the L1 loop may be related to the motion within the S6-S7 loop [26]. In our simulations, both the $p53_{DBD}$ and $p53_{ALL}$ simulations show the conformational change in the L1 loop upon DNA binding. We used two CVs to concurrently describe the conformational dynamics of the S6-S7 loop and L1 loop. The x-axis represents the distance between the CA atoms of residues R158 and F212, and the y-axis represents the distance between the CA atoms of K120 and R280. In the simulation of $p53_{DBD\text{-}DNA}$ (Figure 7A), the distance between the CA atoms of R280 and K120 is 5.70 Å after DNA binds to DBD, and the distance between the CA atoms of F212 and R158 is 12.96 Å. In the $p53_{ALL\text{-}DNA}$ simulation (Figure 7B), the distance between the CA atoms of R280 and K120 further decreases to 4.87 Å after CTD binds to DNA, indicating that the L1 and H2 move closer together. This conformation facilitates the simultaneous insertion of L1 and H2 into the DNA major groove. Meanwhile, the distance between F212 and R156 also shortens to 12.08 Å (Figure 7C,D), and the conformational change in the S6-S7 loop promotes allosteric

https://doi.org/10.3390/ph19010124

regulation within the DBD, which is consistent with previous simulation studies of the DBD region [29]. Interestingly, both K120 and R280 are located at the DNA-binding interface and carry positive charges. The decreased distance between these residues enhances DBD-DNA binding affinity through favorable electrostatic interactions and promotes the second transition of the DBD structure.

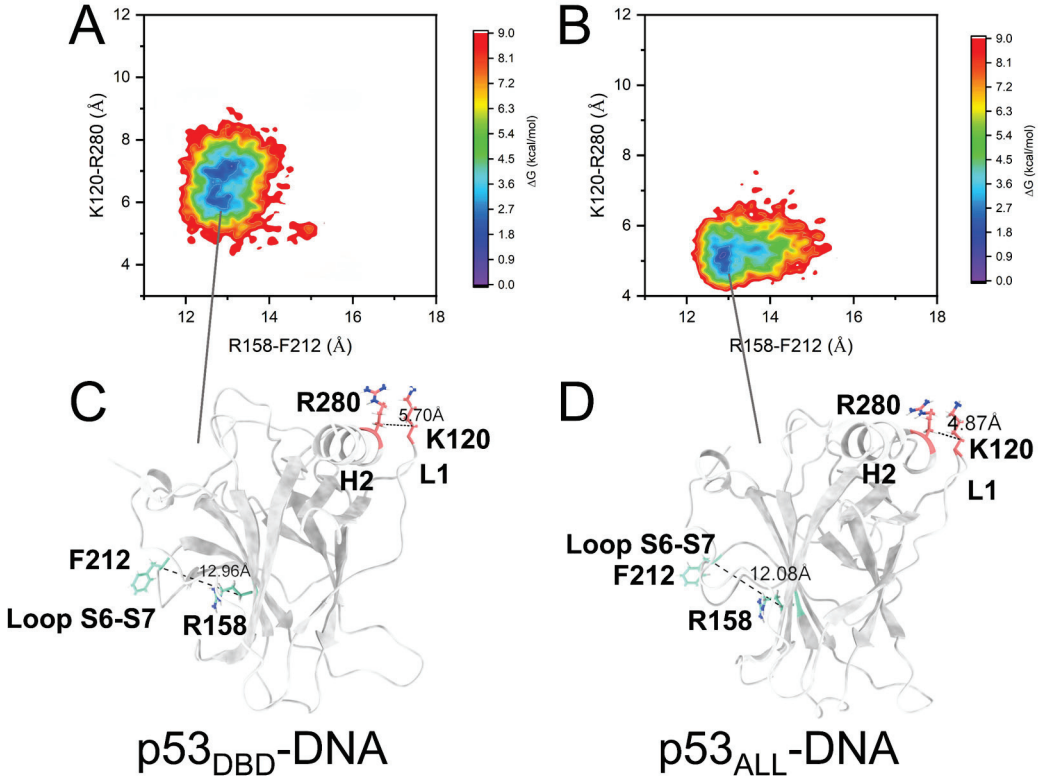


Figure 7. Conformational free-energy landscapes of the S6-S7 and L1 loops in p53 upon DNA binding. (**A**) Two-dimensional free-energy surface for $p53_{DBD\text{-}DNA}$, projected onto the Cα–Cα distances of R158–F212 (x-axis) and K120–R280 (y-axis). This landscape reflects conformational sampling of the S6-S7 and L1 loops. (**B**) Corresponding free-energy landscape for $p53_{ALL\text{-}DNA}$, described by the same distance variables. (**C**,**D**) Representative conformational motions of the S6-S7 and L1 loops extracted from the energy minima in (**C**) $p53_{DBD\text{-}DNA}$ and (**D**) $p53_{ALL\text{-}DNA}$.

The conformational states of the S6-S7 loop and L1 loop upon DNA binding induce the more compact structure of the p53 DBD, which is favorable for the internal network connections. The DCC analysis (Figure 4C) shows the correlation nodes within the DBD of the $p53_{ALL\text{-}DNA}$. The structural network of the DBD is mainly composed of F212 from the S6-S7 loop, E285 from the H2 helix, S166 from H1, and S127 from the L1 loop (Figure 4F). Therefore, the amino acids of S4 and S10 act as connecting units between the S6-S7 loop, H2 helix, and the L1 loop. These results indicate that the conformation of DBD binding to DNA is mediated by the conformational changes in the global network. The conformation shift (S6-S7 loop) is crucial for the recruitment of biological partners. The CTD further promotes the allosteric regulation of the DBD conformation and the subsequent protein–protein interactions. Moreover, the CTD-mediated stabilization of the DBD-DNA complex and the modulation of its conformational dynamics likely facilitate the subsequent recruitment of coactivators (such as MDM2/p300) and other transcription machinery, which are essential for p53′s transcriptional activity [21,24].

https://doi.org/10.3390/ph19010124

2.4. In Silico Mutagenesis Validates the Allosteric Network

To validate the impact of the allosteric network on the interaction between p53 and DNA binding, we calculated the DCC maps and binding free energies for these mutants. The results show a significant reduction in both long-range correlations of DBD (Figure 6C,D) and the favorable binding energy contribution from the mutated regions (Tables S3 and S4). The binding free energy contributions from the domains reveal that the R379A mutation significantly weakens the direct binding free energy contribution of the CTD to DNA (Figure 6B). In the R379A mutation simulation, the binding free energy contribution of the CTD region to DNA decreased from −54.32 kcal/mol to −31.00 kcal/mol. Similarly, in the F212A/R280A double-site mutation simulation, the binding free energy contribution of CTD also decreased to −49.50 kcal/mol. Furthermore, compared to the wild type, the DBD of both mutants, particularly R379A, shows significantly negative correlations. These results suggested the regulatory role of the CTD basic region and the DBD allosteric network in DNA binding. These results suggested the regulatory role of the CTD basic region and the DBD allosteric network in DNA binding.

We also evaluated the structural changes induced in the intrinsically disordered N- and C-terminal domains upon DNA binding. As shown in Figure S10A,B, DNA binding induces a measurable increase in helical content within both terminal domains. Specifically, in the CTD (residues 340–360), the average helicity increases in the $p53_{ALL\text{-}DNA}$ complex compared to the DNA-free state ($p53_{ALL}$). The NTD also shows localized stabilization, particularly in the region spanning residues 12–23. This indicates a shift towards a more ordered conformation in both IDRs upon interaction with DNA. Our site-directed mutagenesis simulations provide clear evidence that destabilizing the allosteric network impairs this ordering. As visualized in Figure S10D, the R379A mutation, targeting a critical CTD anchor point, causes a decrease in helicity (residues 365–370) and increases conformational fluctuations in the NTD (Figure S10C). The F212A/R280A double mutation within the core DBD disrupts the long-range coupling, leading to a substantial reduction in helical stability.

3. Discussion

The IDRs within the protein play a crucial regulatory role in key biological functions by acting as linker units in the binding process of protein complexes, contributing to structural flexibility. The p53 protein contains highly disordered N-terminal and C-terminal sequences, and the DNA-binding domain. The confirmation of DBD is critical for binding biological partners and subsequent transcriptional regulation. Previous studies mainly focus on the conformational and functional changes in the DBD upon DNA binding. The experiment revealed the different properties of p53 NTD and CTD in DNA binding [21–23]. However, there is still a lack of structural details about the regulation of the internal network within DBD and DNA binding.

In this paper, we investigated the regulation of the internal network of the DBD by the CTD upon DNA binding. We employed microsecond-scale MD simulations to establish the correlation between the DBD conformation and the p53 CTD. Some of our results are consistent with the previous enhanced sampling simulations. In addition, the simulations of $p53_{ALL\text{-}DNA}$ elucidate the regulatory roles of the CTD upon DNA binding. Upon DNA binding to the DBD (Figure 8), the internal correlation of the DBD begins to increase, revealing a preliminary allosteric network. Subsequently, as the CTD region interacts with DNA, the binding between the DBD and DNA is further stabilized, accompanied by an expansion of the regions involved in the internal allosteric network. Therefore, DNA binding serves as the first critical step in the allosteric regulation within p53, while CTD-DNA interaction constitutes the second step in modulating the internal allosteric

https://doi.org/10.3390/ph19010124

network. The DBD adopts a more compact conformation upon DNA binding, the S6–S7 loop transits to a closed state, and the L1 loop shifts closer to the H2 helix to facilitate DNA binding. The CTD directly interacts with DNA through electrostatic interactions, expanding the interface of p53 binding with DNA. The CTD modulates the DNA-binding affinity and the structural conformation through global allosteric regulation. Thus, the CTD-induced stabilization of the DBD conformation and the identified allosteric network likely pre-organize or create interfaces that facilitate the recruitment of transcriptional co-activators (like MDM2/p300) [30].

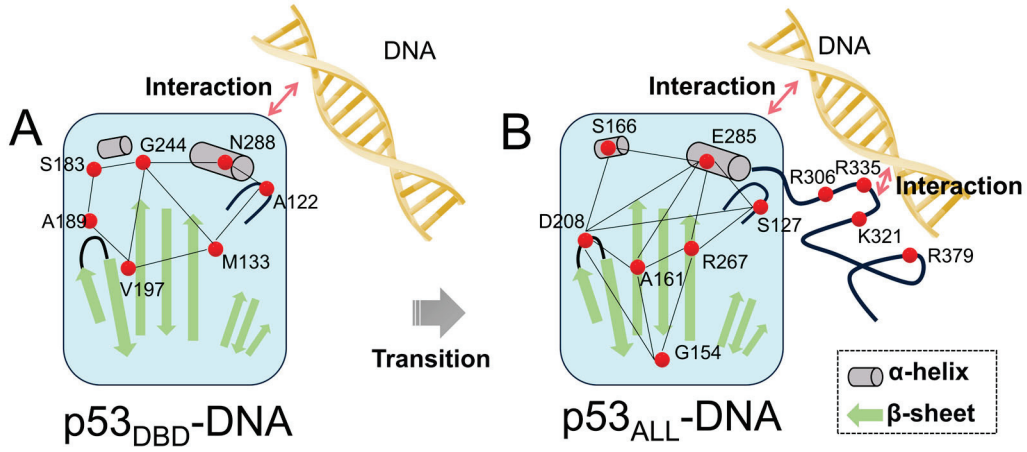


Figure 8. Proposed regulatory model of DNA and CTD binding on p53 DBD stability. (**A**) When DNA binds only to the DBD (CTD unbound), a transient interaction network forms within the DBD, resulting in partial complex stabilization. (**B**) Cooperative binding of DNA to both the DBD and CTD induces extensive strengthening of the DBD internal interaction network, leading to full stabilization of the p53–DNA complex.

Additionally, we discussed the robustness of the predicted p53 structure in the regulatory mechanism of the CTD. We employed Rosetta and Swiss-Model to predict the full-length protein structure of p53 and performed 500 ns MD simulations for each. As shown in Figure S11A, the positions of the CTD region in the initial structural models of p53 predicted by different methods are different. However, the CTD region ultimately binds to DNA in all cases during dynamics simulations (Figure S11B). This is due to the abundance of basic amino acids in the CTD region, and the electrostatic interactions facilitate its binding near DNA. Moreover, this interaction becomes further stabilized as the dynamics trajectory progresses (Figure S11C,D).

The methodology of this work also has some limitations in calculating binding free energy. MM-GBSA provides relative trends rather than absolute binding affinities. The large energy contribution of the CTD should be interpreted qualitatively, highlighting its significant stabilizing role, not as a precise thermodynamic measurement, particularly in handling solvent and ion effects for charged systems like DNA-protein complexes. Therefore, future therapeutic strategies could potentially regulate the transition of the p53 protein between active and inactive states through rational drug design. By inhibiting the allosteric network and DNA-binding activity of p53, it may be possible to suppress disease progression.

4. Materials and Methods

We integrated dynamics and thermodynamic analyses to explore the allosteric mechanism of p53 CTD regulating DBD structure and DNA binding. The workflow of the

https://doi.org/10.3390/ph19010124

unbiased molecular dynamics (MD) method is illustrated in Figure 9. First, we employed multiple structure prediction software to perform structural modeling of the p53 protein, facilitating subsequent assessment of the robustness of the structural mechanisms. Second, unbiased MD simulations were performed for the four p53 complex systems. Third, residue contact maps and atomic fluctuations were analyzed to characterize the binding interactions between the p53 DBD/CTD and DNA. Fourth, dynamic cross-correlation network analysis was applied to examine changes in the internal interaction network of DBD upon DNA and CTD binding. Fifth, the free energy landscape (FEL) analysis was performed to identify key conformations associated with CTD-mediated regulation of the DBD network, followed by cluster analysis. Finally, the allosteric regulatory mechanism of DBD induced by CTD binding was summarized.

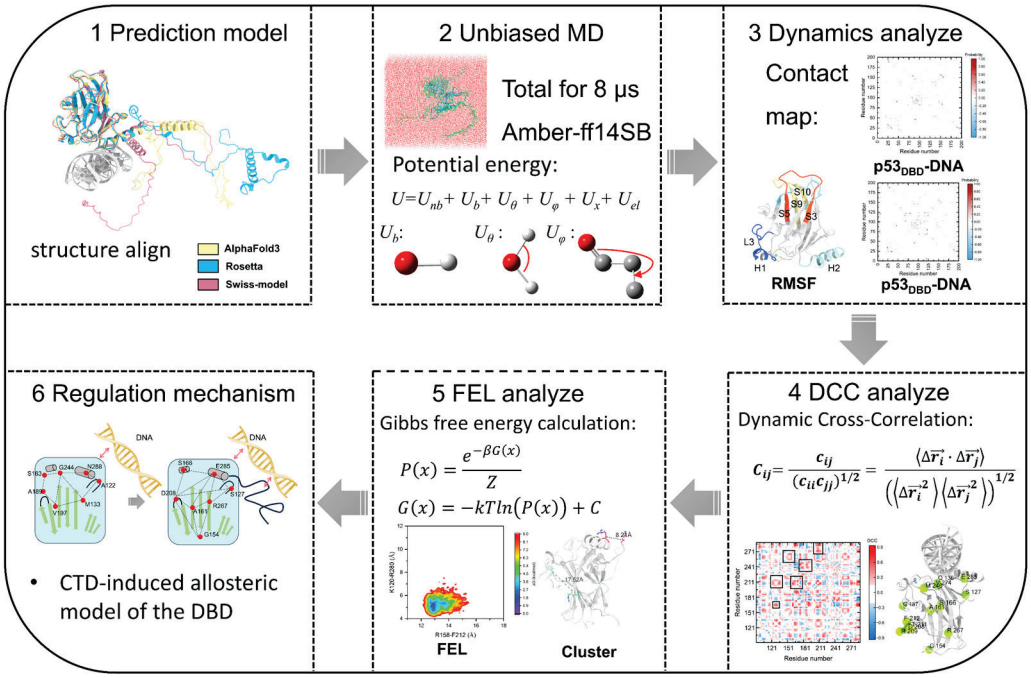


Figure 9. Workflow for unbiased molecular dynamics simulations and post-analysis of conformational dynamics and binding energetics.

4.1. Simulation Protocols

The structure of the DBD ($p53_{DBD}$: 92–290) was obtained from the experimental structure (PDB ID: 2XWR) [31]. The complex structure of the DBD binding to DNA ($p53_{DBD\text{-}DNA}$: 92–290, in complex with DNA) was derived from the X-ray crystal structure (PDB ID: 3KMD) [32]. We utilized AlphaFold3 to predict and construct the full-length p53 protein structure (Figure S1) [33]. Subsequently, we employed Chimera-1.15 to build the zinc finger for the p53 protein ($p53_{ALL}$: 1–393) [34]. Finally, we referred to the experimental structure (PDB ID: 3KMD) and used Chimera-1.15 to construct the complex structure of the $p53_{ALL}$ protein bound to DNA ($p53_{ALL\text{-}DNA}$). At this point, we have constructed four molecular dynamics simulation systems (Table 1) to investigate the conformational changes in the DBD upon DNA binding and the effect of the CTD on the binding interaction between the DBD and DNA. The systems were solvated in a dodecahedral box using the TIP3P water model and neutralized with Na^+ and Cl^- ions to balance the charges within the system [35]. The periodic boundary conditions with a minimum distance were set as 12.5 Å between protein and box edges. We used the AMBER/ff14SB force field, including

https://doi.org/10.3390/ph19010124

parameters for proteins and DNA [36]. The Zinc AMBER force field (ZAFF) was adopted for zinc fingers [37].

Table 1. The four molecular dynamics simulation systems.

Systems	States	Residues	Time
$p53_{DBD}$	Protein	92–290	1 μs × 2
$p53_{DBD\text{-}DNA}$	Protein-DNA	92–290	1 μs × 2
$p53_{ALL}$	Protein	1–393	1 μs × 2
$p53_{ALL\text{-}DNA}$	Protein-DNA	1–393	1 μs × 2

We performed two independent, unbiased molecular dynamics (MD) simulations for each system, with each simulation generating 1 μs of trajectory, resulting in a total of 8 μs of MD simulation data using identical parameters [38]. The key parameter settings are detailed below. The systems were first minimized by using the steepest descent method until the maximum force (1000.0 kJ/mol/nm) or the minimization step reached 20,000 steps. Then, the system was heated to 300 K in two steps with the Nośe-Hoover thermostat method at 100 ps in the NVT ensemble [39]. No restraints were applied during the energy minimization and heating phases of the MD simulations. Next, another 100 ps MD equilibration in the NPT ensemble was performed at the pressure of 1 atm, which was coupled by the Parrinello–Rahman barostat method [40]. The integrating step was 2 fs in all the simulations. The Particle Mesh Ewald method (PME) was used to calculate the long-range electrostatic interactions with a van der Waals cut-off of 10 Å [41]. The hydrogen bonds are constrained by the SHAKE algorithm [42].

To validate the allosteric network in p53, we established additional simulations for two key mutant systems suggested by the network analysis: one with a mutation in the CTD (R379A) and another with a double mutation affecting a DBD interface residue (F212A/R280A).

4.2. Free Energy Analyses

4.2.1. Construction of the Free Energy Landscape

The free energy landscapes (FELs) represent the potential of the mean force of the systems. We strategically employed two distinct sets of collective variables (CVs) to evaluate the dynamic fluctuations within the p53 DBD. The initial set encompasses the root mean square deviation (RMSD) and the radius of gyration (R_g) of the p53 DBD. These CVs can reflect the correlation between the conformational changes in p53 DBD and the stability of the internal network. The CVs are defined by the distances between the CA atoms of residues R158-F212, R156-D208, and K120-R280 within the p53 DBD. The distances between the CA atoms of amino acid residues on the S6-S7 loop and the L1 loop describe the conformational changes in the DBD allosteric network under the regulation of CTD. The calculation of free energy employed the GROMACS tool, which adeptly integrates histogram analysis with Boltzmann statistics to elucidate the free energy landscape of molecular systems [43]. The FELs are derived from the probabilities of the system being in specific states, which are calculated from the histogram data. The relationship is given by the Boltzmann distribution:

$$P(x) = \frac{e^{-\beta G(x)}}{Z} \quad (1)$$

where $P(x)$ is the probability of the system being in state x, $\beta = 1/kT$ (with k being the Boltzmann constant and T the temperature), $G(x)$ is the free energy at state x, and Z is the partition function.

https://doi.org/10.3390/ph19010124

The free energy at each point in the multi-dimensional space is computed using the probabilities derived from the histograms:

$$G(x) = -kTln(P(x)) + C \tag{2}$$

where C is a constant that can be adjusted based on reference states.

4.2.2. Calculation of the Binding Free Energy for Protein-DNA Interactions

The MM-PBSA (Molecular Mechanics Poisson Boltzmann Surface Area) or MM-GBSA (Molecular Mechanics Generalized Born Surface Area) methods in AMBER20 are widely used approaches for estimating the binding free energy of biomolecular complexes, such as protein-ligand interactions [44]. These methods combine molecular mechanics with continuum solvation models to provide a comprehensive evaluation of the free energy changes associated with binding. In this work, the binding free energy was calculated with the MM-GBSA method by the parallel version MMPBSA.py in AMBER20 [45]. The binding free energy (ΔG_{bind}) can be expressed as the difference in free energy between the complexed and uncomplexed states [46]:

$$\Delta G_{bind} = G_{complex} - \left(G_{receptor} - G_{ligand}\right) \tag{3}$$

where $G_{complex}$ is the free energy of the receptor-ligand complex, $G_{receptor}$ is the free energy of the receptor in isolation, and G_{ligand} is the free energy of the ligand in isolation.

4.3. Dynamic Cross-Correlation Analysis

4.3.1. Covariance Calculation

The dynamic cross-correlation analyses were calculated with GROMACS-2022 tools and Python-3.9 scripts [43]. First, we utilized the "gmx covar" tool within the GROMACS-2022 software to construct the dynamic cross-correlation matrices (DCCM) for each system. Subsequently, we employed Python-3.9 scripts to calculate the dynamic cross-correlation coefficients (DCCs) between the CA atoms of each residue. The motion pattern correlation between residues can be determined by calculating the covariance between pairwise residues. The formulas are as follows:

$$c(i,j) = \langle \Delta R_i \cdot \Delta R_j \rangle \tag{4}$$

where

$$\Delta R_i = R_i - \langle R_i \rangle \tag{5}$$

Here, i and j represent the residue indices of the protein amino acids, and $c(i, j)$ is the covariance between residue i and residue j. The angle brackets denote an ensemble average, which means the average over a series of models, including multiple frames from a simulation trajectory or multiple structures obtained from NMR, etc. ΔR_i represents the positional deviation based on the ensemble average position.

4.3.2. Correlation Coefficient Calculation

The dynamic correlation coefficients are the covariance standardized after removing the influence of the magnitude of changes in the two variables, reflecting the pure correlation between the two variables. These coefficients are used to construct the DCCM by converting the covariance between residue pairs into correlation indices and organizing

https://doi.org/10.3390/ph19010124

them into a matrix according to residue numbering. The DCCs ($C(i,j)$) between two residues could be calculated using the following equation [47]:

$$C(i,j) = \frac{c(i,j)}{[c(i,i) \cdot c(j,j)]^{1/2}} \quad (6)$$

4.4. Analyses of Contact Maps

The residue contact map was calculated based on distances between all heavy atoms in both the protein and DNA, with a cutoff distance of 4.5 Å set according to references [8,48]. In the complex structure, any two heavy atoms that are less than 4.5 Å apart were defined as a contact between the corresponding residues. The residue contact probability was then calculated as the ratio of frames containing the contacts to the total number of trajectory frames. For Root Mean Square Fluctuation (RMSF) calculations, we first aligned all structures to the initial reference structure by superimposing the DBD (residues 92–289). The RMSF values were then computed for the heavy atoms of each residue.

5. Conclusions

In summary, this work elucidates a hierarchical model of p53 regulation wherein DNA binding initiates allostery and CTD interaction amplifies and refines it. Future studies exploring the role of the N-terminal domain and the effects of post-translational modifications on this network will further illuminate the full spectrum of p53 regulation. Ultimately, understanding the dynamic interplay among disordered regions and structured domains in p53 opens pathways for allosteric drug discovery and offers a paradigm for studying multi-domain proteins with intrinsically disordered regions.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ph19010124/s1, Figure S1: Domain structure of p53 and the internal interaction of p53 domains; Figure S2: Changes in the interaction interfaces of p53DBD and p53ALL upon DNA binding; Figure S3: Dynamic changes of internal interactions within the p53 DBD following DNA binding; Figure S4: The free energy landscape of p53DBD upon DNA binding; Figure S5: The free energy landscape change of p53 S6-S7 loop and L1 upon DNA binding; Figure S6: Conformational network changes of p53DBD and p53ALL before and after DNA binding; Figure S7: The correlation difference for the p53DBD upon DNA binding; Figure S8: The correlation difference for the p53ALL upon DNA binding; Figure S9: The free energy landscape of p53 p53ALL upon DNA binding; Figure S10: Changes in the secondary structure of the p53 N-terminal and C-terminal domains following DNA binding; Figure S11: Validation of the robustness of different p53 prediction models; Table S1: The binding free energy of p53 DBD and DNA; Table S2: The binding free energy of p53 and DNA; Table S3: The binding free energy of p53_R379A and DNA; Table S4: The binding free energy of p53_F212A-R280A and DNA.

Author Contributions: MD simulations and formal analysis, S.N. and C.Z.; Funding acquisition, S.N. and J.Z.; Investigation, H.W. and Y.X.; Methodology, S.N.; Writing—original draft, S.N.; Writing—review and editing, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Key Research Projects Plan of Colleges and Universities in Henan Province 26A180009 (SN) and 25A310006 (JZ and SN), the National Natural Science Foundation of China 12204154 (HW).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

https://doi.org/10.3390/ph19010124

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no competing financial interests.

Abbreviations

IDRs	Intrinsically disordered regions
CTD	C-terminal domain
DBD	DNA-binding domain
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
MM-GBSA	Molecular mechanics generalized Born surface area
DCCs	Dynamic cross-correlation coefficients
FELs	Free energy landscapes

References

1. Wright, P.E.; Dyson, H.J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321–331. [CrossRef] [PubMed]
2. Holehouse, A.S.; Kragelund, B.B. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat. Rev. Mol. Cell Biol.* **2024**, *25*, 187–211. [CrossRef]
3. Dyson, H.J.; Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208. [CrossRef] [PubMed]
4. Iakoucheva, L.M.; Radivojac, P.; Brown, C.J.; O'Connor, T.R.; Sikes, J.G.; Obradovic, Z.; Dunker, A.K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037–1049. [CrossRef]
5. Dyson, H.J.; Wright, P.E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60. [CrossRef]
6. Oates, M.E.; Romero, P.; Ishida, T.; Ghalwash, M.; Mizianty, M.J.; Xue, B.; Dosztányi, Z.; Uversky, V.N.; Obradovic, Z.; Kurgan, L.; et al. D2P2: Database of disordered protein predictions. *Nucleic Acids Res.* **2013**, *41*, D508–D516. [CrossRef]
7. Galea, C.A.; Wang, Y.; Sivakolundu, S.G.; Kriwacki, R.W. Regulation of Cell Division by Intrinsically Unstructured Proteins: Intrinsic Flexibility, Modularity, and Signaling Conduits. *Biochemistry* **2008**, *47*, 7598–7609. [CrossRef]
8. Liu, N.; Guo, Y.; Ning, S.; Duan, M. Phosphorylation regulates the binding of intrinsically disordered proteins via a flexible conformation selection mechanism. *Commun. Chem.* **2020**, *3*, 123. [CrossRef]
9. Tran, H.T.; Mao, A.; Pappu, R.V. Role of Backbone−Solvent Interactions in Determining Conformational Equilibria of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2008**, *130*, 7380–7392. [CrossRef]
10. Ning, S.; Chao, H.-J.; Li, S.; Zhou, R.; Zou, L.; Zhang, X.; Liu, J.; Yan, D.; Duan, M. The auto-inhibition mechanism of transcription factor Ets-1 induced by phosphorylation on the intrinsically disordered region. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 1132–1141. [CrossRef] [PubMed]
11. Vousden, K.H.; Prives, C. Blinded by the Light: The Growing Complexity of p53. *Cell* **2009**, *137*, 413–431. [CrossRef]
12. Vousden, K.H.; Lane, D.P. p53 in health and disease. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 275–283. [CrossRef]
13. He, F.; Borcherds, W.; Song, T.; Wei, X.; Das, M.; Chen, L.; Daughdrill, G.W.; Chen, J. Interaction between p53 N terminus and core domain regulates specific and nonspecific DNA binding. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 8859–8868. [CrossRef]
14. Walker, K.K.; Levine, A.J. Identification of a novel p53 functional domain that is necessary for efficient growth suppression. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 15335–15340. [CrossRef]
15. Brady, C.A.; Jiang, D.; Mello, S.S.; Johnson, T.M.; Jarvis, L.A.; Kozak, M.M.; Broz, D.K.; Basak, S.; Park, E.J.; McLaughlin, M.E.; et al. Distinct p53 Transcriptional Programs Dictate Acute DNA-Damage Responses and Tumor Suppression. *Cell* **2011**, *145*, 571–583. [CrossRef]
16. Fraser, J.A.; Vojtesek, B.; Hupp, T.R. A Novel p53 Phosphorylation Site within the MDM2 Ubiquitination Signal: I. Phosphorylation at SER269 in vivo is linked to inactivation of p53 function. *J. Biol. Chem.* **2010**, *285*, 37762–37772. [CrossRef]
17. Wei, X.; Wu, S.; Song, T.; Chen, L.; Gao, M.; Borcherds, W.; Daughdrill, G.W.; Chen, J. Secondary interaction between MDMX and p53 core domain inhibits p53 DNA binding. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E2558–E2563. [CrossRef] [PubMed]
18. McKinney, K.; Mattia, M.; Gottifredi, V.; Prives, C. p53 Linear Diffusion along DNA Requires Its C Terminus. *Mol. Cell* **2004**, *16*, 413–424. [CrossRef] [PubMed]

https://doi.org/10.3390/ph19010124

19. Joerger, A.C.; Fersht, A.R. Structural Biology of the Tumor Suppressor p53 and Cancer-Associated Mutants. In *Advances in Cancer Research*; Academic Press: New York, NY, USA, 2007; Volume 97, pp. 1–23.
20. Böttger, V.; Böttger, A.; Garcia-Echeverria, C.; Ramos, Y.F.M.; van der Eb, A.J.; Jochemsen, A.G.; Lane, D.P. Comparative study of the p53-mdm2 and p53-MDMX interfaces. *Oncogene* **1999**, *18*, 189–199. [CrossRef]
21. Ferreon, J.C.; Lee, C.W.; Arai, M.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E. Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6591–6596. [CrossRef] [PubMed]
22. Weinberg, R.L.; Freund, S.M.V.; Veprintsev, D.B.; Bycroft, M.; Fersht, A.R. Regulation of DNA Binding of p53 by its C-terminal Domain. *J. Mol. Biol.* **2004**, *342*, 801–811. [CrossRef] [PubMed]
23. Gu, W.; Roeder, R.G. Activation of p53 Sequence-Specific DNA Binding by Acetylation of the p53 C-Terminal Domain. *Cell* **1997**, *90*, 595–606. [CrossRef]
24. Retzlaff, M.; Rohrberg, J.; Küpper, N.J.; Lagleder, S.; Bepperling, A.; Manzenrieder, F.; Peschek, J.; Kessler, H.; Buchner, J. The Regulatory Domain Stabilizes the p53 Tetramer by Intersubunit Contacts with the DNA Binding Domain. *J. Mol. Biol.* **2013**, *425*, 144–155. [CrossRef] [PubMed]
25. Laptenko, O.; Shiff, I.; Freed-Pastor, W.; Zupnick, A.; Mattia, M.; Freulich, E.; Shamir, I.; Kadouri, N.; Kahan, T.; Manfredi, J.; et al. The p53 C Terminus Controls Site-Specific DNA Binding and Promotes Structural Changes within the Central DNA Binding Domain. *Mol. Cell* **2015**, *57*, 1034–1046. [CrossRef]
26. Pradhan, M.R.; Siau, J.W.; Kannan, S.; Nguyen, M.N.; Ouaray, Z.; Kwoh, C.K.; Lane, D.P.; Ghadessy, F.; Verma, C.S. Simulations of mutant p53 DNA binding domains reveal a novel druggable pocket. *Nucleic Acids Res.* **2019**, *47*, 1637–1652. [CrossRef]
27. Bhattacharjee, S.; Sengupta, J. Hidden electrostatic energy contributions define dynamic allosteric communications within p53 during molecular recognition. *Biophys. J.* **2021**, *120*, 4512–4524. [CrossRef]
28. Degtjarik, O.; Golovenko, D.; Diskin-Posner, Y.; Abrahmsén, L.; Rozenberg, H.; Shakked, Z. Structural basis of reactivation of oncogenic p53 mutants by a small molecule: Methylene quinuclidinone (MQ). *Nat. Commun.* **2021**, *12*, 7057. [CrossRef]
29. Lambrughi, M.; De Gioia, L.; Gervasio, F.L.; Lindorff-Larsen, K.; Nussinov, R.; Urani, C.; Bruschi, M.; Papaleo, E. DNA-binding protects p53 from interactions with cofactors involved in transcription-independent functions. *Nucleic Acids Res.* **2016**, *44*, 9096–9109. [CrossRef]
30. Poyurovsky, M.V.; Katz, C.; Laptenko, O.; Beckerman, R.; Lokshin, M.; Ahn, J.; Byeon, I.-J.L.; Gabizon, R.; Mattia, M.; Zupnick, A.; et al. The C terminus of p53 binds the N-terminal domain of MDM2. *Nat. Struct. Mol. Biol.* **2010**, *17*, 982–989. [CrossRef]
31. Natan, E.; Baloglu, C.; Pagel, K.; Freund, S.M.V.; Morgner, N.; Robinson, C.V.; Fersht, A.R.; Joerger, A.C. Interaction of the p53 DNA-Binding Domain with Its N-Terminal Extension Modulates the Stability of the p53 Tetramer. *J. Mol. Biol.* **2011**, *409*, 358–368. [CrossRef] [PubMed]
32. Chen, Y.; Dey, R.; Chen, L. Crystal Structure of the p53 Core Domain Bound to a Full Consensus Site as a Self-Assembled Tetramer. *Structure* **2010**, *18*, 246–256. [CrossRef]
33. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. [CrossRef] [PubMed]
34. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef] [PubMed]
35. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]
36. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [CrossRef]
37. Peters, M.B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M.N.; Merz, K.M., Jr. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6*, 2935–2947. [CrossRef]
38. Case, D.A.; Cheatham, T.E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688. [CrossRef]
39. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519. [CrossRef]
40. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190. [CrossRef]
41. 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]
42. Ryckaert, J.P.; Ciccotti, G.; Berendsen, H. 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]

 https://doi.org/10.3390/ph19010124

43. 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]
44. Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897. [CrossRef] [PubMed]
45. Miller, B.R.; McGee, T.D.; Swails, J.M.; Homeyer, N.; Gohlke, H.; Roitberg, A.E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321. [CrossRef] [PubMed]
46. Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson−Boltzmann Surface Area Method. *Mol. Inform.* **2012**, *31*, 114–122. [CrossRef]
47. Sethi, A.; Eargle, J.; Black, A.A.; Luthey-Schulten, Z. Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6620–6625. [CrossRef]
48. Ning, S.; Zeng, C.; Zeng, C.; Zhao, Y. The TAR binding dynamics and its implication in Tat degradation mechanism. *Biophys. J.* **2021**, *120*, 5158–5168. [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.

https://doi.org/10.3390/ph19010124

pharmaceuticals

MDPI

Article

Charged Thienobenzo-1,2,3-Triazoles as Especially Potent Non-Selective Cholinesterase Inhibitors: Design, Anti-Inflammatory Activity, and Computational Study

Antonija Jelčić [1,†], Anamarija Raspudić [2,†], Danijela Barić [3], Ana Ratković [1], Ivana Šagud [4], Paula Pongrac [5], Dora Štefok [5], Martina Bosnar [5], Sunčica Roca [6], Zlata Lasić [7], Ilijana Odak [2,*] and Irena Škorić [1,*]

[1] Department of Organic Chemistry, Faculty of Chemical Engineering and Technology, University of Zagreb, Trg Marka Marulića 19, HR-10 000 Zagreb, Croatia; ajelcic@fkit.unizg.hr (A.J.); ana.ratkovic30@gmail.com (A.R.)
[2] Department of Chemistry, Faculty of Science and Education, University of Mostar, Matice Hrvatske bb, 88 000 Mostar, Bosnia and Herzegovina; anamarija.raspudic@fpmoz.sum.ba
[3] Group for Computational Life Sciences, Division of Physical Chemistry, Ruđer Bošković Institute, Bijenička Cesta 54, HR-10 000 Zagreb, Croatia; dbaric@irb.hr
[4] Croatian Agency for Medicinal Products and Medical Devices, Ksaverska Cesta 4, HR-10 000 Zagreb, Croatia; ivana.sagud@halmed.hr
[5] Pharmacology In Vitro, Selvita Ltd., Prilaz Baruna Filipovića 29, HR-10 000 Zagreb, Croatia; paula.pongrac@selvita.com (P.P.); dora.stefok@selvita.com (D.Š.); martina.bosnar@selvita.com (M.B.)
[6] NMR Center, Ruđer Bošković Institute, Bijenička Cesta 54, HR-10 000 Zagreb, Croatia; sroca@irb.hr
[7] TEVA Global R&D, E&L R&D, Pliva Hrvatska d.o.o., Prilaz Baruna Filipovića 25, HR-10 000 Zagreb, Croatia; zlata.lasic01@pliva.com
* Correspondence: ilijana.odak@fpmoz.sum.ba (I.O.); iskoric@fkit.unizg.hr (I.Š.)
† These authors contributed equally to this work.

Abstract: Background/Objectives: This research reports the synthesis and evaluation of novel charged thienobenzo-triazoles as non-selective cholinesterase inhibitors (AChEs and BChEs), their anti-inflammatory properties, and a computational study. **Methods**: Fifteen derivatives were created through photochemical cyclization and quaternization of the triazole core. The compounds were tested for AChE and BChE inhibition. They showed greater potency and selectivity toward BChE. **Results**: The most potent compound, derivative 14, inhibited BChE with an IC_{50} of 98 nM, while derivative **9** also displayed significant anti-inflammatory activity by inhibiting LPS-induced TNF-α production (IC_{50} = 0.66 μM). Molecular docking revealed that triazolinium salts form key π-π and electrostatic interactions within enzyme active sites. In silico predictions indicated favorable ADME-Tox properties for compounds **9** and **11**, including low mutagenicity and moderate CNS permeability. **Conclusions**: These findings highlight the potential of new charged triazolinium salts as peripherally selective cholinesterase inhibitors with additional anti-inflammatory potential.

Keywords: thienobenzo-1,2,3-triazoles; cholinesterase; BChE inhibitors; anti-inflammatory activity; molecular docking; ADME-Tox prediction

1. Introduction

Cholinesterases are essential enzymes that hydrolyze acetylcholine (ACh) and butyrylcholine (BCh), ending cholinergic signaling. The main types are acetylcholinesterase (AChE), mainly located in synaptic clefts, and butyrylcholinesterase (BChE), which is more diffused in plasma and non-neuronal tissues [1]. Pharmacological blocking of these enzymes has therapeutic uses, especially in treating Alzheimer's disease (AD) and acute

organophosphate poisoning [2]. Non-selective cholinesterase inhibitors (ChEIs) affect both AChE and BChE, making them a flexible group of compounds with increasing clinical and research interest. A growing area of focus is the impact of molecular charge, particularly the benefits of positively charged (cationic) molecules.

ChEIs work by preventing the breakdown of ACh, thereby boosting cholinergic signaling. This approach is used in treating symptoms of neurodegenerative diseases like AD, where cholinergic deficits are common [3,4]. Non-selective inhibitors, which target both AChE and BChE, may offer broader benefits. In the later stages of AD, BChE activity rises and may help compensate for the reduced activity of AChE, underscoring the need to target both enzymes [5,6]. Approved ChEIs such as donepezil, galantamine, and rivastigmine have shown effectiveness in the brain, while peripheral inhibitors like neostigmine are important in anesthesia and neuromuscular disorders [7,8].

Non-selective cholinesterase inhibitors include both traditional compounds like physostigmine and neostigmine, as well as newer experimental agents that have reversible and irreversible qualities. Many of these molecules are designed with quaternary ammonium groups, which give them a permanent positive charge. This positive charge makes the molecules hydrophilic and greatly limits their ability to cross the blood–brain barrier (BBB), leading to mainly peripheral effects [9]. Charged ChEIs, especially those with quaternary ammonium groups, provide several pharmacological and safety benefits in clinical use. Because of their polar nature, these molecules poorly penetrate the BBB, which is beneficial in clinical situations where avoiding central nervous system (CNS) side effects is important [10,11].

Furthermore, the limited CNS distribution of charged molecules increases their suitability for peripheral therapeutic targets. This is especially advantageous in conditions such as myasthenia gravis or in reversing neuromuscular blockade after surgery, where a peripheral mode of action is preferred [8,12]. Their restricted distribution in the CNS also helps reduce the risk of neurotoxicity and convulsions, making them safer for use in acute or emergency situations [13,14]. Additionally, their hydrophilic nature results in predictable pharmacokinetics, with distribution mainly confined to extracellular fluids, which allows for consistent plasma levels and simpler dosing adjustments [15]. Despite these benefits, charged ChEIs have some limitations. Their inability to cross the BBB renders them ineffective for treating CNS-related disorders like AD unless delivered through alternative methods or chemically modified to improve CNS penetration [16]. Several charged non-selective ChEIs are currently used in clinical practice [17]. These drugs demonstrate how charged structures can offer targeted therapeutic effects while minimizing central toxicity.

Recent advances in medicinal chemistry aim to overcome the limitations of charged molecules while retaining their advantages. One approach involves designing hybrid or prodrug forms that temporarily mask the charge to facilitate BBB penetration, with activation occurring at the target site [16]. Research also continues into dual-action inhibitors that combine ChE inhibition with anti-oxidant or anti-amyloid properties, which could provide multifunctional benefits in neurodegenerative diseases [5,11]. Efforts are also ongoing to develop tissue-specific delivery systems and nanocarriers that allow controlled distribution of polar ChEIs [18]. Non-selective cholinesterase inhibitors with charged structures offer considerable therapeutic advantages, particularly in the peripheral nervous system, where CNS penetration is undesirable. Their hydrophilic and cationic nature allows for targeted, predictable, and relatively safe pharmacological profiles. However, their limited CNS access remains a significant obstacle in treating brain-related disorders. Ongoing research into prodrug strategies and hybrid molecules holds promise for extending the utility of charged ChEIs in broader clinical contexts.

In our previous research [19], pairs of uncharged thienobenzo-triazoles and their corresponding charged salts were prepared to examine the role of the positive charge on the nitrogen of the triazole ring in interactions within the active site of cholinesterase enzymes. The goal was also to compare the selectivity of 1,2,3-triazolinium salts with that of their uncharged analogs. Uncharged thienobenzo-triazoles I (Figure 1) favored inhibition of BChE over AChE. Conversion to the salt form (structure II, Figure 1) significantly increased the inhibition of AChE (the most active IC_{50} 4.4 μM) and also, to a reasonable degree, that of BChE (the most active IC_{50} 0.47 μM). It was concluded that the key structural feature for thienobenzo-triazoles is the presence or absence of charge. Triazolinium salts [19] were shown to be dual inhibitors, with their potency depending on the substituents on the thiophene and triazole heterocycles.

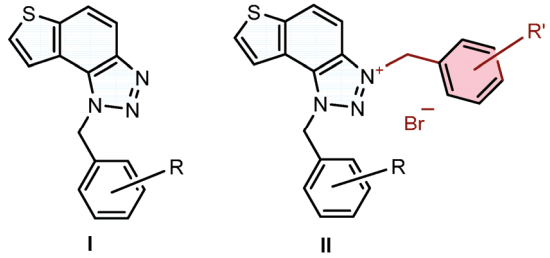
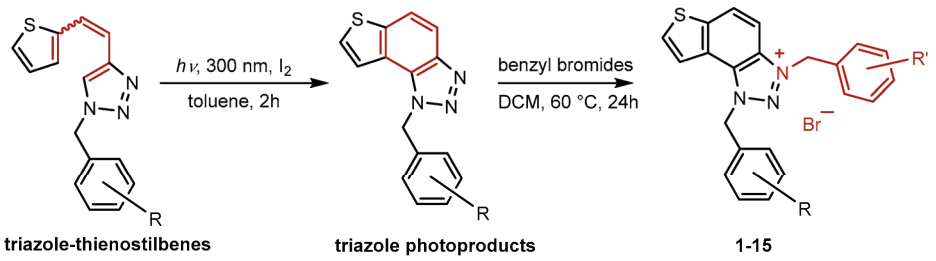


Figure 1. Structures of uncharged thienobenzo-triazoles **I** with preferred inhibition of BChE and their charged triazolinium salts **II** as dual cholinesterase inhibitors [19].

2. Results and Discussion

2.1. Synthesis of Charged Thienobenzo-1,2,3-Triazoles **1–15**

New charged thienobenzo-1,2,3-triazoles **1–15** were synthesized to explore their potential as AChE and BChE inhibitors, building on previous positive results [19]. To accomplish this, targeted triazole-thienostilbenes (Scheme 1) were prepared through a sequence of consecutive reactions [19]. Mixtures of geometrical isomers of triazole-thienostilbenes were subjected to photochemical cyclization, producing triazole photoproducts that served as starting materials for synthesizing charged triazole benzyl salts **1–15**. These photoproducts were then converted into triazolinium salts **1–15** by reacting with the corresponding benzyl bromides. NMR analysis confirmed that all bromide salts 1–15 were successfully synthesized with high yields (Figure 2, isolated yields ranging from 27% to 96%).



Scheme 1. Synthetic steps to charged triazolinium benzyl salts **1–15**.

New charged thienobenzo-1,2,3-triazolinium salts **1–15** were fully spectroscopically characterized (See Section 3 and Supplementary Materials). In the ^{1}H NMR spectra of triazolinium salts **1–15**, a new signal of the second methylene group on the charged triazole nitrogen is visible between 6.1 and 6.4 ppm, undoubtedly confirming the formation of the target structures (Figure 3).

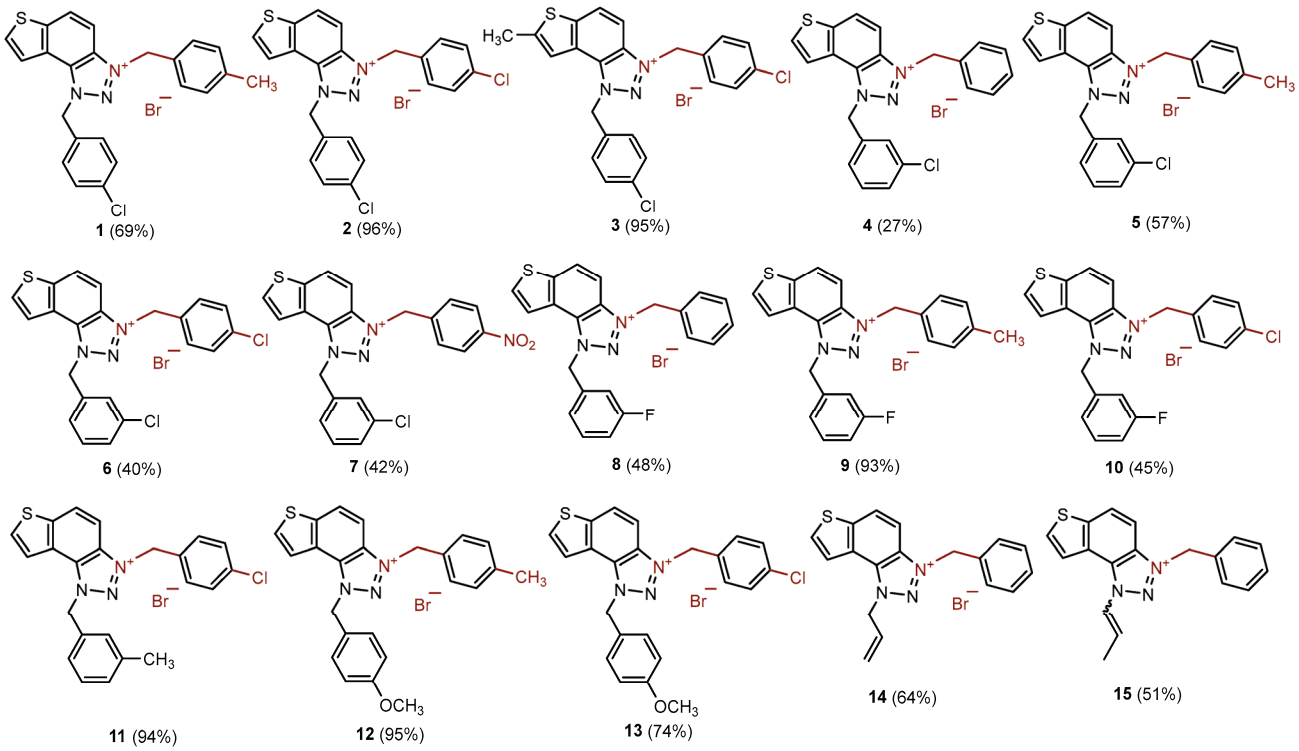


Figure 2. Structures of new charged thienobenzo-1,2,3-triazolinium salts **1**–**15** and their isolated yields in brackets.

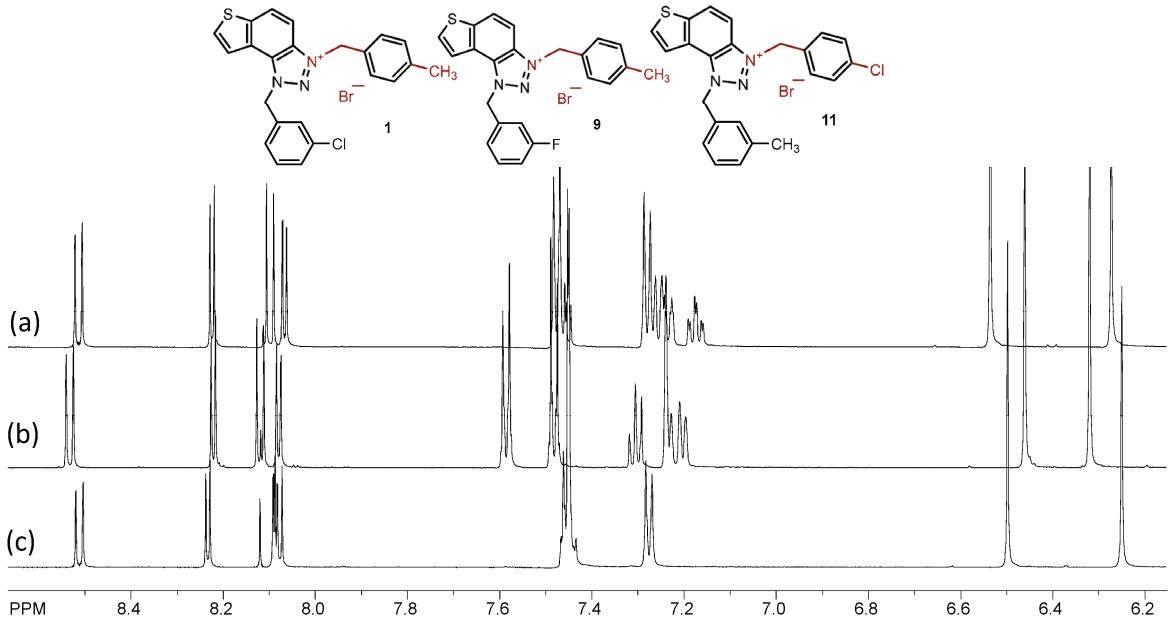


Figure 3. Parts of the ^{1}H NMR spectra of charged triazolinium salts (a) **9**, (b) **11**, and (c) **1**.

2.2. Cholinesterase Inhibition Activity of Triazolinium Salts **1–15**

To assess the biological potency of newly synthesized thienobenzo-triazolinium salts **1**–**15** as inhibitors of acetylcholinesterase (AChE) and butyrylcholinesterase (BChE), a modified Ellman's method [20] was used. The compounds were tested across a broad concentration range, depending on their solubility, and their inhibitory activities were compared to those of the commercially available drug donepezil. The IC_{50} values and the selectivity index for triazolinium salts **1**–**3** and **5**–**15** are listed in Table 1 (the amount

of triazolinium salt **4** obtained was insufficient for cholinesterase inhibition testing). As expected, based on our previous findings [19,21], the tested triazolinium salts inhibited both enzymes, AChE and BChE. Most of the compounds showed stronger inhibition of BChE. However, enzyme selectivity varies and depends on the specific molecular structure (Table 1).

Table 1. Calculated IC_{50} values for the inhibition of AChE and BChE by the charged triazolinium salts **1**–**3** and **5**–**15**.

Compound	AChE IC_{50}/μM	BChE IC_{50}/μM	Selectivity Index
1	12.74 ± 1.85	2.55 ± 0.58	5.0
2	10.71 ± 1.46	4.99 ± 1.10	2.1
3	8.89 ± 3.26	7.60 ± 0.82	1.2
5	55.62 ± 2.36	8.13 ± 0.83	6.8
6	12.90 ± 0.87	0.96 ± 0.14	13.4
7	11.42 ± 0.39	10.03 ± 1.64	1.1
8	22.23 ± 2.83	4.52 ± 0.17	4.9
9	4.80 ± 0.71	0.363 ± 0.038	13.2
10	9.96 ± 0.42	2.45 ± 0.43	4.1
11	6.23 ± 0.58	1.11 ± 0.77	5.6
12	11.81 ± 2.15	3.98 ± 1.25	3.0
13	7.40 ± 0.94	4.05 ± 1.05	1.8
14	2.13 ± 0.13	0.098 ± 0.018	21.7
15	3.62 ± 0.42	0.732 ± 0.137	4.9
Donepezil	0.023 ± 0.004	4.25 ± 0.09	0.005

Among the tested compounds, derivative **14**, which has allyl and benzyl groups on the triazole ring, stands out as the most potent and selective BChE inhibitor. The presence of the allyl group enhances inhibitory activity, as we previously found that among the non-charged thienobenzo-triazoles, the derivative with the allyl substituent was the most potent BChE inhibitor [22,23]. Introducing a charge on the triazole ring with the benzyl group increases the inhibition of both enzymes. The IC_{50} value for BChE is in the nanomolar range at 98.0 nM, which is better than the standard donepezil (IC_{50} for BChE at 4.25 μM), while the IC_{50} for inhibiting AChE is 2.13 μM, which is much weaker than donepezil's 23.0 nM (Table 1). Derivative **15**, with a propenyl instead of allyl, is also a strong BChE inhibitor, with an IC_{50} of 732.0 nM. In contrast, its IC_{50} for AChE is similar to that of **14**, at 3.62 μM.

The remaining tested compounds (**1**–**3** and **5**–**13**) contain two monosubstituted benzyl groups (with –Cl, –F, $–OCH_3$, $–NO_2$, or $–CH_3$) attached to the triazole ring, while only derivative **3** has a methyl substituent on the thiophene ring. Among these derivatives, compound **9**, with a methyl group at the para-position of one benzyl ring and fluorine at the meta-position of another benzyl ring, exhibited the most potent dual inhibitory activity, with IC_{50} values of 363.0 nM for BChE and 4.80 μM for AChE. The next most potent inhibitors were derivatives **6** and **11** (Table 1), while the others exhibited a wide range of inhibitory activity toward BChE and good to moderate activity toward AChE. Overall, the synthesized charged thienobenzo-triazoles proved to be potent inhibitors of both enzymes, with activity and selectivity depending on the type, electron-withdrawing or donating effect, and position of substituents on the rings.

2.3. Anti-Inflammatory Activity of Triazolinium Salts **1**–**15**

To assess the potential anti-inflammatory activity of the compounds, their effects on LPS-stimulated TNF-α production in human peripheral blood mononuclear cells (PBMCs)

were examined. Triazolinium salt **9** inhibited LPS-stimulated TNF-α production with an IC_{50} of 0.66 μM (Figure 4). The compound affected cell viability at the three highest concentrations tested, but at lower concentrations, it inhibited TNF-α production without reducing cell numbers. Conversely, compounds **1**, **2**, **4**, **5**, **8**, and **10**–**13** decreased cell viability, but at concentrations that did not reduce cell number, they did not inhibit TNF-α production. Compounds **3**, **6**, and **7** did not inhibit TNF-α production at any tested concentrations. Consistent with earlier results, the control compound, dexamethasone, inhibited LPS-stimulated TNF-α production with an IC_{50} of 4.6 nM (Figure 4). Notably, among derivatives **1**–**3** and **5**–**13**, all containing two monosubstituted benzyl groups attached to the triazole ring, compound **9** once again showed the most potent dual inhibitory activity, with IC_{50} values of 363.0 nM for BChE and 4.80 μM for AChE (Table 1).

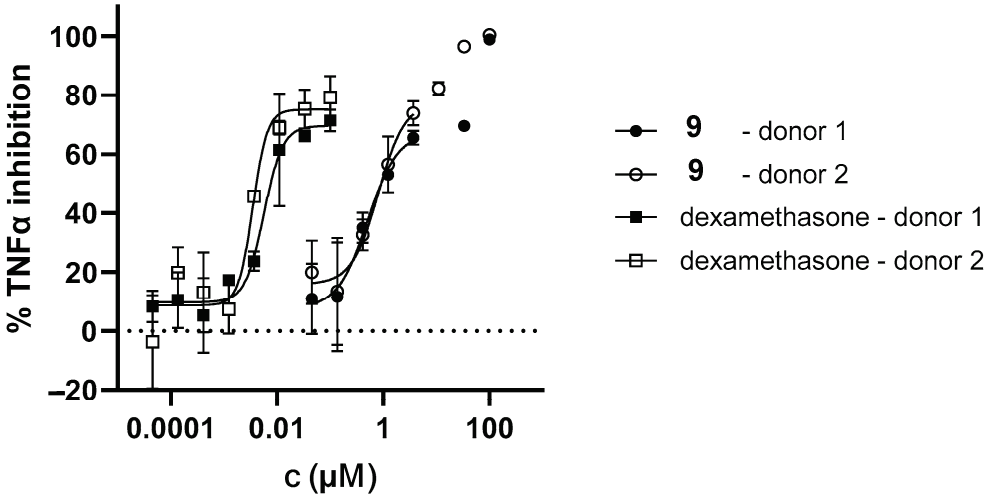


Figure 4. Inhibition of LPS-stimulated TNFα production in PBMCs from two donors for charged triazolinium bromide salt **9**.

2.4. Molecular Docking of Bioactive Triazolinium Salts **9** *and* **11**

The inhibitory activity results presented in Table 1 show that, among the tested compounds (**1**–**3** and **5**–**13**) that contain two monosubstituted benzyl groups, salts **9** and **11** demonstrate notable activity against both cholinesterases. Although not the most potent inhibitors identified, they were selected for molecular docking due to their synthetic accessibility and suitability for further analysis. Docking was performed to elucidate the structures of the non-covalent complexes they form with cholinesterases.

As illustrated in Figure 5, the complexes of salts **9** and **11** with AChE reveal that both ligands adopt a similar orientation within the enzyme's active site. The triazolinium ring is positioned toward residue Asp74 in the peripheral anionic site (PAS) due to electrostatic attraction, while the thienobenzo fragment engages in π-π stacking with Tyr341.

In the case of salt **9**, this orientation allows the meta-fluorobenzyl group attached to the triazole ring to engage in π-π interactions with Trp286 and Tyr72, both of which are located within the PAS. At the same time, the para-methylbenzyl substituent on the opposite side of the triazole ring forms π-π stacking and alkyl-π interactions with Trp86, a residue that belongs to the active site's anionic subsite. For salt **11**, which contains a para-chlorinated benzyl group on the triazole ring, π-π stacking with Trp86 is similarly observed. Additionally, the chlorine at the para position interacts to stabilize dispersion with Trp86 and His447, with interatomic distances of approximately 4.2 Å to the centroid of the tryptophan phenyl ring and 4.7 Å to the centroid of the histidine imidazole ring.

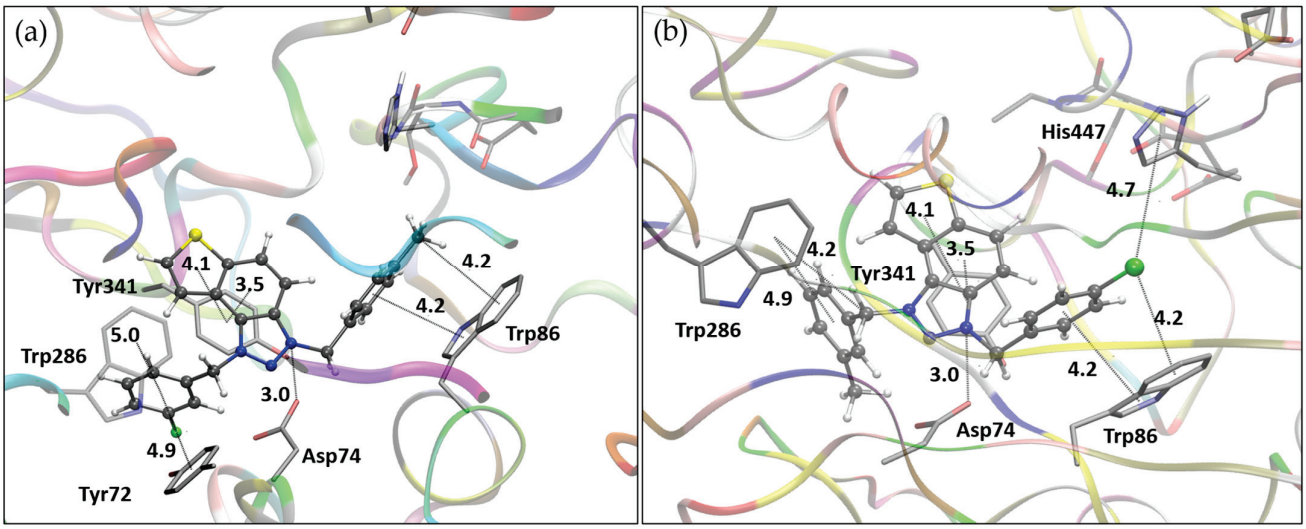


Figure 5. (**a**) The structure of triazolinium salt **9** docked into the active site of AChE. (**b**) Structure of triazolinium salt **11** docked into the active site of AChE. Ligands are presented using a ball-and-stick model, with distances given in angstroms.

The binding modes of salts **9** and **11** with BChE are depicted in Figure 6. Unlike their orientation in AChE, the ligands take on distinct conformations within the BChE active site. For salt **9**, the thienobenzo core engages in π-π stacking with Trp82, situated in the anionic site, while the triazole ring is oriented toward His438 in the catalytic site. The para-methylbenzyl substituent interacts with Phe329, with the methyl group directed toward the acyl pocket, comprising residues Leu286 and Val288. The second substituent, a meta-fluorobenzyl group linked by a flexible –CH_2– spacer, participates in parallel π-π stacking with Tyr332. In the case of salt **11**, the thienobenzo fragment is positioned near the acyl pocket and also interacts with Phe329. This orientation enables the *para*-chlorobenzyl group attached to the triazolinium ring to form π-π stacking interactions with Tyr332, while the second substituent, a *meta*-methylbenzyl group, engages in π-π interactions with Trp82.

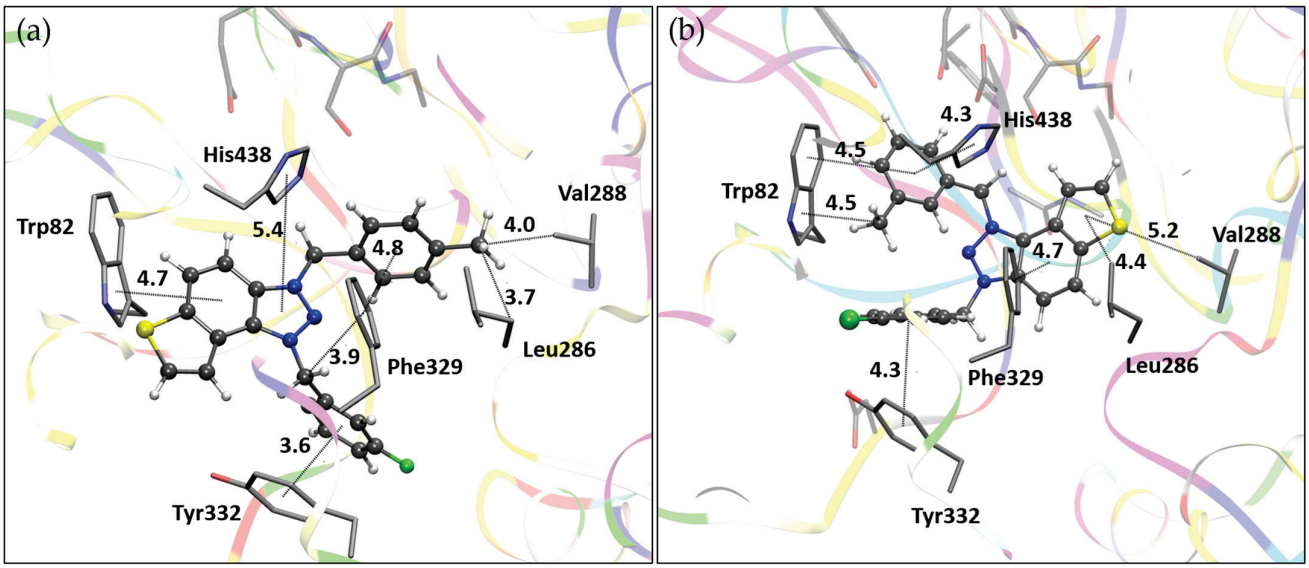


Figure 6. (**a**) The structure of triazolinium salt **9** docked into the active site of BChE. (**b**) Structure of triazolinium salt **11** docked into the active site of BChE. Ligands are presented using a ball-and-stick model, with distances given in angstroms.

2.5. Genotoxicity of Triazolinium Salts **1**–**15**

Two complementary (Q) SAR models were used to predict biological activity based on structural features [24]. In silico analyses are essential in the early stages of drug development. The commonly used software, Lhasa's M7 package, is preferred for its use of complementary models and its practice of expert review of predictions. All compounds intended for synthesis were evaluated for in silico mutagenicity during the synthesis process (Table S3). The results showed that compounds **5**, **7**, **12**, and **13** have the potential for positive mutagenicity. However, since these are only computational predictions, all compounds were still synthesized, though they have not yet been tested for biological activity. Notably, compounds **9** and **11** have shown strong potential for biological activity (as discussed above), and given their negative mutagenicity profiles, they can be considered promising candidates for further development as lead molecules.

Both compounds exhibit limited central nervous system (CNS) permeability (Table 2), as indicated by their negative log p Values (−0.910 for compound **9** and −0.888 for compound **11**), which suggests a reduced ability to cross the blood–brain barrier (BBB) and reach CNS tissues. However, their log BB values (0.883 and 0.81, respectively) are relatively high, indicating some degree of BBB penetration. This apparent contradiction may reflect differences in the modeling methods or how the compounds interact with the BBB. Despite being P-glycoprotein substrates, which can limit CNS exposure through efflux, the moderate log BB values imply that both compounds may still penetrate the CNS to some extent. Overall, both compounds demonstrate limited potential for CNS penetration, with compound **9** showing a slightly higher theoretical capacity to cross the blood–brain barrier.

Table 2. In silico analysis of additional ADME (T) indicators—absorption, distribution, metabolism, excretion, and toxicity—in the human body for new charged thienobenzo-1,2,3-triazolinium salts **9** and **11**.

Property	Model Name	9	11	Unit
Absorption	Water solubility	−5.103	−5.31	log mol/L
	Caco2	0.814	0.812	log Papp in 10^{-6} cm/s
	Intestinal absorption	100	100	% Absorbed
	Skin permeability	−2.732	−2.732	log Kp
	P–glycoprotein substrate	Yes	Yes	
	P–glycoprotein I inhibitor	No	No	
	P–glycoprotein II inhibitor	Yes	Yes	
Distribution	VDss (human)	−0.333	−0.186	log L/kg
	Fraction unbound	0.299	0.289	Fu
	BBB permeability	0.883	0.81	log BB
	CNS permeability	−0.91	−0.888	log PS
Metabolism	CYP2D6 substrate	No	No	
	CYP3A4 substrate	Yes	Yes	
	CYP1A2 inhibitor	Yes	Yes	
	CYP2C19 inhibitor	Yes	No	
	CYP2C9 inhibitor	No	No	
	CYP2D6 inhibitor	Yes	Yes	
	CYP3A4 inhibitor	No	No	
Excretion	Total clearance	0.215	1.231	log ml/min/kg
	Renal OCT2 substrate	No	No	Yes/No

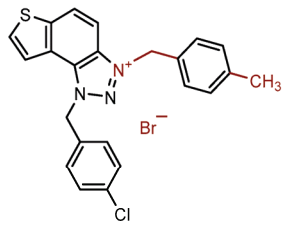
3. Materials and Methods

3.1. General Remarks

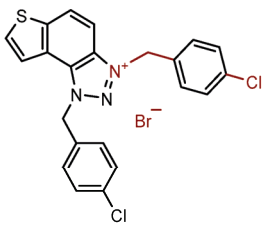
NMR spectra were obtained using either a Bruker AV300 or AV600 spectrometer (Bruker BioSpin GmbH, Rheinstetten, Germany) with a 5 mm probe. Standard ^{1}H and proton-decoupled ^{13}C{1 H} NMR spectra were recorded at a frequency of 600.130 MHz for ^{1}H and 75.432 and 150.903 MHz for ^{13}C. Chemical shifts (δ/ppm) for both ^{1}H and ^{13}C NMR spectra were referenced to the tetramethylsilane (TMS) signal. All spectra were measured in deuterated chloroform (CD_3OD) at 25 °C. Photochemical reactions were carried out in a 50.0 mL solution in quartz cuvettes that allowed light transmission. For this purpose, a Luzchem photochemical reactor equipped with UV lamps (16) with a wavelength of 300 nm was used. All solvents used in this work were purified by distillation and were commercially available. The phosphonium salts were synthesized in our laboratory, and 1-(4-nitrophenyl)-1*H*-1,2,3-triazole-4-carbaldehyde used was previously synthesized in our laboratory [25]. Reaction progress was monitored via thin-layer chromatography (TLC) using silica gel-coated plates (0.2 mm, 60/Kieselguhr F254), placed in 10 mL of the appropriate solvent system. After each synthesis, the reaction mixture was cooled to 0 °C, and then a certain amount of diethyl ether was added to precipitate the product. The resulting suspension was centrifuged (Centrifuge Eba 20, Hettich, Tuttlingen, Germany) at 2 × 3000 rpm for 10 min and then at 5 × 5000 rpm for 10 min. It was then decanted and finally evaporated. High-resolution mass spectrometry (HRMS) analyses were performed on a MALDI TOF/TOF analyzer, using an Nd:YAG laser at 355 nm with a fitting rate of 200 Hz.

3.2. Synthesis of Bromide Salts **1–15**

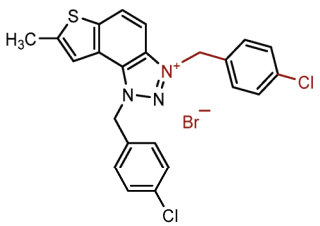
Triazole photoproducts (Scheme 1) previously developed by our group [19,21–23] served as the starting materials for the synthesis of bromide salts **1–15**. In a small vial, triazole analogues were dissolved in 0.6 mL of dry DCM and then purged briefly with argon. Furthermore, 20 eq of the corresponding benzyl bromide was added, and the reaction vial was then stirred in an oil bath at 60 °C for 24 h. Afterward, the reaction mixture was cooled to 0 °C, and then approximately 5 mL of diethyl ether was added, forming a white suspension. At the end, the suspension was centrifuged at 2 × 3000 rpm for 10 min and then at 5 × 5000 rpm for 10 min, before being decanted and evaporated using a rotary evaporator. According to NMR analyses, all bromide salts **1–15** were successfully synthesized in mostly high yields.



1



2

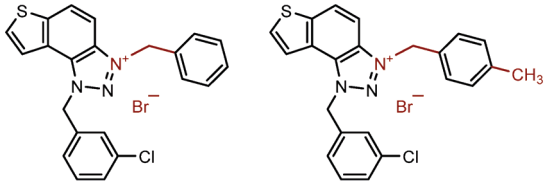


3

1-(4-chlorobenzyl)-3-(4-methylbenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**1**): 9 mg (isolated 69%), white powder; m.p. 117–118 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.51 (d, J = 9.2 Hz, 1H), 8.23 (d, J = 5.5 Hz, 1H), 8.10–8.06 (m, 2H), 7.47–7.43 (m, 6H), 7.28 (d, J = 8.1 Hz, 2H), 6.50 (s, 2H), 6.25 (s, 2H), 2.36 (s, 3H); ^{13}C NMR (CD_3OD, 150 MHz) δ/ppm: 143.0, 139.7, 135.1, 134.0, 133.1, 130.9, 130.7, 129.7, 129.4, 129.2, 129.0, 128.5, 126.9, 123.1, 120.4, 107.9, 55.3, 55.2, 19.8; HRMS (ESI) (m/z) for $C_{23}H_{19}ClN_3S^+$ Br^-: $[M + H]^+_{calcd}$ = 404.0988, and $[M + H]^+_{measured}$ = 404.0997.

1,3-bis(4-chlorobenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**2**): 13 mg (isolated 96%), white powder; m.p. 108–110 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.54 (d, J = 9.2 Hz, 1H), 8.25 (d, J = 5.7 Hz, 1H), 8.13–8.08 (m, 2H), 7.57 (d, J = 8.6 Hz, 2H), 7.50–7.43 (m, 6H), 6.50 (s, 2H), 6.30 (s, 2H); ^{13}C NMR (CD_3OD, 150 MHz) δ/ppm: 143.1, 135.1, 135.40, 134.2, 133.2, 131.0, 130.7, 130.6, 130.3, 129.5, 129.4, 129.2, 127.2, 123.2, 120.4, 107.7, 55.4, 54.4.; HRMS (ESI) (m/z) for $C_{22}H_{16}Cl_2N_3S^+Br^-$: $[M + H]^+_{calcd}$ = 424.0442, and $[M + H]^+_{measured}$ = 424.0448.

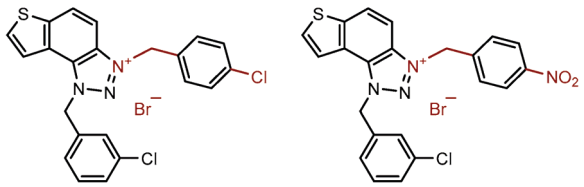
1,3-bis(4-chlorobenzyl)-7-methyl-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**3**): 10 mg (isolated 95%), white powder; m.p. 105–107 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.39 (d, J = 9.2 Hz, 1H), 8.00 (d, J = 9.2 Hz, 1H), 7.83 (s, 1H), 7.54 (d, J = 8.7 Hz, 2H), 7.49–7.43 (m, 6H), 6.45 (s, 2H), 6.26 (s, 2H), 2.77 (s, 3H); ^{13}C NMR (CD_3OD, 75 MHz) δ/ppm: 150.2, 131.7, 131.0, 130.6, 129.8, 129.4, 119.9, 108.0, 55.7, 56.7, 16.1 (several signals are missing due to the small quantity of the sample); HRMS (ESI) (m/z) for $C_{23}H_{18}C_{l2}N_3S^+Br^-$: $[M + H]^+_{calcd}$ = 438.0598, and $[M + H]^+_{measured}$ = 438.0610.



4 **5**

3-benzyl-1-(3-chlorobenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**4**): 5 mg (isolated 27%), white powder; m.p. 119–120 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.53 (d, J = 9.1 Hz, 1H), 8.25 (d, J = 5.7 Hz, 1H), 8.13–8.09 (m, 3H), 7.57 (d, J = 7.8 Hz, 2H), 7.52 (s, 1H), 7.49–7.40 (m, 5H), 6.52 (s, 2H), 6.31 (s, 2H); ^{13}C NMR (CD_3OD, 150 MHz) δ/ppm: 143.1, 135.0, 134.1, 133.2, 132.1, 131.0, 130.6, 129.4, 129.3, 129.1, 128.5, 127.8, 127.0, 126.1, 123.1, 120.4, 107.8, 55.3; MS (ESI) (m/z, %) for $C_{22}H_{17}ClN_3S^+Br^-$: $[M + H]^+$ 391 (100), 250 (85).

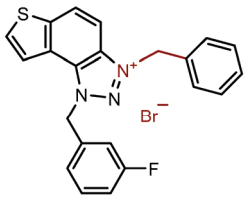
1-(3-chlorobenzyl)-3-(4-methylbenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**5**): 11 mg (isolated 57%), white powder; m.p. 104–105 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.51 (d, J = 9.2 Hz, 1H), 8.24 (d, J = 5.4 Hz, 1H), 8.12–8.08 (m, 2H), 7.53–7.36 (m, 6H), 7.28 (d, J = 8.0 Hz, 2H), 6.51 (s, 2H), 6.26 (s, 2H), 2.36 (s, 3H); ^{13}C NMR (CD_3OD, 150 MHz) δ/ppm: 144.5, 141.1, 136.4, 135.5, 135.4, 134.5, 132.4, 132.0, 131.1, 130.7, 130.0, 129.2, 128.4, 127.5, 124.5, 121.8, 109.3, 56.7, 56.6, 21.2; HRMS (ESI) (m/z) for $C_{23}H_{19}ClN_3S^+Br^-$: $[M + H]^+_{calcd}$ = 404.0988, and $[M + H]^+_{measured}$ = 404.0998.



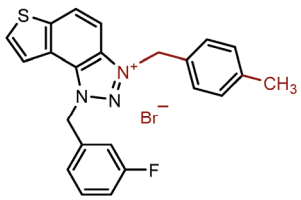
6 **7**

1-(3-chlorobenzyl)-3-(4-chlorobenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**6**): 8 mg (isolated 40%), white powder; m.p. 111–113 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.55 (d, J = 9.2 Hz, 1H), 8.26 (d, J = 5.5 Hz, 1H), 8.14–8.10 (m, 2H), 7.58 (d, J = 8.8 Hz, 2H), 7.52 (t, J = 1.5, 3.4 Hz, 1H), 7.50–7.41 (m, 4H), 7.38 (dt, J = 1.5, 3.3 Hz, 1H), 6.49 (s, 2H), 6.28 (s, 2H); ^{13}C NMR (CD_3OD, 150 MHz) δ/ppm: 143.2, 135.4, 134.9, 134.1, 133.8, 133.2, 130.6, 130.3, 129.2, 129.3, 127.9, 127.2, 126.2, 123.1, 120.3 55.3, 54.3; HRMS (ESI) (m/z) for $C_{22}H_{16}Cl_2N_3S^+Br^-$: $[M + H]^+_{calcd}$ = 424.0442, and $[M + H]^+_{measured}$ = 424.0447.

1-(3-chlorobenzyl)-3-(4-nitrobenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**7**): 9 mg (isolated 42%), white powder; m.p. 125–126 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.57 (d, J = 9.2 Hz, 1H), 8.32 (d, J = 8.7 Hz, 2H), 8.27 (d, J = 5.7 Hz, 1H), 8.15–8.11 (m, 3H), 7.80–7.77 (m, 2H), 7.47–7.38 (m, 3H) 6.53 (s, 2H), 6.47 (s, 2H); ^{13}C NMR (CD_3OD, 75 MHz) δ/ppm: 140.3, 136.4, 135.9, 135.3, 134.8, 132.1, 131.1, 130.6, 130.7, 129.4, 128.8, 127.7, 125.4, 124.6, 123.9, 121.8, 110.0, 56.9, 55.4; HRMS (ESI) (m/z) for $C_{22}H_{16}ClN_4O_2S^+Br^-$: $[M + H]^+_{calcd}$ = 435.0682, and $[M + H]^+_{measured}$ = 435.0690.



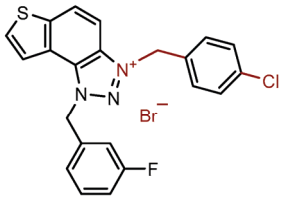
8



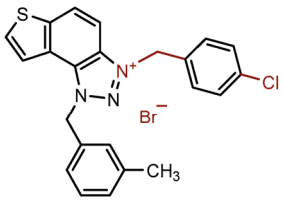
9

3-benzyl-1-(3-fluorobenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**8**): 7 mg (isolated 48%), white powder; m.p. 103–104 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.53 (d, J = 9.1 Hz, 1H), 8.23 (d, J = 5.5 Hz, 1H), 8.13–8.05 (m, 3H), 7.58 (dd, J = 1.8, 8.0 Hz, 2H), 7.49–7.42 (m, 4H), 7.27–7.21 (m, 2H), 6.53 (s, 2H), 6.31 (s, 2H); ^{13}C NMR (CD_3OD, 75 MHz) δ/ppm: 145.3 (d, J_{CF} = 236.7 Hz), 143.7, 139.1, 138.0, 137.1, 134.9, 134.6, 133.6, 133.4, 133.1, 132.9, 132.5, 130.9, 127.1, 124.3, 111.8, 59.3, 59.2 (signals for four quaternary C are missing); HRMS (ESI) (m/z) for $C_{22}H_{17}FN_3S^+Br^-$: $[M + H]^+_{calcd}$ = 374.1127, and $[M + H]^+_{measured}$ = 374.1131.

1-(3-fluorobenzyl)-3-(4-methylbenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**9**): 14 mg (isolated 93%), white powder; m.p. 116–117 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.51 (dd, J = 0.7, 9.2 Hz, 1H), 8.22 (d, J = 5.2 Hz, 1H), 8.12–8.05 (m, 2H), 7.50–7.43 (m, 3H), 7.30–7.22 (m, 4H), 7.19–7.15 (m, 1H), 6.53 (s, 2H), 6.27 (s, 2H), 2.36 (s, 3H); ^{13}C NMR (CD_3OD, 75 MHz) δ/ppm: 167.1 (d, J_{CF} = 246.2 Hz), 147.0, 143.7, 138.4 (d, J_{CF} = 8.0 Hz), 138.0, 137.0, 135.0 (d, J_{CF} = 8.7 Hz), 133.6, 132.9, 132.5, 130.9, 127.4, 127.3, 127.0, 124.3, 119.9 (d, J_{CF} = 21.7 Hz), 118.5 (d, J_{CF} = 23.3 Hz), 59.3, 59.2, 23.7; HRMS (ESI) (m/z) for $C_{23}H_{19}FN_3S^+Br^-$: $[M + H]^+_{calcd}$ = 388.1284, and $[M + H]^+_{measured}$ = 388.1292.



10

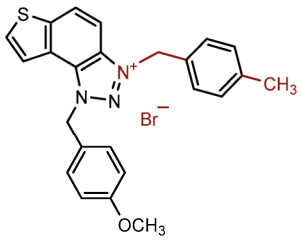


11

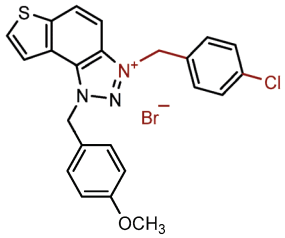
3-(4-chlorobenzyl)-1-(3-fluorobenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**10**): 7 mg (isolated 45%), white powder; m.p. 110–112 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.55 (dd, J = 0.7, 9.2 Hz, 1H), 8.24 (d, J = 5.5 Hz, 1H), 8.14–8.10 (m, 1H), 8.08–8.06 (m, 1H), 7.58 (d, J = 8.6 Hz, 2H), 7.50–7.44 (m, 3H), 7.29–7.15 (m, 3H), 6.53 (s, 2H), 6.31 (s, 2H); ^{13}C NMR (CD_3OD, 150 MHz) δ/ppm: 143.2, 136.3 (d, J_{CF} = 216.2 Hz), 135.4, 134.2, 133.6, 133.2, 132.7, 131.1 (d, J_{CF} = 8.1 Hz), 130.7, 130.3, 129.2, 127.2, 123.5, 123.1, 120.4, 115.9 (d, J_{CF} = 21.9 Hz), 114.6 (d, J_{CF} = 23.2 Hz), 107.7, 55.4, 54.5 (signals for 2 quaternary C are missing); HRMS (ESI) (m/z) for $C_{22}H_{16}ClFN_3S^+Br^-$: $[M + H]^+_{calcd}$ = 408.0738, and $[M + H]^+_{measured}$ = 408.0745.

3-(4-chlorobenzyl)-1-(3-methylbenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**11**): 8 mg (isolated 94%), white powder; m.p. 122–123 °C; ^{1}H NMR (CD_3OD,

600 MHz) δ/ppm: 8.53 (d, J = 9.2 Hz, 1H), 8.22 (d, J =5.5 Hz, 1H), 8.14–8.06 (m, 2H), 7.58 (d, J = 8.6 Hz, 2H), 7.48 (d, J = 8.5 Hz, 2H), 7.32–7.17 (m, 4H), 6.46 (s, 2H), 6.32 (s, 2H), 2.32 (s, 3H); ^{13}C NMR (CD_3OD, 75 MHz) δ/ppm: 144.5, 140.7, 136.8, 135.6, 134.5, 133.2, 132.4, 131.7, 131.2, 130.6, 130.4, 129.4, 128.5, 125.9, 124.6, 122.0, 109.1, 55.8, 57.5, 21.3; HRMS (ESI) (m/z) for $C_{23}H_{19}ClN_3S^+Br^-$: $[M + H]^+_{calcd}$ = 404.0988, and $[M + H]^+_{measured}$ = 404.0996.

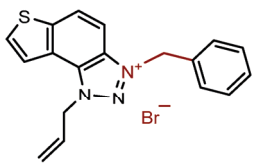


12

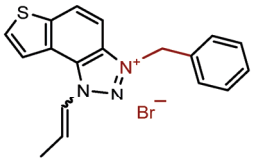


13

1-(4-methoxybenzyl)-3-(4-methylbenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**12**): 9 mg (isolated 95%), white powder; m.p. 120–122 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.49 (d, J = 9.5 Hz, 1H), 8.22 (d, J = 5.5 Hz, 1H), 8.15–8.11 (m, 1H), 8.06 (d, J = 9.2 Hz, 1H), 7.44 (d, J = 8.2 Hz, 2H), 7.39 (d, J = 8.9 Hz, 2H), 7.27 (d, J = 8.2 Hz, 2H), 6.97 (d, J = 8.9 Hz, 2H), 6.42 (s, 2H), 6.24 (s, 2H), 3.79 (s, 3H), 2.35 (s, 3H); ^{13}C NMR (CD_3OD, 150 MHz) δ/ppm: 142.9, 139.7, 134.3, 133.9, 132.9, 130.8, 129.7, 129.2, 129.1, 128.4, 126.8, 123.6, 123.2, 120.6, 114.3, 107.8, 65.4, 54.3, 53.4, 19.8; HRMS (ESI) (m/z) for $C_{24}H_{22}N_3OS^+Br^-$: $[M + H]^+_{calcd}$ = 400.1484, and $[M + H]^+_{measured}$ = 400.1487.
3-(4-chlorobenzyl)-1-(4-methoxybenzyl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**13**): 7 mg (isolated 74%), white powder; m.p. 124–125 °C; ^{1}H NMR (CD_3OD, 600 MHz) δ/ppm: 8.52 (dd, J = 0.7, 9.2 Hz, 1H), 8.24 (d, J = 5.5 Hz, 1H), 8.15 (d, J = 5.5 Hz, 1H), 8.09 (d, J = 9.2 Hz, 1H), 7.36 (d, J = 8.6 Hz, 2H), 7.47 (d, J = 8.6 Hz, 2H), 7.40 (d, J = 8.8 Hz, 2H), 7.00 (d, J = 8.7 Hz, 2H), 6.42 (s, 2H), 6.29 (s, 2H), 3.80 (s, 3H); ^{13}C NMR (CD_3OD, 75 MHz) δ/ppm: 134.5, 132.3, 131.6, 130.7, 130.6, 128.5, 125.5, 124.9, 122.1, 115.8, 109.1, 57.3, 55.8, 55.7 (signals for 5 quaternary C are missing); HRMS (ESI) (m/z) for $C_{23}H_{19}ClN_3OS^+Br^-$: $[M + H]^+_{calcd}$ = 420.0937, and $[M + H]^+_{measured}$ = 420.0943.



14



15

1-allyl-3-benzyl-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**14**): 8 mg (isolated 64%), white powder; m.p. 101–102 °C; ^{1}H NMR (CD_3OD, 300 MHz) δ/ppm: 8.52 (d, J = 9.2 Hz, 1H), 8.29 (d, J = 5.5 Hz, 1H), 8.18 (d, J = 5.6 Hz, 1H), 8.14–8.05 (m, 2H), 7.61–7.53 (m, 2H), 7.49–7.41 (m, 2H), 6.40–6.32 (m, 1H), 6.30 (s, 2H), 5.92 (d, J = 5.5 Hz, 2H), 5.52 (d, J = 10.4 Hz, 1H), 5.41 (d, J = 17.1 Hz, 1H); ^{13}C NMR (CD_3OD, 75 MHz) δ/ppm: 142.9, 133.9, 133.02, 132.1, 129.3, 129.1, 128.8, 128.5, 126.9, 123.2, 120.7, 120.2, 114.5, 107.7, 55.2, 54.9; MS (EI) (m/z, %) for $C_{18}H_{16}N_3S^+Br^-$: 306 (100).
3-benzyl-1-(prop-1-en-1-yl)-1*H*-thieno[3′,2′:3,4]benzo[1,2-*d*][1,2,3]triazol-3-ium bromide (**15**): 11 mg (isolated 51%), yellow oil; mixture of *cis*- and *trans*-isomer; MS (EI) (m/z, %) for $C_{18}H_{16}N_3S^+Br^-$: 306 (100).

3.3. In Vitro Cholinesterase Inhibition Activity Measurements

The inhibition of AChE and BChE was determined using a modified Ellman's spectrophotometric assay [20]. Enzymes AChE (EC 3.1.1.7, *Electrophorus electricus*, Type V-S) and BChE (EC 3.1.1.8, equine serum) were purchased from Sigma-Aldrich (St. Louis, MO, USA). All the remaining chemicals, Trisma base (2-amino-2-(hydroxymethyl)-1,3-propanediol), acetylthiocholine iodide (ATChI), *S*-butyrylthiocholine iodide (BTChI), and standard Donepezil, were also purchased from Sigma-Aldrich, while Ellman's reagent 5,5-dithiobis-(2-nitrobenzoic acid) (DTNB) was sourced from Zwijndrecht (Antwerpen, Belgium). The various concentrations of the tested compounds dissolved in ethanol were added (10 μL) to 180 μL of Tris buffer (50 mM, pH 8.0), 10 μL of enzyme solution (final concentration 0.03 U/mL), and 10 μL of DTNB (final concentration 0.3 mM). The enzymatic reaction was initiated by the addition of 10 μL of substrate ATChI or BTChI, resulting in a final substrate concentration of 0.5 mM. Non-enzymatic hydrolysis was measured as a blank for control measurement without inhibitors. The non-enzymatic hydrolysis reaction with added inhibitor was used as a blank for the samples. The enzyme was replaced with an equivalent amount of buffer. Absorbance was measured at 405 nm over 5 min at room temperature using a 96-well microplate reader (BioTek 800TSUV Absorbance Reader, Agilent, Santa Clara, CA, USA). All measurements were conducted in triplicate. The percentage of inhibition was calculated using the following equation: Inhibition (%) = $[(A_C - A_T)/A_C] \times 100$, where AC represents the enzyme activity in the absence of the test sample, and AT represents the enzyme activity in the presence of the test compound. Inhibition data were used to calculate the IC_{50} value by nonlinear fitting of the inhibitor concentration versus response. The inhibitory activity of ethanol was also measured, and its contribution to inhibition was subtracted. The IC_{50} value is represented as a mean value of three measurements ± standard deviation. The selectivity index is calculated as IC_{50}(AChE)/IC_{50}(BChE).

3.4. Anti-Inflammatory Activity

The effect of triazolinium salts **1**–**15** on lipopolysaccharide (LPS)-stimulated tumor necrosis factor alpha (TNF-α) production was analyzed as described before [26]. Peripheral blood mononuclear cells (PBMCs) were isolated from buffy coats obtained from healthy adult volunteers and resuspended in RPMI1640 medium (Capricorn Scientific, Ebsdorfergrund, Germany) supplemented with 10% heat-inactivated FBS (Biowest, Riverside, MO, USA), 1% GlutaMAX (Gibco, Waltham, MA, USA), and 1% Antibiotic-Antimycotic (Gibco). In a 96-well plate, 2×10^5 PBMCs were seeded per well. Triazolinium salts **1**–**15** were dissolved in dimethyl sulfoxide solvent (DMSO, Sigma), and three-fold serial dilutions in DMSO were assembled and added to the cells, starting at a concentration of 100 μM. After a 1-h pre-incubation with triazolinium salts **1**–**15**, cells were stimulated with 1 ng/mL LPS from *E. coli* 0111:B4 (Sigma). Cells were incubated for 24 h at 37 °C in 5% CO_2, followed by the collection of supernatants for TNF-α measurement and cell viability assessment. To study cell viability, CellTiter-Glo reagent (Promega, Madison, WI, USA) was applied according to the manufacturer's instructions, and signals obtained from compound-treated wells were compared with those in LPS-stimulated, vehicle-treated samples.

TNF-α concentration in supernatants was measured by ELISA utilizing antibodies and recombinant human TNF-α protein (standard) from R&D Systems (Minneapolis, MN, USA). Lumitrac 600 384-well plates (Greiner Bio-One, Kremsmünster, Austria) were coated with 1 μg/mL of TNF-α capture antibody diluted in phosphate-buffered saline (PBS; Gibco) overnight at 4 °C. The next day, plates were blocked with 5% sucrose (Kemika, Ovada, Italy) in assay diluent (1% bovine serum albumin [BSA; Sigma] in PBS) for 4 h at room

temperature (RT). After blocking, samples and standards from triazolinium salts **1**–**15** were joined to the plates and incubated overnight at 4 °C. Subsequently, 250 ng/mL of TNF-α detection antibody was added to the wells, followed by a 2-h incubation at RT. Finally, after incubation with streptavidin-HRP (Invitrogen, Waltham, MA, USA), chemiluminescence ELISA Substrate (Roche, Basel, Switzerland) was added, and luminescence was measured using an EnVision 2105 multilabel reader (Revvity, Waltham, MA, USA). Measured luminescence was used to determine the concentrations of TNF-α in the supernatants by interpolation from the standard curve. Percentages of inhibition were calculated from the cytokine concentrations obtained, and IC_{50} values were determined using GraphPad Prism v9 software with a nonlinear regression curve fit (four parameters with variable slope).

3.5. Computational Details

Conformational analysis and geometry optimization of two compounds chosen for the docking study were carried out using the Gaussian16 program package [27], at the B3LYP/6-31G(d) level of theory. The most stable conformers were used as ligands in the molecular docking process, conducted with the AutoDock4 suite [28]. Crystal structures 4EY7.pdb and 1P0I.pdb, representing AChE and BChE, respectively, were downloaded from the Protein Data Bank [29,30] and prepared for docking (non-amino acid residues removed, polar hydrogens added, and Kollman charges assigned). The Lamarckian Genetic Algorithm was used for docking, with 25 runs per ligand, while the enzyme residues remained rigid.

3.6. ADME-Tox Predictions

pkCSM is a free-to-use machine learning programme that anticipates small-molecule pharmacokinetic values using graph-based signatures. It covers 28 models with key ADMET effects, such as permeability, solubility, absorption, distribution in the body, interactions with metabolic enzymes, excretion, and various toxicity measures. Swiss ADME permits computing physicochemical descriptors for free and predicts ADME parameters, pharmacokinetic properties, drug-like nature, and medicinal chemistry friendliness for one or multiple small molecules to support drug discovery. AdmetSAR 3.0 is a free-access chemical risk assessment tool [31]. ADMETLAb 3.0 provides easy access to comprehensive, accurate, and efficient predictions of ADMET profiles for compounds. These predictions are based on a high-quality database of 0.37 million entries spanning 77 endpoints and the Directed Message Passing Neural Network (DMPNN) framework. For this paper, all triazolinium salts **1**–**15** were evaluated using multiple tools, and predictions of key features for the lead molecules were presented.

4. Conclusions

The findings of this study demonstrate that the newly synthesized charged thienobenzo-1,2,3-triazole derivatives exhibit strong inhibitory activity against both acetylcholinesterase (AChE) and butyrylcholinesterase (BChE), with a clear tendency toward higher potency and selectivity for BChE. This makes them promising candidates for treating disorders that require peripheral cholinesterase inhibition. Introducing a permanent positive charge on the triazole ring was a key structural modification that significantly enhanced enzyme binding affinity, particularly for BChE. Among all of the tested compounds, derivative 14 stood out with sub-micromolar inhibition of both enzymes and excellent selectivity. Meanwhile, compound **9** demonstrated not only potent dual inhibition but also notable anti-inflammatory activity in a TNF-α suppression assay. These dual actions suggest that certain charged triazole derivatives could serve as multifunctional drug candidates,

combining cholinesterase inhibition with immunomodulatory effects. This is especially relevant for neurodegenerative diseases involving both cholinergic deficits and chronic inflammation, such as Alzheimer's disease. Additionally, molecular docking studies provided structural insights into ligand-enzyme interactions, supporting the biological data and confirming stable and specific binding of ligands to key residues in the active sites of AChE and BChE. Besides biological testing, in silico ADME-Tox profiling of selected lead compounds indicated favorable pharmacokinetic and safety profiles, including good solubility, moderate permeability, non-mutagenicity, and predictable metabolism. These properties, along with their synthetic accessibility and chemical stability, enhance the potential of these molecules for further development in preclinical studies. In summary, this work lays a solid foundation for designing new charged heterocyclic compounds with dual therapeutic functions. The promising in vitro and in silico results of derivatives **9** and **14** highlight the importance of structural charge modulation in optimizing cholinesterase inhibitors that are selective for the periphery and offer additional therapeutic benefits.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/ph18071032/s1: ^{1}H and ^{13}C NMR spectra of new charged triazolinium salts **1**–**15** (Figures S1–S30); Mass spectra and HRMS analyses of new charged triazolinium salts **1**–**15** (Figures S31–S43); Cartesian coordinates of ligands **9** and **11** docked into the active site of AChE; Cartesian coordinates of ligands **9** and **11** docked into the active site of BChE; Free energies of binding, the number of conformational clusters, and distribution of conformations obtained by molecular docking (Table S1. Free energies of binding, ΔG_{bind}, obtained by molecular docking of ligands **9**, **11**, and reference ligand donepezil into the active site of AChE (4EY7.pdb), along with the number of conformational clusters and distribution of conformations and Table S2. Free energies of binding, ΔG_{bind}, obtained by molecular docking of ligands **9**, **11**, and reference ligand donepezil into the active site of BChE (1P0I.pdb), along with the number of conformational clusters and distribution of conformations); The mutagenic potential of new charged thienobenzo-1,2,3-triazolinium salts **1**–**15** through Lhasa M7 evaluation (Table S3).

Author Contributions: Conceptualization, I.O. and I.Š. (Irena Škorić); methodology, A.J., A.R. (Ana Ratković), A.R. (Anamarija Raspudić), D.B., M.B., and I.Š. (Irena Škorić); formal analysis, Z.L. and S.R.; investigation, P.P., D.Š., A.R. (Anamarija Raspudić), and A.J.; resources, M.B., I.O., D.B., I.Š. (Ivana Šagud), and I.Š. (Irena Škorić); writing—original draft preparation, A.J., M.B., I.O., D.B., I.Š. (Ivana Šagud), and I.Š. (Irena Škorić); writing—review and editing, D.B. and I.Š. (Irena Škorić); supervision, I.Š. (Irena Škorić). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Sample collection was approved by the Ethics Committee of the Croatian Institute of Transfusion Medicine (251-541-06/6-20-2, 20 March 2020), and all donors provided written informed consent.

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

Data Availability Statement: Dataset available on request from the authors.

Acknowledgments: This work was supported by grants from the University of Zagreb for short-term scientific support for 2024 under the title Synthesis and Biological Activity of New Heteropolycycle Systems. We thank the University of Zagreb (Croatia) Computing Centre (SRCE) for granting computational time on the Supercomputer Supek.

Conflicts of Interest: The authors Paula Pongrac, Dora Štefok, and Martina Bosnar were employed by the company Selvita Ltd. Zlata Lasić is employed by the Pliva Hrvatska d.o.o. The remaining authors

declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Masson, P.; Lockridge, O. Butyrylcholinesterase for protection from organophosphorus poisons: Catalytic complexities and therapeutic promises. *Arch. Biochem. Biophys.* **2010**, *15*, 107–120. [CrossRef] [PubMed]
2. Colović, M.B.; Krstić, D.Z.; Lazarević-Pašti, T.D.; Bondžić, A.M.; Vasić, V.M. Acetylcholinesterase inhibitors: Pharmacology and toxicology. *Curr. Neuropharmacol.* **2013**, *11*, 315–335. [CrossRef] [PubMed]
3. Anand, P.; Singh, B. A review on cholinesterase inhibitors for Alzheimer's disease. *Arch. Pharmacal Res.* **2013**, *36*, 375–399. [CrossRef] [PubMed]
4. Talesa, V.N. Acetylcholinesterase in Alzheimer's disease. *Curr. Drug Targets* **2001**, *2*, 363–373. [CrossRef]
5. Greig, N.H.; Utsuki, T.; Yu, Q.S.; Holloway, H.W. A new therapeutic target in Alzheimer's disease treatment: Selective butyryl-cholinesterase inhibition. *Curr. Med. Chem.* **2005**, *12*, 237–243.
6. Mesulam, M.M.; Guillozet, A.; Shaw, P.; Levey, A. Acetylcholinesterase knockouts establish centrality of cholinergic networks. *Ann. Neurol.* **2002**, *52*, 253–256.
7. Lane, R.M.; Potkin, S.G.; Enz, A. Targeting acetylcholinesterase and butyrylcholinesterase in dementia. *Int. J. Neuropsychopharmacol.* **2006**, *1*, 101–124. [CrossRef]
8. Waghmare, A.A.; Kadam, P.; Sharma, H. Clinical application of neostigmine and its peripheral selectivity. *Indian J. Anaesth.* **2022**, *66*, 738–745.
9. Silman, I.; Sussman, J.L. Acetylcholinesterase: 'Classical' and 'non-classical' functions and pharmacology. *Curr. Opin. Pharmacol.* **2005**, *5*, 293–302. [CrossRef]
10. Taylor, P. Anticholinesterase agents. In *Basic & Clinical Pharmacology*, 12th ed.; Katzung, B.G., Ed.; McGraw-Hill: New York, NY, USA, 2011; pp. 123–134.
11. Wu, D.; Chen, Q.; Chen, X.; Han, F.; Chen, Z.; Wang, Y. The blood–brain barrier: Structure, regulation and drug delivery. *Sig. Transduct. Target Ther.* **2023**, *8*, 217. [CrossRef]
12. Bajgar, J.; Kuca, K.; Fusek, J.; Karasova, J.; Kassa, J.; Cabal, J.; Jun, D.; Blaha, V. Inhibition of blood cholinesterases following intoxication with VX and its derivatives. *J. Appl. Toxicol.* **2007**, *5*, 458–463. [CrossRef] [PubMed]
13. Seidl, S.E.; Potashkin, J.A. The promise of neuroprotective agents in Parkinson's disease. *Front. Neurol.* **2011**, *2*, 68. [CrossRef] [PubMed]
14. Cacabelos, R.; Martínez-Iglesias, O.; Cacabelos, N.; Carrera, I.; Corzo, L.; Naidoo, V. Therapeutic Options in Alzheimer's Disease: From Classic Acetylcholinesterase Inhibitors to Multi-Target Drugs with Pleiotropic Activity. *Life* **2024**, *14*, 1555. [CrossRef] [PubMed]
15. Žužek, M.C. Advances in Cholinesterase Inhibitor Research-An Overview of Preclinical Studies of Selected Organoruthenium(II) Complexes. *Int. J. Mol. Sci.* **2024**, *25*, 9049. [CrossRef]
16. Singh, M.; Kaur, M.; Kukreja, H.; Chugh, R.; Silakari, O.; Singh, D. Acetylcholinesterase inhibitors as Alzheimer therapy: From nerve toxins to neuroprotection. *Eur. J. Med. Chem.* **2013**, *70*, 165–188. [CrossRef]
17. Roth, B.L.; Driscol, J.; Glennon, R.A. Drugs with anticholinergic properties: Functional roles and side effects. *Pharmacol. Rev.* **2002**, *54*, 364–385.
18. Krátký, M.; Vinšová, J.; Buchta, V.; Stolaříková, J. Quaternary ammonium-based cholinesterase inhibitors: Synthesis and biological evaluation. *Bioorganic Med. Chem. Lett.* **2016**, *26*, 1125–1130.
19. Mlakić, M.; Sviben, M.; Ratković, A.; Raspudić, A.; Barić, D.; Šagud, I.; Lasić, Z.; Odak, I.; Škorić, I. Efficient Access to New Thienobenzo-1,2,3-Triazolium Salts as Preferred Dual Cholinesterase Inhibitors. *Biomolecules* **2024**, *14*, 1391. [CrossRef]
20. Ellman, G.L.; Courtnex, K.D.; Andres, V.; Featherstone, R.M. A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochem. Pharmacol.* **1961**, *7*, 88–95. [CrossRef]
21. Mlakić, M.; Barić, D.; Ratković, A.; Šagud, I.; Čipor, I.; Piantanida, I.; Odak, I.; Škorić, I. New Charged Cholinesterase Inhibitors: Design, Synthesis, and Characterization. *Molecules* **2024**, *29*, 1622. [CrossRef]
22. Mlakić, M.; Odak, I.; Faraho, I.; Talić, S.; Bosnar, M.; Lasić, K.; Barić, D.; Škorić, I. New naphtho/thienobenzo-triazoles with interconnected anti-inflammatory and cholinesterase inhibitory activity. *Eur. J. Med. Chem.* **2022**, *241*, 114616. [CrossRef] [PubMed]
23. Mlakić, M.; Selec, I.; Ćaleta, I.; Odak, I.; Barić, D.; Ratković, A.; Molčanov, K.; Škorić, I. New Thienobenzo/Naphtho-Triazoles as Butyrylcholinesterase Inhibitors: Design, Synthesis and Computational Study. *Int. J. Mol. Sci.* **2023**, *24*, 5879. [CrossRef] [PubMed]

24. Hasselgren, C.; Bercu, J.; Cayley, A.; Cross, K.; Glowienke, S.; Kruhlak, N.; Muster, W.; Nicolette, J.; Vijayaraj Reddy, M.; Saiakhov, R.; et al. Management of Pharmaceutical ICH M7 (Q)SAR Predictions–The Impact of Model Updates. *Regul. Toxicol. Pharmacol.* **2020**, *118*, 104807. [CrossRef] [PubMed]
25. Ratković, A.; Mlakić, M.; Dehaen, W.; Opsomer, T.; Barić, D.; Škorić, I. Synthesis and photochemistry of novel 1,2,3-triazole di-heterostilbenes. An experimental and computational study. *Spectrochim. Acta. Part A Mol. Biomol. Spectrosc.* **2021**, *261*, 120056. [CrossRef]
26. Mlakić, M.; Faraho, I.; Odak, I.; Kovačević, B.; Raspudić, A.; Šagud, I.; Bosnar, M.; Škorić, I.; Barić, D. Cholinesterase Inhibitory and Anti-Inflammatory Activity of the Naphtho- and Thienobenzo-Triazole Photoproducts: Experimental and Computational Study. *Int. J. Mol. Sci.* **2023**, *24*, 14676. [CrossRef]
27. 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 16, Revision C01*; Gaussian, Inc.: Wallingford, CT, USA, 2016.
28. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDock-Tools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [CrossRef]
29. Cheung, J.; Rudolph, M.; Burshteyn, F.; Cassidy, M.; Gary, E.; Love, J.; Height, J.; Franklin, M. Crystal Structure of Recombinant Human Acetylcholinesterase in Complex with Donepezil. *J. Med. Chem.* **2012**, *55*, 10282–10286. [CrossRef]
30. Nicolet, Y.; Lockridge, O.; Masson, P.; Fontecilla-Camps, J.C.; Nachon, F. Crystal Structure of Human Butyrylcholinesterase. *J. Biol. Chem.* **2003**, *278*, 41141–41147. [CrossRef]
31. Yang, H.; Lou, C.; Sun, L.; Li, J.; Cai, Y.; Wang, Z.; Li, W.; Liu, G.; Tang, Y. AdmetSAR 2.0: Web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics* **2019**, *35*, 1067–1069. [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.

pharmaceuticals

MDPI

Article

Exploring Brazilian Green Propolis Phytochemicals in the Search for Potential Inhibitors of B-Raf^{600}E Enzyme: A Theoretical Approach

Garcia Ferreira de Souza [1], Airis Farias Santana [1,2], Fernanda Sanches Kuhl Antunes [2], Ramon Martins Cogo [3], Matheus Dornellas Pereira [3], Daniela Gonçales Galasse Rando [3] and Carolina Passarelli Gonçalves [3,*]

[1] Biological Science Department, Instituto Universitario Italiano de Rosario, Corrientes 720, PB y 1er Piso Rosario S2000CTT, Santa Fe, Argentina; garciafdesouza@gmail.com (G.F.d.S.); iris.farias@icloud.com (A.F.S.)
[2] Cosmobeauty, Avenida Aruanã 1160, Barueri 06460-010, SP, Brazil; fernanda.sanches.dra@gmail.com
[3] Postgraduate Program in Chemical Biology, Universidade Federal de São Paulo, Diadema 09913-030, SP, Brazil; ramon.cogo@unifesp.br (R.M.C.); matheus.dornellas16@unifesp.br (M.D.P.); dgrando@unifesp.br (D.G.G.R.)
* Correspondence: carolina.passarelli@unifesp.br

Abstract: Background/Objectives: Melanoma is one of the most aggressive forms of skin cancer and is frequently associated with the B-Raf^{600}E mutation, which constitutively activates the MAPK signaling pathway. Although selective inhibitors such as Vemurafenib offer clinical benefits, their long-term efficacy is often hindered by resistance mechanisms and adverse effects. In this study, twelve phytochemicals from Brazilian green propolis were evaluated for their potential as selective B-Raf^{600}E inhibitors using a computational approach. **Methods:** Physicochemical, ADME, and electronic properties were assessed, followed by molecular docking using the B-Raf^{600}E crystal structure (PDB ID: 3OG7). Redocking validation and 500 ns molecular dynamics simulations were performed to investigate the stability of the ligand-protein complexes, and free energy calculations were then computed. Results: Among the tested compounds, Artepillin C exhibited the strongest binding affinity (−8.17 kcal/mol) in docking and maintained stable interactions with key catalytic residues throughout the simulation, also presenting free energy of binding ΔG of −20.77 kcal/mol. HOMO-LUMO and electrostatic potential analyses further supported its reactivity and selectivity. Notably, Artepillin C remained bound within the ATP-binding site, mimicking several critical interactions observed with Vemurafenib. **Results:** Among the tested compounds, Artepillin C exhibited the strongest binding affinity (−8.17 kcal/mol) and maintained stable interactions with key catalytic residues throughout the simulation. HOMO-LUMO and electrostatic potential analyses further supported its reactivity and selectivity. Notably, Artepillin C remained bound within the ATP-binding site, mimicking several critical interactions observed with Vemurafenib. **Conclusions:** These findings indicate that Artepillin C is a promising natural compound for further development as a selective B-Raf^{600}E inhibitor and suggest its potential utility in melanoma treatment strategies. This study reinforces the value of natural products as scaffolds for targeted drug design and supports continued experimental validation.

Keywords: Brazilian green propolis; Artepillin C; B-Raf^{600}E; melanoma; molecular docking; molecular dynamics; natural products; MAPK signaling pathway; selective inhibition; in silico drug discovery

1. Introduction

Melanoma is one of the most aggressive and deadly forms of skin cancer, notorious for its rapid progression and resistance to traditional treatments [1]. Understanding the molecular drivers behind its growth has led to significant breakthroughs in targeted therapies. Among the key players in melanoma development is the enzyme B-Raf, particularly its mutated form known as B-Raf^{600}E [1,2]. This mutation occurs when valine at position 600 is replaced by glutamic acid, causing the enzyme to become permanently active. As a result, the MAPK signaling pathway, which controls cell growth and survival, is continuously turned on, leading to uncontrolled cell division and tumor growth. The discovery of B-Raf^{600}E's role in melanoma was a turning point, highlighting it as a crucial target for drug development [3]. Researchers have been focusing on designing inhibitors that could specifically bind to this mutant form without affecting the non-mutated form of the protein, thus minimizing side effects. Vemurafenib was one of the first successful drugs to achieve this [2]. Although Vemurafenib represented a significant advance in the treatment of melanoma with the B-Raf^{600}E mutation, its long-term efficacy is often hindered by important limitations.

One of the most pressing challenges is the rapid development of resistance. While many patients initially respond favorably, tumors frequently acquire adaptive mechanisms that reactivate the MAPK signaling pathway, the very cascade that Vemurafenib aims to inhibit [4]. Beyond resistance, Vemurafenib is also associated with several adverse effects, particularly dermatological toxicities. Commonly reported reactions include photosensitivity and, in some cases, the development of cutaneous squamous cell carcinomas. Arthralgia is another frequent side effect, which can significantly impact the patient's quality of life and may necessitate dose adjustments or even treatment discontinuation [5]. Given these limitations, ongoing research has increasingly focused on developing next-generation therapies, including more selective BRAF inhibitors and combination regimens, with the goal of enhancing efficacy while minimizing toxicity and delaying the onset of resistance [6]. In this context, natural extracts have also been investigated as complementary or alternative therapeutic strategies for melanoma.

Among those showing promising results is Brazilian green propolis extract, which has demonstrated antitumor, pro-apoptotic, and anti-angiogenic effects in preclinical studies, suggesting potential as an adjunct in melanoma treatment [7–12]. Due to the encouraging outcomes observed with Brazilian green propolis, further studies are warranted to isolate and identify the specific bioactive compounds responsible for these effects. Such investigations could help determine which substances have the greatest therapeutic potential and, importantly, whether they are capable of selectively inhibiting the mutant B-Raf^{600}E, offering a novel and potentially less toxic approach to melanoma treatment. Brazilian green propolis is a resinous substance collected by *Apis mellifera* bees from the buds and exudates of plants, especially *Baccharis dracunculifolia*. This propolis is widely recognized for its biological properties, including antioxidant, antibacterial and antiproliferative activities [13]. Studies have shown that Brazilian green propolis has a distinct chemical composition, rich in prenylated derivatives of *p*-coumaric acid and flavonoids, especially Artepillin C, which is found in high concentration in this variety of propolis [13,14]. In addition to *p*-coumaric acid (**1**) and Artepillin C (**3**), its main chemical marker, several other compounds have been identified in Brazilian green propolis, including drupanin (**2**) [15], culifolin (**4**) [16], ferulic acid (**5**) and caffeic acid (**6**) [17], kaempferol (**7**) and kaempferide (**8**) [18], ermanin (**9**) [19], (*E*)-2-isopropenyl-7-isopentenyl-2,3-dihydrobenzofuran-5-acrylic acid (**10**), 2,2-dimethyl-2*H*-1-benzopyran-6-carboxylic acid (**11**) and (*E*)-3-(2,2-dimethyl-2*H*-1-benzopyran-6-yl) propenoic acid (**12**) [20]. The compounds selected for this study

represent the main chemical constituents of Brazilian green propolis, including both major markers, such as *p*-coumaric acid and Artepillin C, and secondary bioactive compounds. The selection prioritized compounds are representative of the chemical diversity of green propolis, which has a history of description in the literature and potential biological activity, allowing a more comprehensive analysis of their structural correlations and relevant properties for the development of medicinal formulations. In this preliminary study, these compounds were evaluated for their potential as selective inhibitors of the B Raf^{600}E mutant protein using bioinformatics and cheminformatics approaches. The investigation focused on structural and conformational analyses, employing molecular modeling techniques such as molecular docking to predict binding affinities and interaction modes, as well as molecular dynamics simulations to assess the stability and behavior of the ligand-protein complexes over time.

2. Results

2.1. Physicochemical Properties and ADME Studies

The physicochemical properties of twelve phytochemicals present in Brazilian green propolis were calculated theoretically using the SWISSADME platform, with a focus on their possible drug-like characteristics. The molecular properties analyzed included molecular weight (MW), number of heavy atoms, number of aromatic heavy atoms, fraction of sp^3 hybridized carbons (Csp^3 fraction), number of rotational bonds, number of hydrogen bond acceptors (HBAs), number of hydrogen bond donors (HBDs), molar refractivity (MR) and topological polar surface area (TPSA). These parameters are essential for predicting the pharmacokinetic behavior of the compounds, such as solubility, permeability and bioavailability. The results are shown in Table 1.

Table 1. Calculated physicochemical properties of the phytochemicals considered in this study.

Ligand	MF	MW (g/mol)	nHA	nAHA	F.Csp^3	nRB	nHBA	nHBD	MR	TPSA (Å^2)
1	$C_9H_8O_3$	164.16	12	6	0	2	3	2	45.13	57.53
2	$C_{14}H_{16}O_3$	232.28	17	6	0.21	4	3	2	68.85	57.53
3	$C_{19}H_{24}O_3$	300.39	22	6	0.32	6	3	2	92.57	57.53
4	$C_{19}H_{22}O_3$	298.38	22	6	0.32	4	3	1	90.95	46.53
5	$C_{10}H_{10}O_4$	194.18	14	6	0.1	3	4	2	51.63	66.76
6	$C_9H_8O_4$	180.16	13	6	0	2	4	3	47.16	77.76
7	$C_{15}H_{10}O_6$	286.24	21	16	0	1	6	4	76.01	111.13
8	$C_{16}H_{12}O_6$	300.26	22	16	0.06	2	6	3	80.48	100.13
9	$C_{17}H_{14}O_6$	314.29	23	16	0.12	3	6	2	84.95	89.13
10	$C_{19}H_{22}O_3$	298.38	22	6	0.32	6	3	1	90.12	46.53
11	$C_{12}H_{12}O_3$	204.22	15	6	0.25	1	3	1	57.52	46.53
12	$C_{14}H_{14}O_3$	230.26	17	6	0.21	2	3	1	67.23	46.53

MF—Molecular Formula; MW (g/mol)—Molecular weight; nHA—Number of heavy atoms; nAHA—Number of aromatic heavy atoms; F.Csp^3—Fraction of sp^3 hybridized carbon atoms; nRB—Number of rotatable bonds; nHBA—Number of hydrogen bond acceptors; nHBD—Number of hydrogen bond donors; MR—Molar refractivity; TPSA (Å^2)—Topological polar surface area. Ligands—*p*-coumaric acid (**1**), drupanin (**2**), Artepillin C (**3**), culifolin (**4**), ferulic acid (**5**), caffeic acid (**6**), kaempferol (**7**), kaempferide (**8**), ermanin (**9**), (*E*)-2-isopropenyl-7-isopentenyl-2,3-dihydrobenzofuran-5-acrylic acid (**10**), 2,2-dimethyl-2*H*-1-benzopyran-6-carboxylic acid (**11**), and(*E*)-3-(2,2-dimethyl-2*H*-1-benzopyran-6-yl) propenoic acid (**12**).

The pharmacokinetic parameters of the evaluated phytochemicals were assessed to predict their ADME behavior. All ligands demonstrated high gastrointestinal (GI) absorption, indicating favorable oral bioavailability. The blood–brain barrier (BBB) permeability was observed for most compounds, except for ligands **6**, **7**, **8**, and **9**. None of the ligands were identified as substrates for *p*-glycoprotein (P-gp), suggesting a low potential for efflux-mediated drug resistance. The LogP values ranged from 0.93 to 4.27, with the majority falling between **1** and **3**. Notably, ligands **3**, **4**, and **10** had values greater than **4**. Regarding cytochrome P450 enzyme inhibition, ligands **1**, **2**, and **5** showed no inhibitory activity against any of the tested isoenzymes. Ligand **3** inhibited CYP3A4, while ligand **4** inhibited both CYP3A4 and CYP2D6. Ligand **10** inhibited CYP1A2, CYP2D6, and CYP3A4. Ligands **7**, **8**, and **9** inhibited CYP2C19, CYP2C9, CYP2D6, and CYP3A4. These results are summarized in Table 2.

Table 2. Predicted ADME properties of ligands based on in silico pharmacokinetic analysis.

Ligand	LogP	GI	BBB	P-gp Substrate	Cytochrome Inhibitor P450				
					CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4
1	1.26	High	Yes	No	No	No	No	No	No
2	2.83	High	Yes	No	No	No	No	No	No
3	4.27	High	Yes	No	No	Yes	Yes	No	No
4	4.02	High	Yes	No	No	Yes	Yes	Yes	No
5	1.36	High	Yes	No	No	No	No	No	No
6	0.93	High	No	No	No	No	No	No	No
7	1.58	High	No	Yes	No	No	Yes	Yes	No
8	2.00	High	No	Yes	No	No	Yes	Yes	No
9	2.35	High	No	Yes	No	Yes	Yes	Yes	No
10	4.11	High	Yes	Yes	Yes	Yes	No	Yes	Yes
11	2.33	High	Yes	Yes	No	No	No	Yes	No
12	2.67	High	Yes	Yes	Yes	No	No	Yes	Yes

LogP—Measures lipophilicity, affecting membrane permeability; GI—Indicates absorption in the gastrointestinal tract; BBB—Shows ability to cross the blood–brain barrier; P-gp Substrate—Identified by P-glycoprotein, influencing drug transport and distribution. Ligands: *p*-coumaric acid (**1**), drupanin (**2**), Artepillin C (**3**), culifolin (**4**), ferulic acid (**5**), caffeic acid (**6**), kaempferol (**7**), kaempferide (**8**), ermanin (**9**), (*E*)-2-isopropenyl-7-isopentenyl-2,3-dihydrobenzofuran-5-acrylic acid (**10**), 2,2-dimethyl-2*H*-1-benzopyran-6-carboxylic acid (**11**), and (*E*)-3-(2,2-dimethyl-2*H*-1-benzopyran-6-yl) propenoic acid (**12**).

2.2. Molecular Modeling

2.2.1. Geometry Optimization

To obtain the most stable conformations of the major phytochemicals identified in Brazilian green propolis, a hybrid conformational optimization protocol was applied. This approach combined three levels of molecular modeling: initial geometry optimizations were performed using classical molecular mechanics with the Universal Force Field (UFF), followed by semi-empirical calculations using the AM1 method, and finally refined using an ab initio Hartree–Fock (HF/6-31G*) level of theory. This stepwise strategy aimed to maximize accuracy by combining the efficiency of classical methods with the precision of quantum chemical approaches. The optimized 3D structures are presented in Figure 1. Among the selected phytochemicals, compounds **3** and **10** exhibited the highest degree of conformational flexibility, each containing six rotatable bonds. These structural features indicate a greater potential for dynamic adaptation during molecular recognition. The

conformational planarity of compound **7** (a flavonol with hydroxyl groups at positions 3, 5, 7, and 4′) was also notable, suggesting enhanced intramolecular conjugation. In contrast, methoxylated derivatives, such as compounds **8** and **9**, showed a decreased planarity and increased hydrophobicity due to substitutions on the B and C rings.

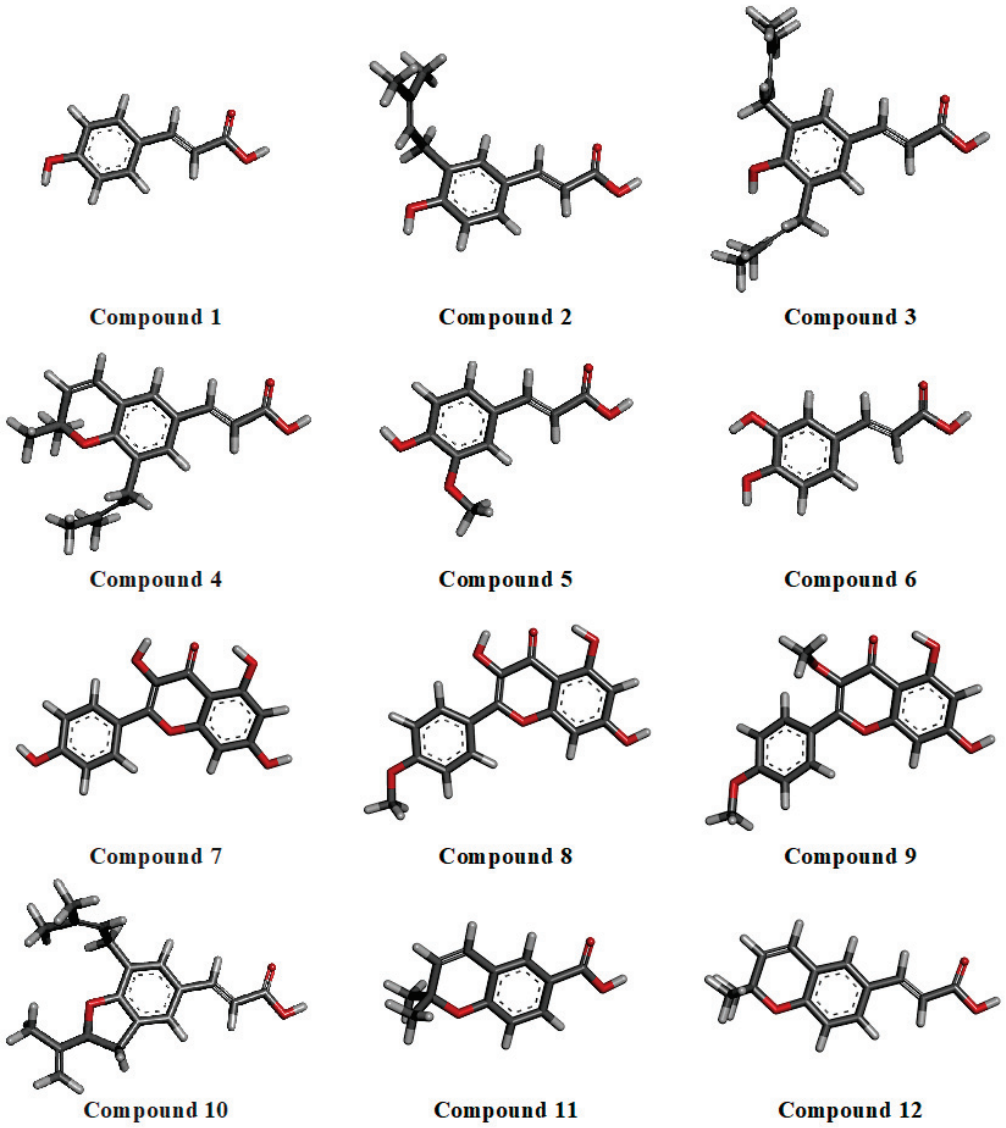


Figure 1. Optimized 3D structures of major phytochemicals found in Brazilian green propolis.

2.2.2. HOMO and LUMO Orbitals

The highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals, known as frontier orbitals, are important for understanding the electrical and optical properties and reactivity of molecules. The energy difference between HOMO and LUMO, known as the energy gap, is particularly important as it directly influences the chemical reactivity of the molecule. Compounds with smaller energy gaps tend to be more reactive, which can facilitate interaction with the biological target, while those with larger gaps have greater chemical stability, allowing for more controlled interactions. In addition, frontier orbitals are involved in electron transfer processes, such as oxidation and reduction reactions, which are often essential for the biological activity of active molecules [21]. The graphical depiction of the HOMO and LUMO orbitals for the compounds studied is presented in Figure 2.

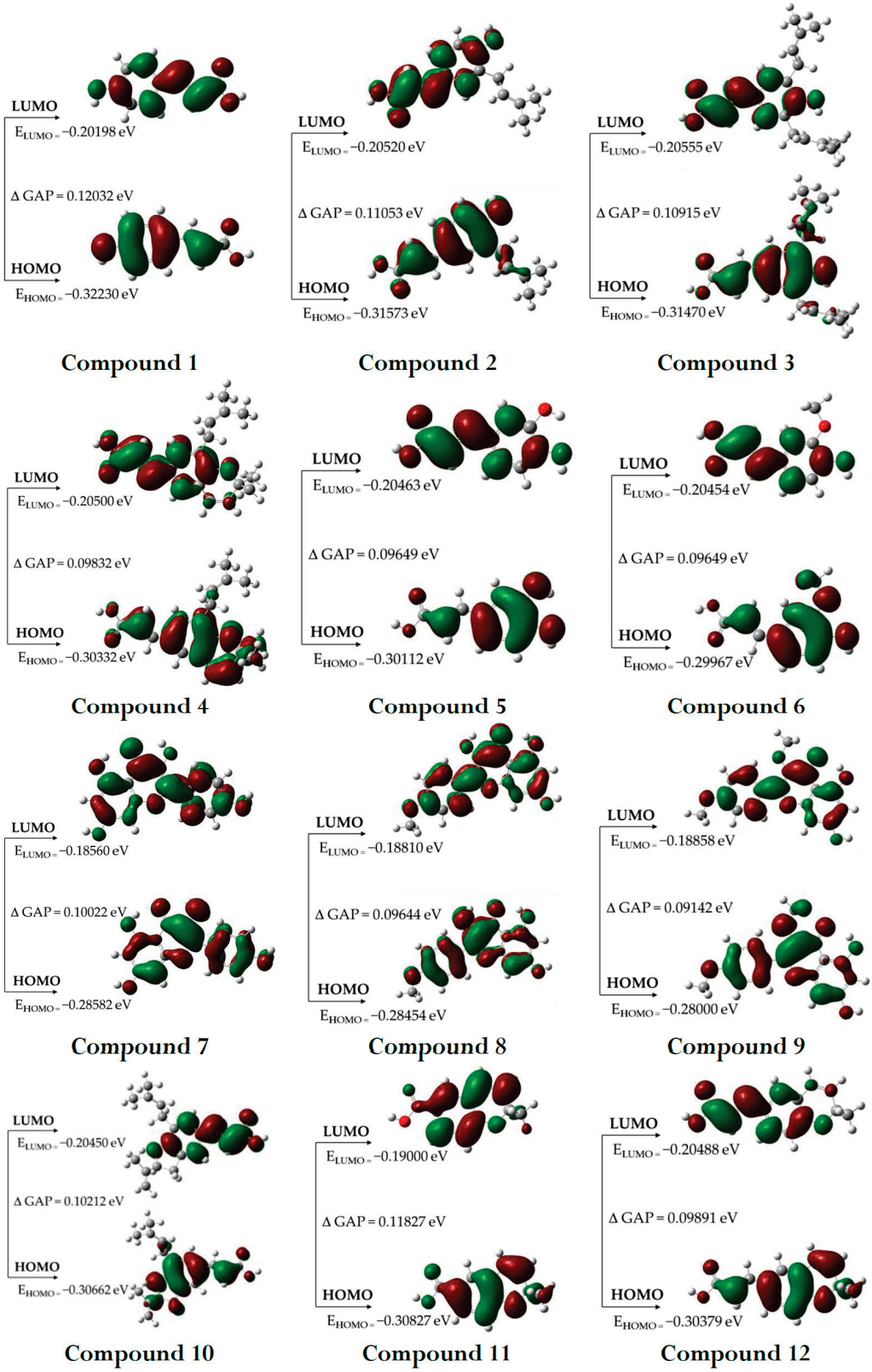


Figure 2. Frontier molecular orbitals (HOMO and LUMO) of the selected compounds.

2.2.3. Electrostatic Potential Map (EPM)

The electrostatic potential map (EPM) (Figure 3) is a valuable tool for understanding how molecules interact with biological targets. It visualizes the distribution of electrostatic

charges across the molecular surface, allowing for the identification of regions most likely to engage in interactions that influence binding affinity and selectivity. This information is crucial for optimizing the therapeutic potential of bioactive compounds. Electron-rich (highly electronegative) regions often act as hydrogen bond acceptors, while electron-deficient (electropositive) regions typically function as hydrogen bond donors [22]. Beyond predicting molecular interactions, EPMs also contribute to evaluating pharmacokinetic properties such as solubility and membrane permeability. While highly polar regions can enhance aqueous solubility, they may simultaneously hinder passive diffusion through biological membranes, ultimately affecting bioavailability. Furthermore, mapping the electrostatic profile of a molecule provides insights into its metabolic stability by highlighting areas susceptible to enzymatic transformations, factors that can impact efficacy or lead to the formation of toxic metabolites [23].

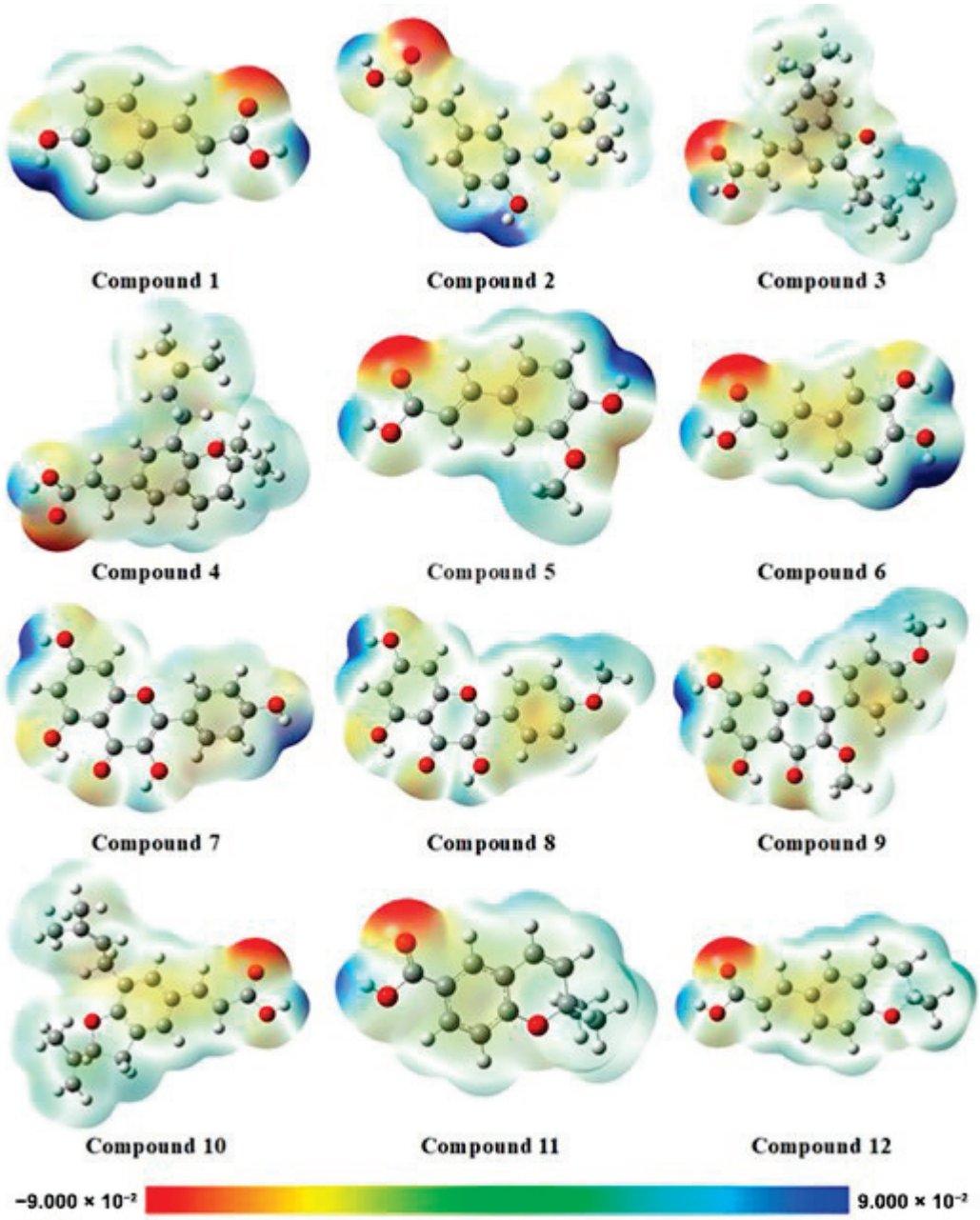


Figure 3. Electrostatic potential maps of the studied compounds highlighting reactive and interaction-prone regions.

2.3. Molecular Docking

The B-Raf^{600}E enzyme is a key regulator of the MAPK signaling pathway, playing a central role in controlling cell growth and survival. Its structure is finely tuned to support its function as a serine/threonine kinase. The enzyme comprises two principal lobes: a smaller N-terminal lobe, composed mainly of β-sheets, and a larger C-terminal lobe, rich in α-helices. These lobes are connected by an ATP-binding cleft—the catalytic core where phosphorylation of substrate proteins occurs. This region is especially critical, as it positions ATP for phosphate transfer during kinase activity. In the B-Raf^{600}E mutant, a single amino acid substitution—valine to glutamic acid at position 600—has significant structural and functional consequences. Valine is a small, nonpolar residue, while glutamic acid is bulkier and carries a negative charge. This alteration destabilizes the inactive conformation of the kinase, favoring a constitutively active state even in the absence of upstream signals. The result is continuous MAPK pathway activation, driving uncontrolled cell proliferation and contributing to melanoma development. Structural analysis, particularly based on the 3OG7 crystal structure, reveals how this mutation reconfigures the ATP-binding pocket, creating new opportunities for selective drug binding. Vemurafenib, a potent B-Raf^{600}E inhibitor, binds precisely within this pocket, forming a network of hydrogen bonds and hydrophobic interactions. It locks the enzyme in an inactive conformation, effectively preventing ATP binding and blocking downstream signaling.

The Glu600 residue, in combination with adjacent positively charged residues, helps shape a unique hydrophobic pocket that Vemurafenib exploits, explaining its high selectivity for the mutant over the wild-type enzyme—thus minimizing off-target effects. These structural insights have been instrumental in the development of targeted melanoma therapies. However, despite the initial success of B-Raf inhibitors, resistance often emerges through secondary mutations or alternative activation of the MAPK pathway. To address this challenge, combination therapies—such as co-administering B-Raf and MEK inhibitors—have been investigated to enhance efficacy and delay resistance. In this context, a molecular docking approach was employed to investigate the potential interactions between phytochemical ligands from Brazilian green propolis and the active site of the B-Raf^{600}E mutant. The docking results, which provide insights into binding affinities and interaction profiles, are summarized in Table 3.

Table 3. Physicochemical properties of the phytochemicals considered in this study.

Ligand	Binding Energy (kcal/mol)	Hydrogen Interacting Residue of Target Along with Their Bond Length	Hydrophobic Interacting Residue of Target Along with Their Bond Length
1	−5.11	CYS532 (2.01), GLY534 (1.88), GLN530 (1.72)	TRP531, CYS532
2	−7.02	CYS532 (1.76), ILE527 (2.62), GLN530 (2.05)	THR529, PHE595, LEU505, LEU505, LEU514, ALA481, LYS483
3	−8.17	LYS483 (2.09), CYS532 (2.07), GLN530 (2.02)	VAL471, ILE463, VAL471, LEU505, LEU505, LEU514, PHE468, PHE595, VAL471
4	−7.57	CYS532 (1.85)	CYS532, PHE583, VAL471, ALA481, ALA481, LEU514, VAL471, LYS483, VAL471, LYS483, VAL471, PHE468, PHE583, PHE583, VAL471, ALA481
5	−5.51	PHE595 (2.94)	GLY596, ALA481, VAL471, VAL471, ALA481, LYS483

Table 3. *Cont.*

Ligand	Binding Energy (kcal/mol)	Hydrogen Interacting Residue of Target Along with Their Bond Length	Hydrophobic Interacting Residue of Target Along with Their Bond Length
6	−5.41	LYS483 (2.62), PHE595 (2.59), GLY596 (2.13), ALA481 (2.21), ILE527 (2.04)	VAL471, ALA481, LYS483
7	−7.15	THR529 (2.59), CYS532 (1.77), ILE527 (2.04), GLN530 (2.03), CYS532 (2.29)	TRP531, TRP531, PHE583, PHE583, VAL471, ALA481, CYS532, ILE463, VAL471, VAL471, ALA481, LYS483
8	−7.38	LYS483 (2.34), CYS532 (1.69), GLN530 (1.98), CYS532 (2.23)	TRP531, TRP531, PHE583, PHE583, VAL471, ALA481, CYS532, ILE463, LYS483, LEU514
9	−7.29	CYS532 (1.71), GLN530 (2.04)	ASP594, LYS483, TRP531, LEU505, VAL471, ALA481, VAL471, ALA481, CYS532, LYS483
10	−7.74	LYS483 (1.73), PHE595 (2.99), GLY596 (1.96)	VAL471, ALA481, ALA481, LEU514, CYS532, VAL471, TRP531, TRP531, TRP531, PHE583, VAL471, ALA481, LYS483, LEU514
11	−6.74	LYS483 (1.80), ASP594 (2.22), PHE595 (2.99), GLY596 (2.24)	LYS483, ALA481, ALA481, ALA481, LEU514, LEU514, VAL471, PHE583, LYS483, LEU514
12	−8.49	LYS483 (1.70), ASP594 (2.29), PHE595 (2.99), GLY596 (2.21), PHE583 (2.85)	VAL471, ALA481, ALA481, LEU514, CYS532, VAL471, VAL471, ILE463, VAL471, PHE468, TRP531, TRP531, PHE583, PHE583, VAL471, LEU514

Ligands: *p*-coumaric acid (**1**), drupanin (**2**), Artepillin C (**3**), culifolin (**4**), ferulic acid (**5**), caffeic acid (**6**), kaempferol (**7**), kaempferide (**8**), ermanin (**9**), (E)-2-isopropenyl-7-isopentenyl-2,3-dihydrobenzofuran-5-acrylic acid (**10**), 2,2-dimethyl-2H-1-benzopyran-6-carboxylic acid (**11**), and (E)-3-(2,2-dimethyl-2H-1-benzopyran-6-yl) propenoic acid (**12**).

2.4. Molecular Dynamics and Free Energy of Binding

Molecular dynamics (MD) simulations were performed for 500 nanoseconds (ns), starting from the most favorable ligand-enzyme complexes identified through molecular docking. To enable comparative analysis, simulations of the free enzyme and Vemurafenib were also conducted under the same theoretical conditions described in the "Molecular Dynamics" section compound **3**. Artepillin C remained stably bound in the active site throughout the entire 500 ns simulation. Moreover, the protein exhibited enhanced structural stability from 200 ns onward, with the Root Mean Square Deviation (RMSD) stabilizing around 0.05 nm. The radius of the gyration profile also revealed a consistent compaction pattern, which means that the structural integrity of the protein remained preserved throughout the whole simulation, as shown in Figure 4. In the absence of the ligand (Figure 4, red lines), the enzyme appeared more compact until 300 ns but less stable, also expanding from 300 ns onwards. This suggests that Artepillin C contributes to the structural stabilization of the enzyme. The same patterns seen for Artepillin C are observed on Vemurafenib MD. The radius of gyration starts with contraction followed by relaxation from 300 ns onwards. On the other hand, for RMSD, Vemurafenib presents a less stabilizing capability than Artepillin C, with a delta of 1 nm from 200 ns to the end of the simulation. This variability is closer to the free enzyme RMSD variation, meaning Artepillin C is a better stabilizer of the protein.

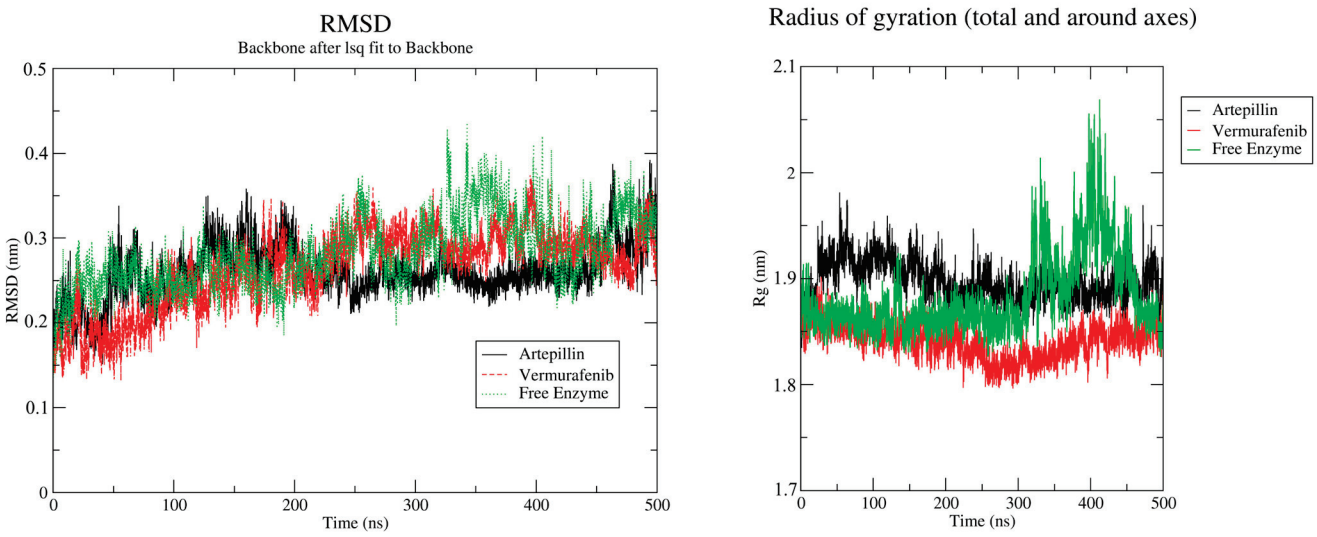


Figure 4. Radius of gyration and RMSD behavior along the molecular dynamic simulation time in ns. Red lines—Vemurafenib-enzyme complex simulation; Black lines—Artepillin C-enzyme complex simulation; Green lines—free enzyme simulation.

The Artepillin C hydrogen bond (hbond) analysis revealed that the complex was maintained by two persistent hbonds throughout the simulation, occasionally reaching 3 hbonds. Vemurafenib yet presents a crescent number of hbond interactions along the simulation, preserving seven hbonds from 200 ns onward. Additionally, analysis of the most representative pose from the simulation confirmed that Artepillin C remained bound to the ATP-binding site even after 300 ns (Figure 5).

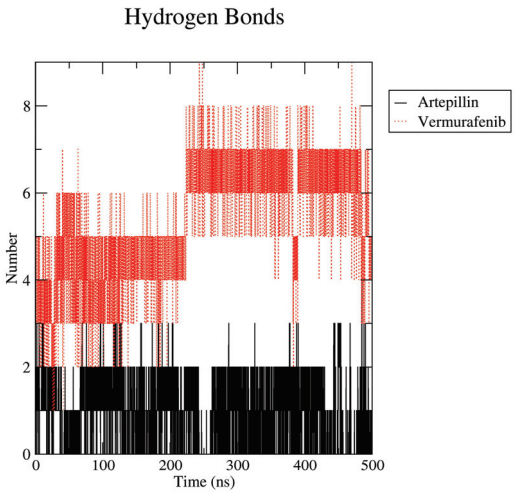


Figure 5. Hydrogen bond pattern along the molecular dynamic's simulation.

Compared to the crystal structure, the cavity appeared more open due to system relaxation, yet the compound occupied regions associated with Vemurafenib's binding mode, especially near PHE583 (Figure 6). Interaction analysis from the most representative pose in the clusters revealed two key hydrogen bonds with ASP594 and a critical ion–ion interaction between the carboxylate group of Artepillin C and LYS473 of B-Raf (Figure 7A,C). Hydrophobic interactions were also observed, involving Artepillin's prenyl branches and apolar residues such as LEU514, VAL471, ALA481, LYS483 and ILE463. Other surrounding residues also contributed via potential van der Waals interactions, as repre-

sented in Figure 7C. The hydrophobic surface map confirmed that the prenyl branches oriented toward hydrophobic regions of the pocket (Figure 7B), while the carboxylate group projected outward to interact with polar regions and solvent-accessible water molecules. The hydroxyl group was located deeper within the binding site, interacting with ASP594.

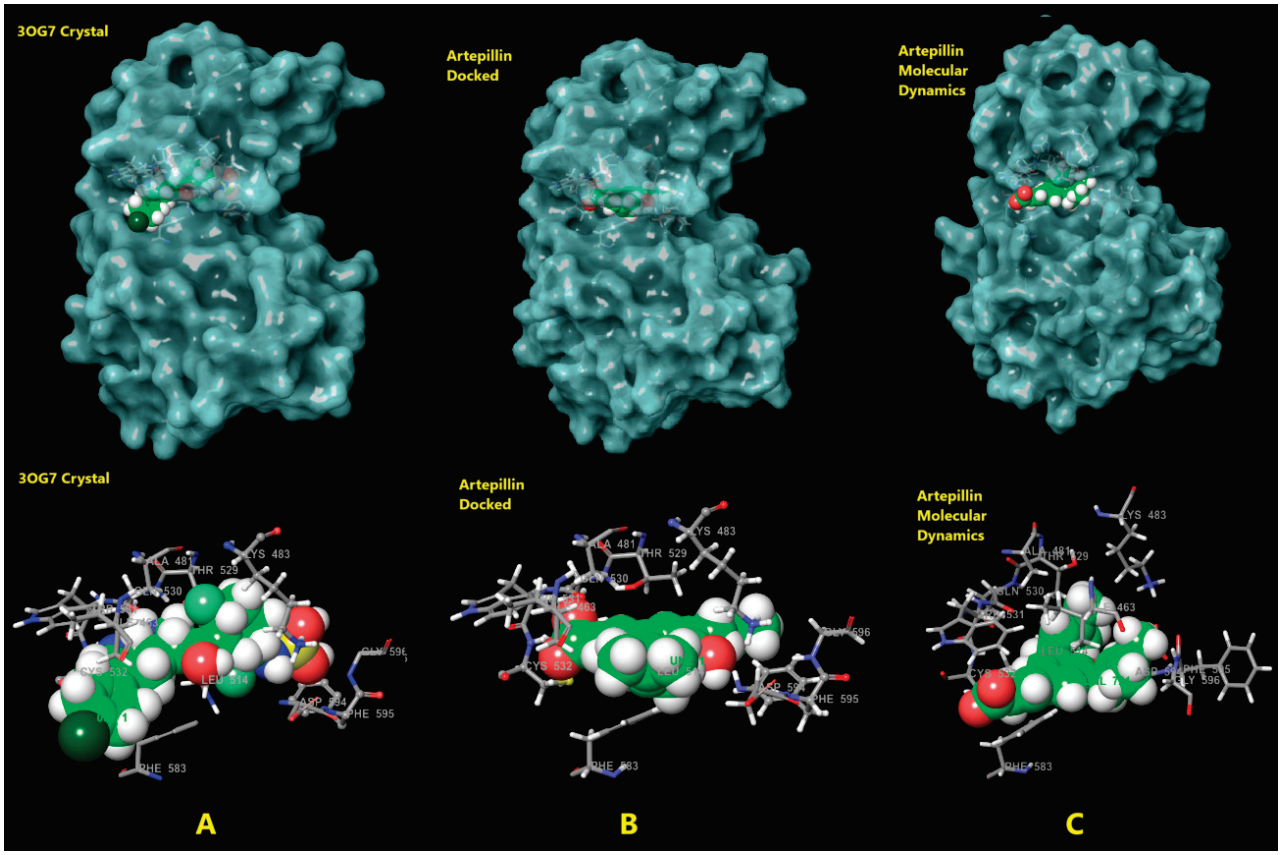


Figure 6. Molecular representation of the complexes and a zoomed-in view highlighting the closest residues of the protein. (**A**) Crystal structure from the input obtained in the presence of the ligand Vemurafenib; (**B**) Best-docked pose obtained to compound **3**; (**C**) Most representative pose obtained to Artepillin after molecular dynamics simulation.

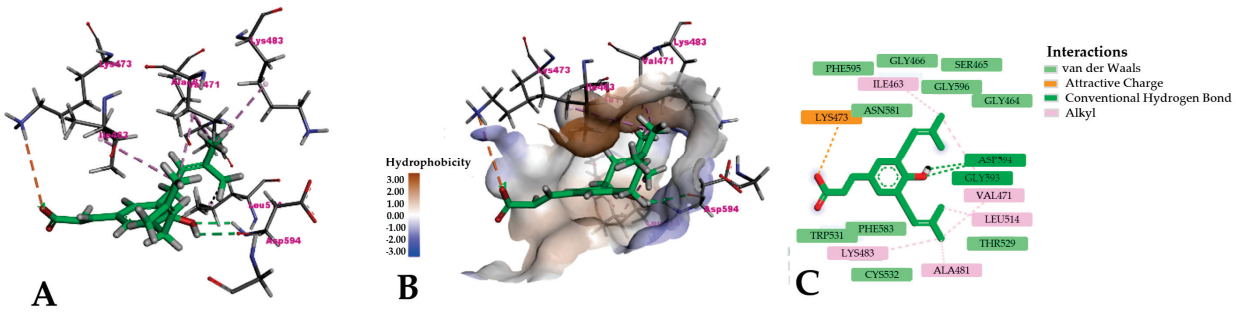
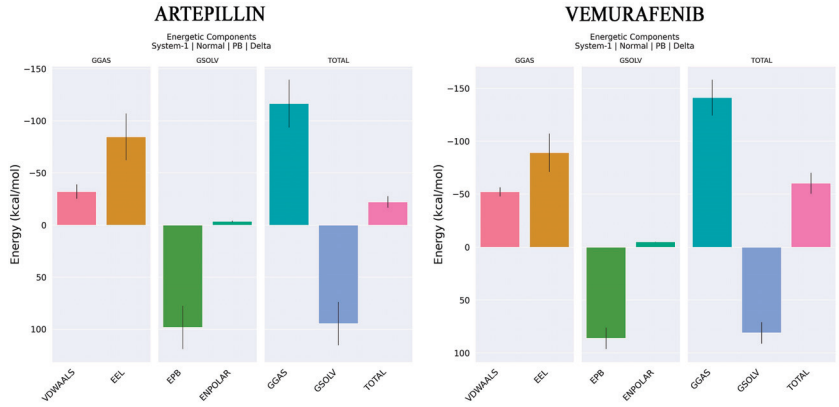


Figure 7. Interaction pattern of Artepillin C at the binding site. (**A**) Tridimensional view of the interactions; (**B**) Interactions located inside the hydrophobic surface map; (**C**) Two-dimensional view of the interactions.

As shown in Table 4, the major energy contributions for both Artepillin C and Vemurafenib come from electrostatic (EEL) and Van der Waals (VDW) components. However, ENPOLAR, which is also associated with hydrophobic interaction along with VDW, provides a minor contribution to enthalpic energy in both complexes. These data reinforce what

was already observed during interaction analysis, i.e., for Artepillin C, the hydrophobic and VDW interactions observed along MD for ILE463, VAL471 and ALA481 are confirmed. Additionally, ASP594 is identified as an unfavorable interaction due to a desolvation penalty. For Vemurafenib, CYS532, GLN530, VAL471, ALA481, PHE583, PHE595 present modest contributions.

Table 4. Free energy of biding from both Artepillin C and Vemurafenib.



ΔE (VDW)	−32.03	−52.13
ΔE (EEL)	−84.62	−89.14
ΔE (EPB)	98.26	86.12
ΔE (ENPOLAR)	−3.65	−5.1
GGAS	−116.65	−141.27
GSOLV	94.6	81.02
ΔH (total)	−22.05	−60.25
−TΔS	1.28	42.70
ΔG	−20.77	−17.55

The components are defined as ΔE (VDW), the energy from van der Waals interactions, and ΔE (EEL), the energy from electrostatic interactions. ΔE (EPB) is the energy from polar solvation, and ΔE (ENPOLAR) is the repulsive component of the nonpolar solvation energy. The total free energy in the gas phase, GGAS, is calculated as GGAS = ΔEVDW + ΔEEL. The total solvation-free energy, GSOLV, is determined as GSOLV = ΔEPB + ΔENPOLAR. Other terms used are ΔH for enthalpy of binding; −TΔS for conformational entropy loss contribution upon ligand binding; and ΔG (binding), represented as ΔH–TΔS.

Furthermore, PHE595 also contributes unfavorable interaction due to desolvation penalty (Figure 8). Regarding affinities, Vemurafenib presents a lower ΔH of −60.25 kcal/mol than Artepillin C, which has a ΔH of −22.05 kcal/mol. This difference can be associated with the quantity of interactions observed for Vemurafenib, as shown in Figures 5 and 8. With respect to the entropy contribution, Vemurafenib displays a higher value. Although this entropic loss can be interpreted as thermodynamically favorable due to the associated gain in disorder, its resulting free energy is higher than that of Artepillin C. This difference suggests that the overall thermodynamic stability of Vemurafenib is lower, making Artepillin C more spontaneously favorable.

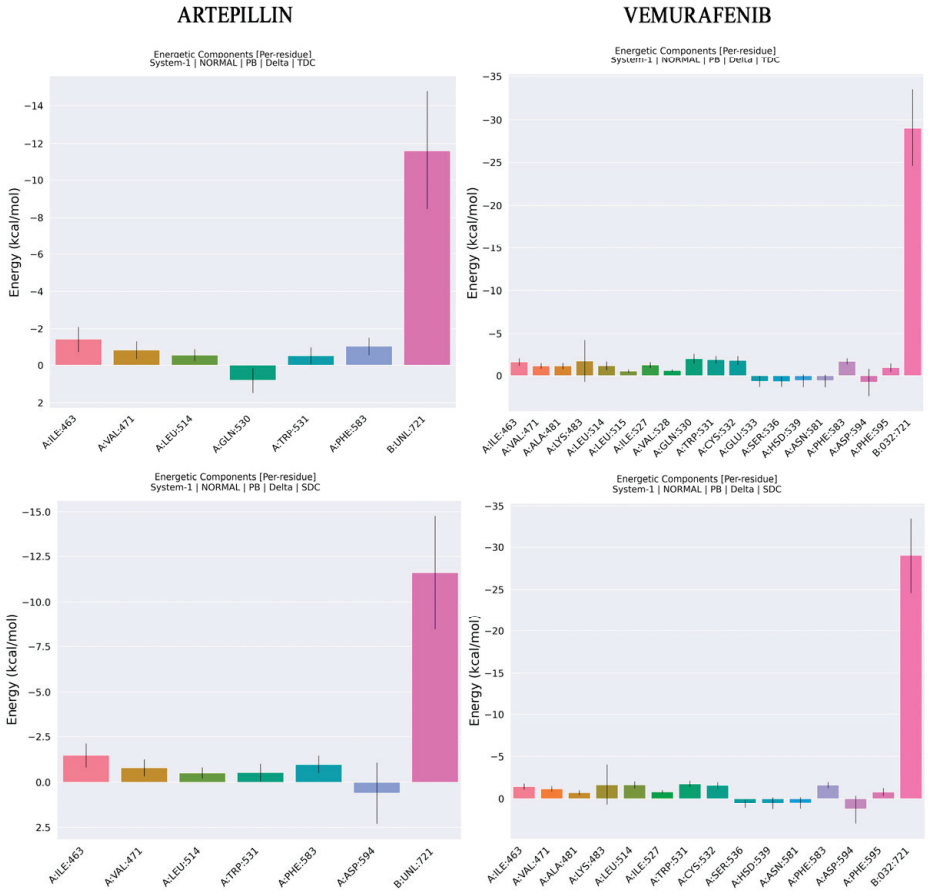


Figure 8. Decomposition interaction analysis for Artepillin C and Vemurafenib. TDC (Total decomposition), total interaction energy contribution for each residue. SDC (Sidechain decomposition) is the only sidechain contribution for interaction energy.

3. Discussion

3.1. Physicochemical Properties and ADME Studies

The molecular weights (MW) of the studied compounds ranged from 164.16 to 314.29 g/mol, with ligand **1** having the lowest MW and ligand **9** the highest. All compounds fall within the acceptable range for drug-like molecules (<500 g/mol), suggesting favorable pharmacokinetic properties. Aromatic heavy atoms were prominent across the series, varying from 6 to 16, contributing to planar and rigid structures. The fraction of sp^3 hybridized carbons (Fraction Csp^3), which indicates molecular complexity and three-dimensionality, varied among the compounds. Ligands **1** and **6** exhibited a fraction Csp^3 of 0, indicating completely planar structures, while compound **3** showed the highest value (0.32), suggesting increased structural diversity. Hydrogen bonding capacity, which affects solubility and permeability, was evaluated through hydrogen bond acceptors (HBAs) and donors (HBDs). The HBA values ranged from 3 to 6, while HBDs varied from **1** to **4.** Molecules **7**, **8**, and **9** showed the highest numbers of both HBAs and HBDs, consistent with their polyphenolic structures. These molecules also exhibited the highest topological polar surface area (TPSA), with ligand **7** having the maximum value of 111.13 $Å^2$, increasing solubility but may limit passive membrane permeability.

Molar refractivity (MR) ranged from 45.13 in ligand **1** to 92.57 in ligand **3**, correlating with molecular size and complexity. The extensive conjugation present in these molecules suggests a strong potential for interactions with aromatic residues in protein targets, potentially enhancing binding affinity. Overall, these findings provide a detailed overview

of the physicochemical properties influencing the drug-likeness of the studied molecules. The balance between molecular weight, aromaticity, hydrogen bonding capacity, and structural flexibility suggests promising potential for these compounds as medicinal candidates. Molecules with moderate MW, balanced TPSA, and optimal Fraction Csp^3, like ligand **3**, exhibit the most favorable profile for further development.

The high GI absorption values observed across all compounds support their potential as orally active agents. Most ligands' ability to cross the BBB suggests possible central nervous system activity. The lack of P-gp substrate characteristics indicates a lower risk of drug resistance via efflux mechanisms. The range of LogP values suggests generally favorable mem-brane permeability, while compounds with higher values may benefit from increased BBB penetration. The distinct CYP inhibition profiles highlight varying degrees of potential metabolic interaction. Compounds with minimal or no CYP inhibition, such as Ligands **1**, **2**, and **5**, are less likely to participate in drug-drug interactions. In contrast, Ligands **7**, **8**, and **9** exhibited broad-spectrum CYP inhibition, which may influence the metabolism of co-administered drugs and necessitate caution in further development. These findings underscore the importance of early pharmacokinetic screening in guiding compound selection for drug discovery.

3.2. Molecular Modeling

3.2.1. Geometry Optimization

The main phytochemicals already identified in Brazilian green propolis are derivatives of simple or prenylated cinnamic acids and flavonoids. Cinnamic acids have a basic structure composed of an aromatic ring linked to an unsaturated aliphatic chain (CH=CH–COOH). The presence of the double bond in the side chain (CH=CH) introduces geometric isomerism (E/Z), with the E isomer generally being more stable due to the lower steric repulsion between the substituent groups. Although Z isomers are less common, they can be more reactive due to the functional proximity between the groups [24]. The single bond between the aromatic ring and the vinyl group (CH=CH–), as well as the C–C bond between the side chain and the carboxylic acid (–COOH), are rotatable, creating different conformations that influence molecular interactions, such as the formation of hydrogen bridges or π–π interactions [25]. In addition, these bonds can generate intramolecular interactions, such as hydrogen bonds, which alter the conformational stability and polarity of the compound [26]. In addition to Lipinski's Rule of Five [27], other criteria, such as Veber's Rule [28] and Muegge's Rule [29] suggest that small bioactive molecules should not exceed 10 and 15 rotatable bonds, respectively, to ensure favorable pharmacokinetic properties.

It is well known that compound **3**, the majority compound present in Brazilian green propolis, is one of the most active compounds ever reported [30–32]. However, it's reported in the literature as an unstable compound, and its instability is mainly attributed to the reactivity of the prenyl group and phenol with some bioactive properties, such as antioxidant activity or the ability to interact with biological targets [33–35]. The conformational flexibility of flavonoids is directly related to the characteristics of each ring and the interactions between them, and that determines the overall conformation of the molecule, an essential characteristic for its biological activity [36,37]. The presence of hydroxyl groups also increases the ability to form hydrogen bonds, both intra- and intermolecular, which contributes to structural stabilization [38]. In addition, methoxylation can cause a small increase in the angle between the A and B rings, slightly decreasing the planarity in relation to compound **7** [39]. Double methoxylation confers greater hydrophobicity, reduces polarity and the ability to form hydrogen bonds, as well as increases volume and conformational flexibility around the C ring [40].

3.2.2. HOMO and LUMO Orbitals

The HOMO-LUMO gap values obtained for the Brazilian green propolis phytochemicals considered for this study range from 0.09142 to 0.12032 eV, reflecting the electronic diversity of these natural compounds. Their moderate ΔE values suggest an ideal balance between reactivity and stability, allowing efficient interactions with biomolecular targets such as bacterial cell walls and inflammatory mediators. The HOMO-LUMO gap values observed follow the general trends reported for bioactive polyphenols, flavonoids and prenylated phenolics, reinforcing the medicinal potential of the phytochemicals in Brazilian green propolis. Compounds such as compound **9** (0.09142 eV), compound **8** (0.09644 eV), and compound **6** (0.09649 eV) showed the lowest ΔGAP HOMO-LUMO values, suggesting greater electronic reactivity. This characteristic is often associated with high antioxidant potential, as these molecules can donate electrons to neutralize reactive oxygen species (ROS), recognized for their antioxidant and cytoprotective properties [41–44].

3.2.3. Electrostatic Potential Map (EPM)

Analysis of the electrostatic potential maps revealed that all cinnamic acid derivatives exhibit a pronounced negative electronic charge on the carbonyl oxygen of the carboxylic acid group. This high electron density suggests a strong capacity for hydrogen bond acceptance, which could contribute to their known bioactivities, such as antioxidant and anti-inflammatory effects. The localization of the lowest electronic density on the phenolic group indicates potential sites for electrophilic attack, which may play a role in their radical scavenging activity. These electronic characteristics align with previous studies [45] that link the antioxidant properties of cinnamic acids to the electron-withdrawing nature of the carboxyl group, which stabilizes phenoxyl radicals formed during redox reactions. In contrast, the flavonoids **7**, **8** and **9** did not exhibit prominent regions of negative electrostatic potential. Instead, the low electronic density was localized on the phenolic group in the A ring. This suggests a distinct electronic distribution compared to the cinnamic acids, potentially influencing their interaction with biological targets. The absence of highly negative regions may reduce the strength of hydrogen-bonding interactions compared to the cinnamic acid derivatives. However, the conjugated π-electron system in the flavonoid structures could enhance π–π stacking interactions with proteins or nucleic acids, contributing to their broad spectrum of biological activities, including anti-inflammatory and anticancer properties [46]. Overall, the observed electrostatic potential patterns provide a molecular basis for understanding the diverse biological activities of these compounds. The distinct electronic characteristics of cinnamic acids and flavonoids could be correlated with their different mechanistic pathways in antioxidant, anti-inflammatory, and anticancer activities. Subsequently, our studies involving molecular docking and dynamic simulations further demonstrate the interactions between these compounds and their respective biological targets, B-RAF600, providing a more comprehensive understanding of their pharmacological potential.

3.3. Molecular Docking

The molecular docking simulation of twelve chemical compounds in the B-Raf^{600}E enzyme revealed intriguing interactions within the active site. Critical amino acids in the active site, including GLN530, CYS532, PHE595, and GLY596, play significant roles in ligand binding and catalytic activity. The compounds exhibited varied calculated binding energies and interaction profiles, shedding light on their potential as inhibitors. Compound 3 also demonstrated a significant binding affinity of −8.17 kcal/mol, interacting through hydrogen bonds with LYS483, CYS532, and GLN530. These interactions occur near key

catalytic residues, possibly hindering ATP binding. Notably, hydrophobic contacts with VAL471, LEU505, and PHE595 suggest a stable docking pose within the active site. The binding pattern reflects a strategic occupation of the hydrophobic pocket, indicative of selective inhibition potential. This compound's interaction profile aligns with its previously reported bioactivity, supporting its continued investigation as a therapeutic agent. Other compounds, such as compound 10 and compound 4, also displayed favorable binding energies of −7.74 and −7.57 kcal/mol, respectively, but much lower than the ones observed with compound 3 and compound 12. Both compounds interacted with CYS532 and PHE595 as well as through hydrophobic interactions, particularly with residues like VAL471 and ALA481. Compound 2, compound 7, compound 8, and compound 9 presented moderate binding affinities ranging from −7.02 to −7.38 kcal/mol. Their interactions were characterized by hydrogen bonds with CYS532 and GLN530, along with hydrophobic contacts involving TRP531 and PHE583. These compounds consistently engaged with critical residues within the ATP-binding pocket; however, their binding energies, although favorable, were lower compared to compound 3 and compound 12, suggesting they could not be good ligands of the target.

Conversely, cinnamic acid derivatives like compounds 1, 5, and 6 showed the lowest binding energies, ranging from −5.11 to −5.51 kcal/mol. Although they formed hydrogen bonds with essential residues such as CYS532 and GLN530, their hydrophobic interactions were limited. This lack of extensive hydrophobic contact might contribute to their reduced binding affinities. Their weaker interaction profiles suggest a lower potential for effective kinase inhibition. The results of the molecular docking simulation highlight several positive aspects. However, some limitations must be considered. Compounds with lower binding energies, despite forming hydrogen bonds, lacked extensive hydrophobic interactions, potentially reducing their stability within the active site. Additionally, the simulation does not account for dynamic conformational changes in the kinase, which may influence binding efficiency. Therefore, molecular dynamics simulations and in vitro kinase assays are necessary to validate these findings.

3.4. Molecular Dynamics and Free Energy

The molecular dynamics simulations demonstrated that Artepillin C maintained a stable and specific interaction with the enzyme B-Raf^{600}E over a 500 ns timescale. The stabilization of the Artepillin-B-Raf complex, particularly from 200 ns onward, suggests favorable conformational accommodation and binding. The consistent RMSD and radius of gyration values indicate that the enzyme structure remains compact and stable in the presence of Artepillin C, while the free enzyme fluctuated more significantly within the same timeframe, reinforcing the hypothesis that ligand binding enhances structural stability while preserving structural integrity, a behavior also observed for Vemurafenib. The persistence of hydrogen bonds and the prominent ion-ion interaction with LYS473 suggest that electrostatic forces play a significant role in complex stability. Notably, this interaction appears to be a distinguishing feature between the docking and the most representative dynamics pose, with a repositioning of the interaction site from LYS483 (in docking) to LYS473 (in dynamics), possibly due to structural relaxation and optimization of the binding conformation. Artepillin C also reproduced some of the key interactions observed for Vemurafenib, including hydrogen bonds with ASP594 and a network of hydrophobic contacts within the ATP-binding pocket. While Vemurafenib relies on a π-cation interaction with LYS483, Artepillin's stronger ionic interaction with LYS473 may provide an alternative mode of stabilization. Although LYS473 is not present in decomposition analysis as a contributing residue, it does not mean this interaction is not happening, as can be seen for

the representative pose from the cluster analysis. Moreover, this strong interaction is not possible for Vemurafenib in its binding mode. Artepillin C also presents the slightest free energy of binding compared to that of Vemurafenib, mostly due to entropic loss differences. These free energy of binding interactions, combined with favorable docking energy and dynamic stability, support the hypothesis that artepillin C is a promising candidate for further investigation as a selective B-Raf^{600}E inhibitor.

4. Materials and Methods

4.1. Physicochemical Properties and ADME Studies

The 2D structures of the phytochemicals selected for this study were designed using ChemDraw Ultra 8.0. The resulting compounds were then processed to obtain their SMILES representations, which were subsequently analyzed using the SwissADME Web tool (https://www.swissadme.ch/; accessed on 23 December 2024) [47] to evaluate their pharmacological and toxicological properties in silico. The data generated were used for further analysis.

4.2. Electronic Properties Calculations

The initial 2D structures were converted to their respective 3D projections, thus attributing the intra- and intermolecular interactions. Then, geometry optimization was performed using a molecular mechanics (UFF) force field using Avogadro 1.2.0v software [48]. Partial atomic charges were calculated employing Austin Model 1 (AM1) semi-empirical methodology, implemented in MOPAC2016 [49], and subsequently, the geometry optimization and electrostatic partial atomic charges (CHELPG) were computed using the ab initio method HF/6-31G* [49,50] implemented at ORCA 5.0.3 software [51]. To visualize the electronic properties of the molecules, including the frontier molecular orbitals (HOMO and LUMO) and the molecular electrostatic potential (MEP), a conversion process was necessary, as ORCA's native output formats are not directly compatible with GaussView. After completion of the ORCA calculations, the binary wavefunction file (.gbw) was processed using the auxiliary tool (orca_2mkl), which generated a Molden input file (.molden.input). This file was then opened in the Molden software, where the molecular orbitals and electron density data could be further examined and exported in a format compatible with Gaussian [52].

4.3. Preparation of Target Protein

The crystal structure coordinates of human B-Raf Kinase V600E oncogenic mutant (BRAF V600E) (PDB ID: 3OG7) in the bound state with inhibitor Vemurafenib® determined by X-ray diffraction were obtained from the Protein Data Bank [53] in PDB format. AutoDockTools 1.5.6 [54] were used for the preparation of the target. The protein structure was prepared for docking analysis by removing the compound and water molecules, attached compounds and ions. Polar hydrogen atoms and Kollman charges were added to the structure. Chain A of BRAF V600E was selected for docking study as the inhibitor Vemurafenib® is in chain A [55].

4.4. Molecular Docking

The three-dimensional geometry models of B-Raf Kinase V600E Oncogenic Mutant were retrieved from the Protein Data Bank (PDB) (PDB ID: 3OG7, resolution 2.45 Å) and used as geometry reference to perform the redocking validation test. The target was prepared by the removal of heteroatoms (water, ions, and ligands) and by checking for potential structural errors, mainly those located at the recognition sites to be explored in

this study. Molecular docking was performed employing AutoDock Tools-1.5.6 software. The compounds were then redocked considering a grid box with dimensions 40 × 40 × 40 and centered at the coordinates X = −1.385, Y = −12.941 and Z = −18.992. The redocking protocol was validated using the same software and target, employing Vemurafenib as the reference ligand. Fifty runs with ten resulting poses were generated and analyzed. The parameters of the Genetic Algorithm were kept as default in both studies, and the docking resulting poses were ordered according to their score values and analyzed using the freeware visualization program Discovery Studio Visualizer. All possible interactions and the steric complementarity obtained were analyzed considering a maximum distance of 3.0 Å to hydrogen bond interactions as well as minimum angles for hydrogen bond donors of 120 °C and for hydrogen bond acceptors of 90 °C. Interactions involving electronic π systems such as π-cation, π-stacking dipole-dipole and ion-dipole interactions were also considered, according to pre-settled criteria in the Discovery Studio Visualizer program.

4.5. Molecular Dynamics

Molecular dynamics studies were performed employing the best docking pose of Compound 3 into the target as the input file. The complex was prepared using CHARMM-GUI [56] web-based software considering a pH of 7.4 to calculate the amino acids ionization states, followed by the addition of the counterions Na^+ and Cl^- to stabilize the charges. A cubic water box with a side length of 10 Å was constructed around the complex using the TIP3P water model, and periodic boundary conditions were added. The molecular dynamics simulation was accomplished using the GROMACS 2024 software [57]. The system underwent minimization using the steepest descent algorithm for 50.000 steps, reaching convergence at 1000 kJ/(mol·nm). For equilibration, NVT ensemble was applied at a temperature of 310.15 K for 5 ns using the v-rescale thermostat. Finally, the production phase was carried out using an NPT ensemble for 500 ns, under the temperature of 310.15 K, with the v-rescale thermostat, and at a pressure of 1 atm, maintaining the C-rescale barostat.

4.6. Free Energy of Binding

The binding affinities were calculated with the Molecular Mechanics Poission-Boltzmann Surface Area (MMPBSA) method for both Artepillin C and Vemurafenib complexes from 200 ns onwards, i.e., after observable stabilization. The software gmxMMPBSA [58] was employed. At first, the contributing residues were established within a radius of 6 Å from the ligands. The method, then, decomposes the free energy of binding as a function of the constituent residues of the protein. Therefore, E_{MM}, G_{polar} and $G_{nonpolar}$ energy components are computed for each individual atom of the contributing residues in each bound and unbound state. Thereafter, the procedure calculates the interaction energy contribution of the residue *x*, as demonstrated:

$$\Delta R_x^{BE} = \sum_{i=0}^{n} (A_i^{bound} - A_i^{free})$$

where A_i^{bound} and A_i^{free} are the energy of atom *i* from residue *x*, and *n* stands for the total number of atoms in the residue. Therefore, the summation of the energy contribution from all residues is equivalent to the interaction energy, meaning $\Delta G_{binding} = \sum_{x=0}^{m} \Delta R_x^{BE}$, *m* being the total number of residues contributing to the complex. Yet only the enthalpic component ΔH, could be obtained using the gmxMMPBSA tool; for the entropic component $-T\Delta S$, quasiharmonic analysis was computed in gromacs. For that, a covariance

matrix was computed to obtain the eigenvectors and eigenvalues. Vibrational entropy was then calculated:

$$S = k_B \sum_i \left[\ln\left(\frac{k_B T}{h v_i} \right) + 1 \right]$$

where k_B is the Boltzmann constant, T is temperature, h is the Planck constant, and v_i is the frequency associated with the mode i.

5. Conclusions

This study highlights the promising potential of Artepillin C, a major bioactive compound from Brazilian green propolis, as a selective inhibitor of the oncogenic B-Raf^{600}E kinase. Through a combination of molecular docking and long-scale molecular dynamics simulations, Artepillin C demonstrated favorable binding affinity, persistent interactions with key catalytic residues, and structural stability within the ATP-binding site. These computational results suggest that Artepillin C could interfere with the MAPK signaling pathway, offering a natural and potentially less toxic alternative for melanoma treatment. Furthermore, the in silico analysis of physicochemical and ADME properties reinforces its drug-likeness and pharmacokinetic viability. Taken together, these findings support the continued investigation of Artepillin C and related phytochemicals as scaffolds for the development of novel targeted therapeutics against melanoma driven by B-Raf^{600}E mutation.

Author Contributions: Conceptualization, C.P.G. and G.F.d.S.; software, R.M.C. and M.D.P.; validation, A.F.S.; formal analysis, C.P.G. and G.F.d.S.; investigation, G.F.d.S.; data curation, C.P.G. and G.F.d.S.; writing—original draft preparation, C.P.G.; writing—review and editing, C.P.G. and D.G.G.R.; supervision, C.P.G.; project administration, C.P.G.; funding acquisition, F.S.K.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors are grateful to the State of São Paulo Research Foundation, FAPESP (FAPESP 2019/26686-2), for funding the Gaussian software license.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors gratefully acknowledge the developers of the freely available computational tools used in this study. We thank the Scripps Research Institute for providing the AutoDock and AutoDockTools software for molecular docking analyses. We are also grateful to the ORCA development team for access to the ORCA quantum chemistry program, which enabled our electronic structure calculations. Additionally, we acknowledge the developers of MOPAC for making semi-empirical calculations accessible through their software. The availability of these powerful open-source tools has been essential to the completion of this research.

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

References

1. Imani, S.; Roozitalab, G.; Emadi, M.; Moradi, A.; Behzadi, P.; Kaboli, P.J. The Evolution of BRAF-Targeted Therapies in Melanoma: Overcoming Hurdles and Unleashing Novel Strategies. *Front. Oncol.* **2024**, *14*, 1504142. [CrossRef] [PubMed]
2. Czarnecka, A.M.; Bartnik, E.; Fiedorowicz, M.; Rutkowski, P. Targeted therapy in melanoma and mechanisms of resistance. *Int. J. Mol. Sci.* **2020**, *21*, 4576. [CrossRef]
3. Śmiech, M.; Leszczyński, P.; Kono, H.; Wardell, C.; Taniguchi, H. Emerging BRAF Mutations in Cancer Progression and Their Possible Effects on Transcriptional Networks. *Genes* **2020**, *11*, 1342. [CrossRef] [PubMed]
4. Amanuel, B.; Grieu, F.; Kular, J.; Millward, M.; Iacopetta, B. Incidence of BRAF p.Val600Glu and p.Val600Lys Mutations in a Consecutive Series of 183 Metastatic Melanoma Patients from a High Incidence Region. *Pathology* **2012**, *44*, 357–359. [CrossRef]
5. Silva, G.d.B.; Mendes, A.P.; Macedo, M.P.d.; Pinto, C.A.L.; Gibbons, I.L.; Duprat Neto, J.P. Vemurafenib and Cutaneous Adverse Events—Report of Five Cases. *An. Bras. Dermatol.* **2015**, *90*, 242–246. [CrossRef] [PubMed]

6. Singh, A.K.; Sonawane, P.; Kumar, A.; Singh, H.; Naumovich, V.; Pathak, P.; Grishina, M.; Khalilullah, H.; Jaremko, M.; Emwas, A.-H.; et al. Challenges and Opportunities in the Crusade of BRAF Inhibitors: From 2002 to 2022. *ACS Omega* **2023**, *8*, 27819–27844. [CrossRef]
7. de Carvalho, F.M.d.A.; Schneider, J.K.; de Jesus, C.V.F.; de Andrade, L.N.; Amaral, R.G.; David, J.M.; Krause, L.C.; Severino, P.; Soares, C.M.F.; Caramão Bastos, E.; et al. Brazilian Red Propolis: Extracts Production, Physicochemical Characterization, and Cytotoxicity Profile for Antitumor Activity. *Biomolecules* **2020**, *10*, 726. [CrossRef]
8. Pereira, F.A.N.; Barboza, J.R.; Vasconcelos, C.C.; Lopes, A.J.O.; Ribeiro, M.N.d.S. Use of Stingless Bee Propolis and Geopropolis against Cancer—A Literature Review of Preclinical Studies. *Pharmaceuticals* **2021**, *14*, 1161. [CrossRef]
9. Batista, C.M.; de Queiroz, L.A.; Alves, Â.V.F.; Reis, E.C.A.; Santos, F.A.; Castro, T.N.; Lima, B.S.; Araújo, A.A.S.; Godoy, C.A.P.; Severino, P.; et al. Photoprotection and Skin Irritation Effect of Hydrogels Containing Hydroalcoholic Extract of Red Propolis: A Natural Pathway against Skin Cancer. *Heliyon* **2022**, *8*, e08893. [CrossRef]
10. Saito, Y.; Tsuruma, K.; Ichihara, K.; Shimazawa, M.; Hara, H. Brazilian Green Propolis Water Extract Up-Regulates the Early Expression Level of HO-1 and Accelerates Nrf2 after UVA Irradiation. *BMC Complement. Altern. Med.* **2015**, *15*, 421. [CrossRef]
11. Szliszka, E.; Helewski, K.J.; Mizgala, E.; Krol, W. Ethanolic Extract of Brazilian Green Propolis Sensitizes Prostate Cancer Cells to TRAIL-Induced Apoptosis. *Int. J. Oncol.* **2011**, *38*, 941–953. [PubMed]
12. Gastaldello, G.H.; Cazeloto, A.C.V.; Ferreira, J.C.; Rodrigues, D.M.; Bastos, J.K.; Campo, V.L.; Zoccal, K.F.; Tefé-Silva, C. Green Propolis Compounds (Baccharin and *p*-Coumaric Acid) Show Beneficial Effects in Mice for Melanoma Induced by B16f10. *Medicines* **2021**, *8*, 20. [CrossRef]
13. Quintino, R.L.; Reis, A.C.; Fernandes, C.C.; Martins, C.H.G.; Colli, A.C.; Crotti, A.E.M.; Squarisi, I.S.; Ribeiro, A.B.; Tavares, D.C.; Miranda, M.L.D. Brazilian Green Propolis: Chemical Composition of Essential Oil and Their In Vitro Antioxidant, Antibacterial and Antiproliferative Activities. *Braz. Arch. Biol. Technol.* **2020**, *63*, e20190408. [CrossRef]
14. Xu, X.; Yang, B.; Wang, D.; Zhu, Y.; Miao, X.; Yang, W. The Chemical Composition of Brazilian Green Propolis and Its Protective Effects on Mouse Aortic Endothelial Cells against Inflammatory Injury. *Molecules* **2020**, *25*, 4612. [CrossRef] [PubMed]
15. de Moura, S.A.L.; Negri, G.; Salatino, A.; Lima, L.D.d.C.; Dourado, L.P.A.; Mendes, J.B.; Andrade, S.P.; Ferreira, M.A.N.D.; Cara, D.C. Aqueous Extract of Brazilian Green Propolis: Primary Components, Evaluation of Inflammation and Wound Healing by Using Subcutaneous Implanted Sponges. *Evid.-Based Complem Altern. Med.* **2011**, *1*, 748283. [CrossRef]
16. Yamaga, M.; Nohara, M.; Hata, A.; Ito, T.; Furumoto, K.; Ohta, N.; Miyamae, J.; Tani, H.; Yamaki, A.; Fujitani, N. Tissue Distribution and Accumulation of Cinnamic Acid Derivatives from Brazilian Green Propolis in Mice. *NFS J.* **2025**, *38*, 100222. [CrossRef]
17. Celińska-Janowicz, K.; Zaręba, I.; Lazarek, U.; Teul, J.; Tomczyk, M.; Pałka, J.; Miltyk, W. Constituents of Propolis: Chrysin, Caffeic Acid, *p*-Coumaric Acid, and Ferulic Acid Induce PRODH/POX-Dependent Apoptosis in Human Tongue Squamous Cell Carcinoma Cell (CAL-27). *Front. Pharmacol.* **2018**, *9*, 336. [CrossRef]
18. Ueda, T.; Inden, M.; Shirai, K.; Sekine, S.; Masaki, Y.; Kurita, H.; Ichihara, K.; Inuzuka, T.; Hozumi, I. The Effects of Brazilian Green Propolis That Contains Flavonols against Mutant Copper-Zinc Superoxide Dismutase-Mediated Toxicity. *Sci. Rep.* **2017**, *7*, 2882. [CrossRef]
19. Son, N.T.; Ribeiro, V.P.; Bastos, J.K. Flavonoids from Green Propolis of the Northeastern Brazilian Caatinga *Mimosa tenuiflora* (Willd.) Poir.: A Chemotaxonomic Aspect. *Biochem. Syst. Ecol.* **2022**, *104*, 104473. [CrossRef]
20. Bankova, V.S.; de Castro, S.L.; Marcucci, M.C. Propolis: Recent Advances in Chemistry and Plant Origin. *Apidologie* **2000**, *31*, 3–15. [CrossRef]
21. Huang, Y.; Rong, C.; Zhang, R.; Liu, S. Evaluating Frontier Orbital Energy and HOMO/LUMO Gap with Descriptors from Density Functional Reactivity Theory. *J. Mol. Model.* **2017**, *23*, 3. [CrossRef] [PubMed]
22. Rathi, P.C.; Ludlow, R.F.; Verdonk, M.L. Practical High-Quality Electrostatic Potential Surfaces for Drug Discovery Using a Graph-Convolutional Deep Neural Network. *J. Med. Chem.* **2020**, *63*, 8778–8790. [CrossRef]
23. Suresh, C.H.; Remya, G.S.; Anjalikrishna, P.K. Molecular Electrostatic Potential Analysis: A Powerful Tool to Interpret and Predict Chemical Reactivity. *WIREs Comp. Mol. Sci.* **2022**, *12*, e1601. [CrossRef]
24. Dugave, C.; Demange, L. Cis–Trans Isomerization of Organic Molecules and Biomolecules: Implications and Applications. *Chem. Rev.* **2003**, *103*, 2475–2532. [CrossRef]
25. Aakeröy, C.B.; Panikkattu, S.; Chopade, P.D.; Desper, J. Competing Hydrogen-Bond and Halogen-Bond Donors in Crystal Engineering. *Cryst. Eng. Comm.* **2013**, *15*, 3125–3136. [CrossRef]
26. Nishio, M. The CH/π Hydrogen Bond in Chemistry. Conformation, Supramolecules, Optical Resolution and Interactions Involving Carbohydrates. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13873. [CrossRef] [PubMed]
27. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25. [CrossRef]

28. Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623. [CrossRef]
29. Muegge, I.; Heald, S.L.; Brittelli, D. Simple Selection Criteria for Drug-like Chemical Matter. *J. Med. Chem.* **2001**, *44*, 1841–1846. [CrossRef]
30. Beserra, F.P.; Gushiken, L.F.S.; Hussni, M.F.; Ribeiro, V.P.; Bonamin, F.; Jackson, C.J.; Pellizzon, C.H.; Bastos, J.K. Artepillin C as an Outstanding Phenolic Compound of Brazilian Green Propolis for Disease Treatment: A Review on Pharmacological Aspects. *Phytother. Res.* **2021**, *35*, 2274–2286. [CrossRef]
31. de Freitas Meirelles, L.E.; de Assis Carvalho, A.R.B.; Ferreira Damke, G.M.Z.; Souza, R.P.; Damke, E.; de Souza Bonfim-Mendonça, P.; de Oliveira Dembogurski, D.S.; da Silva, D.B.; Consolaro, M.E.L.; da Silva, V.R.S. Antitumoral Potential of Artepillin C, a Compound Derived from Brazilian Propolis, against Breast Cancer Cell Lines. *Anticancer. Agents Med. Chem.* **2024**, *24*, 117–124. [CrossRef] [PubMed]
32. Valivand, N.; Aravand, S.; Lotfi, H.; Esfahani, A.J.; Ahmadpour-Yazdi, H.; Gheibi, N. Propolis: A Natural Compound with Potential as an Adjuvant in Cancer Therapy—A Review of Signaling Pathways. *Mol. Biol. Rep.* **2024**, *51*, 931. [CrossRef]
33. Shahinozzaman, M.; Basak, B.; Emran, R.; Rozario, P.; Obanda, D.N. Artepillin C: A Comprehensive Review of Its Chemistry, Bioavailability, and Pharmacological Properties. *Fitoterapia* **2020**, *147*, 104775. [CrossRef] [PubMed]
34. Hirata, Y.; Motoyama, M.; Kimura, S.; Takashima, M.; Ikawa, T.; Oh-Hashi, K.; Kamatari, Y.O. Artepillin C, a major component of Brazilian green propolis, inhibits endoplasmic reticulum stress and protein aggregation. *Eur. J. Pharmacol.* **2021**, *5*, 174572. [CrossRef] [PubMed]
35. Arruda, C.; Pena Ribeiro, V.; Oliveira Almeida, M.; Aldana Mejía, J.A.; Casoti, R.; Kenupp Bastos, J. Effect of Light, Oxygen and Temperature on the Stability of Artepillin C and *p*-Coumaric Acid from Brazilian Green Propolis. *J. Pharm. Biomed. Anal.* **2020**, *178*, 112922. [CrossRef]
36. Oteiza, P.I.; Fraga, C.G.; Mills, D.A.; Taft, D.H. Flavonoids and the Gastrointestinal Tract: Local and Systemic Effects. *Mol. Aspects Med.* **2018**, *61*, 41–49. [CrossRef]
37. Panche, A.N.; Diwan, A.D.; Chandra, S.R. Flavonoids: An Overview. *J. Nutr. Sci.* **2016**, *5*, e47. [CrossRef]
38. de Souza Farias, S.A.; da Costa, K.S.; Martins, J.B.L. Analysis of Conformational, Structural, Magnetic, and Electronic Properties Related to Antioxidant Activity: Revisiting Flavan, Anthocyanidin, Flavanone, Flavonol, Isoflavone, Flavone, and Flavan-3-Ol. *ACS Omega* **2021**, *6*, 8908–8918. [CrossRef]
39. Jan, R.; Khan, M.; Asaf, S.; Lubna; Asif, S.; Kim, K.-M. Bioactivity and Therapeutic Potential of Kaempferol and Quercetin: New Insights for Plant and Human Health. *Plants* **2022**, *11*, 2623. [CrossRef]
40. Echeverri, F.; Cardona, G.; Torres, F.; Pelaez, C.; Quiñones, W.; Renteria, E. Ermanin: An Insect Deterrent Flavonoid from Passiflora Foetida Resin. *Phytochemistry* **1991**, *30*, 153–155. [CrossRef]
41. Lu, L.; Luo, K.; Luan, Y.; Zhao, M.; Wang, R.; Zhao, X.; Wu, S. Effect of Caffeic Acid Esters on Antioxidant Activity and Oxidative Stability of Sunflower Oil: Molecular Simulation and Experiments. *Food Res. Int.* **2022**, *160*, 111760. [CrossRef] [PubMed]
42. Sharma, N.; Biswas, S.; Al-Dayan, N.; Alhegaili, A.S.; Sarwat, M. Antioxidant Role of Kaempferol in Prevention of Hepatocellular Carcinoma. *Antioxidants* **2021**, *10*, 1419. [CrossRef] [PubMed]
43. Okińczyc, P.; Widelski, J.; Szperlik, J.; Żuk, M.; Mroczek, T.; Skalicka-Woźniak, K.; Sakipova, Z.; Widelska, G.; Kuś, P.M. Impact of Plant Origin on Eurasian Propolis on Phenolic Profile and Classical Antioxidant Activity. *Biomolecules* **2021**, *11*, 68. [CrossRef]
44. Wang, Y.; Li, C.; Li, Z.; Moalin, M.; Hartog, G.J.M.d.; Zhang, M. Computational Chemistry Strategies to Investigate the Antioxidant Activity of Flavonoids—An Overview. *Molecules* **2024**, *29*, 2627. [CrossRef]
45. Sumayya, P.C.; Mujeeb, V.M.A.; Muraleedharan, K. Radical Scavenging Capacity, UV Activity, and Molecular Docking Studies of 2′,5′,3,4-Tetrahydroxychalcone: An Insight into the Photoprotection. *Chem. Phys. Imp.* **2022**, *5*, 100126. [CrossRef]
46. Rana, J.N.; Mumtaz, S. Prunin: An Emerging Anticancer Flavonoid. *Int. J. Mol. Sci.* **2025**, *26*, 2678. [CrossRef] [PubMed]
47. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci. Rep.* **2017**, *7*, 42717. [CrossRef]
48. Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminformatics* **2012**, *4*, 17. [CrossRef]
49. Dewar, M.J.S.; Thiel, W. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907. [CrossRef]
50. Boys, S.F. Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of any Molecular System. *Proc. R. Soc.* **1950**, *200*, 542–554.
51. Neese, F. The ORCA Program System. *WIREs Comp. Mol. Sci.* **2012**, *2*, 73–78. [CrossRef]
52. 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 16, Revision C.01*; Gaussian, Inc.: Wallingford, CT, USA, 2016.

53. Berman, H.M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]
54. Goodsell, D.S.; Olson, A.J. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins* **1990**, *8*, 195–202. [CrossRef] [PubMed]
55. Laskowski, R.A. PDBsum: Summaries and Analyses of PDB Structures. *Nucleic Acids Res.* **2001**, *29*, 221–222. [CrossRef]
56. Brooks, B.R.; Brooks, C.L.; Mackerell, A.D.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614. [CrossRef]
57. Bekker, H.; Berendsen, H.; Dijkstra, E. *Physics Computing*; DeGroot, R., Nadrchal, J., Eds.; World Scientific Publishing: Praha, Czech Republic, 1993; pp. 252–256.
58. Bill, R.; Miller, I.; McGee, D., Jr.; Swails, J.; Homeyer, N.; Gohlke, H.; Roitberg, A.E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comp.* **2012**, *8*, 3314–3321.

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.

pharmaceuticals

MDPI

Article

Design, Synthesis, and Computational Evaluation of 3,4-Dihydroquinolin-2(*1H*)-One Analogues as Potential VEGFR2 Inhibitors in Glioblastoma Multiforme

Shafeek Buhlak [1,†], Nadeem Abad [1,†], Jihane Akachar [1], Sana Saffour [1], Yunus Kesgun [1], Sevval Dik [1], Betul Yasin [1], Gizem Bati-Ayaz [1], Essam Hanashalshahaby [1], Hasan Türkez [2] and Adil Mardinoglu [3,4,*]

[1] Trustlife Labs, Drug Research & Development Center, Istanbul 34774, Turkey; shafeekbuhlak@outlook.com (S.B.); abadnadeem3@gmail.com (N.A.); jihane.akachar@trustlifelabs.com (J.A.); sasaffour@gmail.com (S.S.); yunus.kesgun@trustlifelabs.com (Y.K.); sevvaldik8@gmail.com (S.D.); btlyasin@hotmail.com (B.Y.); batigizem@gmail.com (G.B.-A.); essam.hanash@trustlifelabs.com (E.H.)
[2] Department of Medical Biology, Faculty of Medicine, Atatürk University, Erzurum 25240, Turkey; hturkez@atauni.edu.tr
[3] Science for Life Laboratory, KTH-Royal Institute of Technology, SE-17165 Stockholm, Sweden
[4] Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London SE1 9RT, UK
* Correspondence: adilm@scilifelab.se
† These authors contributed equally to this work.

Abstract: Background/Objectives: Glioblastoma multiforme (GBM), an aggressive and deadly brain tumour, presents significant challenges in achieving effective treatment due to its resistance to current therapies and poor prognosis. This study aimed to synthesise and evaluate 23 novel analogues of 3,4-dihydroquinolin-2(1*H*)-one, designed to enhance druggability and solubility, and to investigate their potential as VEGFR2 inhibitors for GBM treatment. **Methods**: The synthesised compounds were analysed using *in silico* methods, including molecular docking and dynamics studies, to assess their interactions with key residues within the VEGFR2 binding pocket. *In vitro* evaluations were performed on U87-MG and U138-MG GBM cell lines using MTT assays to determine the IC50 values of the compounds. **Results**: Among the tested compounds, **4u** (IC50 = 7.96 μM), **4t** (IC50 = 10.48 μM), **4m** (IC50 = 4.20 μM), and **4q** (IC50 = 8.00 μM) demonstrated significant antiproliferative effects against both the U87-MG and U138-MG cell lines. These compounds exhibited markedly higher efficacy compared to temozolomide (TMZ), which showed IC50 values of 92.90 μM and 93.09 μM for U87-MG and U138-MG, respectively. Molecular docking and dynamics studies confirmed strong interactions between the compounds and VEGFR2 kinase, supporting their substantial anti-cancer activity. **Conclusions**: This study highlights the promising potential of 3,4-dihydroquinolin-2(1*H*)-one analogues, particularly **4m**, **4q**, **4t**, and **4u**, as VEGFR2-targeting therapeutic agents for GBM treatment. Further detailed research is warranted to validate and expand upon these findings.

Keywords: glioblastoma multiforme; 3,4-dihydroquinolin-2(*1H*)-one; therapeutic efficacy; molecular docking; molecular dynamics; VEGFA–VEGFR2 pathway; anti-cancer

1. Introduction

Glioblastoma multiforme (GBM) is the most common and aggressive form of primary brain cancer, designated as a grade IV glioma. It encompasses a genetically and

phenotypically diverse group of tumours characterised by rapid growth, widespread infiltration into surrounding brain tissue, and a notable resistance to existing therapies. Despite advancements in surgical techniques, radiation therapy, and chemotherapy, the prognosis for patients with GBM has not significantly improved, with median survival times lingering at 12–15 months over the last two decades [1,2]. The treatment of GBM faces several challenges, primarily due to the high heterogeneity of glioblastoma cells, which vary in cell type, density, size, mitotic activity, and the complex molecular mechanisms that drive their growth, vascularisation, necrosis, and survival. Another significant hurdle in treating GBM is the blood–brain barrier (BBB), a selective and sophisticated physiological barrier that regulates the passage of molecules between the brain and the bloodstream, effectively preventing drugs from reaching their target sites at therapeutic concentrations. Successful treatment strategies require therapeutic agents to penetrate the brain interstitial fluid, achievable either through receptor-mediated transport or lipid-mediated free diffusion across the BBB. Currently, almost all compounds in clinical trials for brain conditions are lipid-soluble small molecules with a molecular weight of 400–600 Da and low hydrogen bonding, typically equal to or less than seven hydrogen bonds with water as a solvent [3–8]. The therapeutic landscape for GBM has evolved significantly in recent years, marked by the emergence of novel treatments targeting specific molecular pathways within tumour cells. The conventional standard of care for GBM involves surgical removal of the tumour followed by concurrent chemotherapy and radiotherapy with temozolomide (TMZ) (Figure 1), an orally administered alkylating agent that is generally well-tolerated by most patients. However, the effectiveness of TMZ is limited by several factors. GBM cells frequently develop resistance to TMZ, largely due to the action of the repair enzyme O6-methylguanine-DNA-methyltransferase (MGMT). MGMT's activity, which includes the induction of DNA methylation, impedes the replication of tumour cells, undermining the efficacy of TMZ. Additionally, the limited capacity of TMZ to cross the BBB and its associated toxicity at therapeutic doses underscore the need for alternative therapeutic approaches [9–13]. The ongoing challenges in GBM treatment, such as limited BBB penetration, resistance mechanisms, dosage constraints, and toxicity, necessitate the exploration of novel therapeutic strategies. This has led to increased research interest in small molecule drugs capable of crossing the BBB and offering more selective and potent therapeutic effects. The quinoline scaffold has received significant attention in pharmaceutical development for its diverse pharmacological properties [14–17]. Mefloquine (Figure 1), a compound initially known for its antimalarial effects, has shown promise in cancer treatment by inhibiting tumour cell growth [18]. Notably, studies investigating mefloquine's ability to cross the BBB discovered that one of its enantiomers can do so in a stereoselective manner, highlighting its potential utility in treating brain tumours [19]. Previous studies have explored the relationship between antimalarial drugs and their effects on glioma cells, uncovering that quinine and chloroquine (Figure 1) exhibit notable anti-cancer properties. These compounds have been found to enhance sensitivity to TMZ and counteract resistance in newly diagnosed and recurrent TMZ-resistant gliomas through mechanisms like DNA replication interference and autophagy inhibition [20]. Additionally, aripiprazole (Figure 1), a drug primarily used for its antipsychotic effects in treating bipolar disorder and schizophrenia, has gained attention for its potential role in GBM therapy. Research conducted by Mi Seon et al. demonstrated that aripiprazole could reduce glioma U251 cell cytotoxicity by inhibiting the phosphorylation of Src and STAT3, impacting cell migration and apoptosis [21]. This suggests aripiprazole's ability to modulate crucial signalling pathways involved in GBM progression, warranting further exploration of its therapeutic potential against this formidable brain cancer [22–24]. Concurrently, cilostazol (Figure 1), initially developed

for cardiovascular conditions, has emerged as a compound of interest in GBM research. It has shown promise in delaying tumour growth in vivo by activating the calcium-activated potassium (BK) channel. The inactivation of this channel contributes to glioblastoma cells' resistance to radiation, implying that reactivating the BK channel could enhance therapeutic outcomes, positioning cilostazol as a potential repurposed therapy for GBM [25]. One of the primary mechanisms leading GBM progression is angiogenesis, a process that encourages tumour vascularisation, mainly regulated by vascular endothelial growth factor receptor 2 (VEGFR2) and its ligand VEGF-A. Overexpression of VEGF-A in GBM leads to irregular blood vessel formation, boosting tumour growth, which subsequently accelerates the tumour's invasive capacity. For these purposes, VEGFR2 was selected as a target gene based on prior research involving the designed peptide VGB4, known for its ability to bind to VEGFR2 and inhibit VEGF-A, thereby reducing cancer cell proliferation in the U87-MG glioblastoma cell line [26]. Targeting VEGFR2 has accordingly stood out as a critical therapeutic approach in the treatment of GBM [27,28]. There are many inhibitors for VEGFR2, such as avastin (bevacizumab), an anti-VEGF-A monoclonal antibody, which have been employed to slow down angiogenesis in GBM. However, the clinical efficacy of avastin has been limited due to the resistance development of GBM cells [29–31]. Sorafenib and sunitinib, small molecule VEGFR2 tyrosine kinase inhibitors (TKIs), block VEGFR2-mediated signalling, which is crucial for endothelial cell proliferation and migration [32,33]. However, their application in treating GBM has been restricted because of difficulties in penetrating the blood–brain barrier (BBB). These limitations emphasise the need for novel VEGFR2 inhibitors with improved BBB penetration and efficacy in targeting the growth of GBM cells. Leveraging insights from biological studies on existing derivatives, we aim to contribute to the ongoing search for innovative and effective treatments for GBM. Here, we focus on designing and synthesising novel 6-hydroxy-3,4-dihydroquinolin-2(1H)-one (Figure 1) derivatives. This scaffold has been shown to possess diverse pharmacological properties, including the possibility to inhibit VEGFR2. Hydrazone derivatives, in particular, have illustrated potency as VEGFR2 inhibitors, leading to improvements in their ability to induce cell death in GBM cell lines [34–37]. Moreover, research on the crystal structure of the VEGFR2 kinase domain in complex with PF-00337210 (N,2-dimethyl-6-(7-(2-morpholinoethoxy)quinolin-4-yloxy) benzofuran-3-carboxamide) has promoted our understanding of these interactions [38]. The discoveries presented in this research not only build upon existing knowledge but also pave the way for exploring new therapeutic strategies. By joining the collective effort to address the challenges posed by GBM, we hope to make a significant impact. A molecular docking study targeting VEGFR2 was conducted [39]. Additionally, the in silico ADME (Absorption, Distribution, Metabolism, and Excretion) properties of the derivatives indicated their potential to efficiently cross the BBB and exhibit antitumour activity within the central nervous system (CNS). This research highlights the significance of 6-hydroxy-3,4-dihydroquinolin-2(1H)-one derivatives as promising candidates for GBM therapy, contributing to the development of more selective and potent treatments for this devastating disease. This endeavour seeks to contribute to the ongoing search for more effective treatments for GBM, highlighting the potential of these novel compounds in the broader context of cancer research and therapy development.

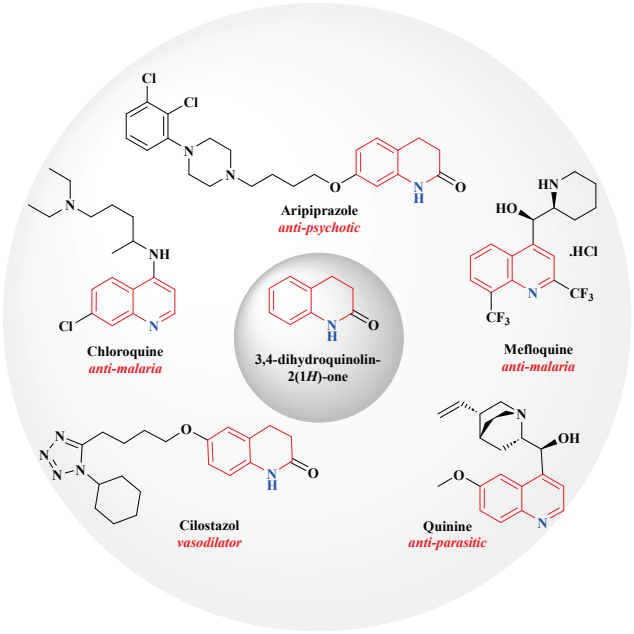
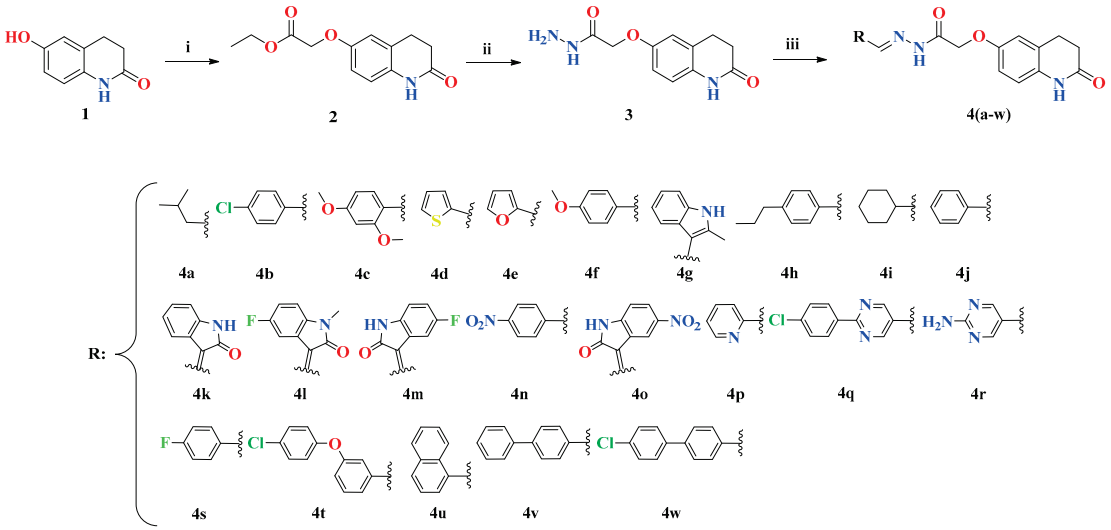


Figure 1. Some FDA-approved quinoline and quinolinone derivative drugs (tetrahydroquinoline core in red color).

2. Results and Discussion

2.1. Chemistry

The synthesis of the designed compounds was accomplished using a carefully adapted chemical procedure. This procedure not only yielded the desired compounds but also provided opportunities for introducing diverse functional groups, each with potential implications for the efficacy of the compounds against GBM. We outline the synthesis process and examine the structural differences among these compounds, emphasising the impact of electron-withdrawing (EWG) and electron-donating groups (EDG) on their potential activity against cancer cells. The journey toward these new compounds represents a significant step in exploring novel therapeutic options. Scheme 1 started with the conversion of 6-hydroxy-3,4-dihydroquinolin-2(*1H*)-one in THF, with drops of DMF. The hydroxyl group was activated by stirring it with base-like potassium carbonate (K_2CO_3) or triethylamine (TEA), for one hour. Subsequently, ethyl 2-bromoacetate was added drop-wisely into the reaction mixture. At the end of this step, ethyl 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetate was obtained. The subsequent transformation involved heating ethyl 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetate with hydrazine hydrate (99%) in ethanol at 70 °C (reflux) to yield 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide. Finally, the 23 hydrazone compounds were synthesised by reacting various hydrazides with aldehydes/ketones in ethanol at 70 °C (reflux), with the addition of drops of acetic acid, to obtain 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide. The reactions were carried out at reflux temperature for 4–12 h.



Scheme 1. Synthesis of 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide-hydrazone analogues **4**(**a**–**w**). Reagents and conditions: (**i**) K_2CO_3, THF, DMF, and ethyl 2-bromoacetate; (**ii**) hydrazine hydrate and EtOH, at reflux; (**iii**) corresponding aldehydes or ketones in EtOH and acetic acid.

The chemical structure of all the synthesised compounds was confirmed by LC-MS, ^{1}H NMR, and ^{13}C NMR spectral analysis (see Supplementary Materials). For compound **2**, the ^{1}H NMR spectral exhibit quartet at δ 4.15 ppm related to methylene protons (O-C**H$_2$**-CH$_3$), and there was a triplet at δ 1.20 ppm corresponding to methyl protons (O-CH$_2$-C**H$_3$**). That means a new ethylene group was formed from compound **1** to compound **2**. These findings could be approved by ^{13}C NMR spectral analysis, where there were new signals at δ 14.24 ppm and at δ 65.97 ppm, related to the methyl and methylene carbon, respectively. In addition, for compound **3**, the ethoxy group signals disappeared in ^{1}H NMR and ^{13}C NMR spectral analyses. Instead, there were new signals exhibited at δ 9.92 ppm for (C-N**H**-NH$_2$) and at δ 4.31 ppm for (-N**H$_2$**). The same finding was discovered for ^{13}C NMR spectral analysis, where there were no signals for the ethoxy group that was in compound **2**. Related to the desire compounds **4**(**a**–**w**), there were new signals appearing in ^{1}H NMR related to the methine group (N=C**H**-); for instance, compound **4a** exhibited a multiplet between δ (6.67–6.64) ppm and singlets at δ (8.33, 8.58, 8.18, 8.24, and 8.28) ppm, related to the compounds **4b**, **4c**, **4d**, **4e**, and **4f**, respectively. For ^{13}C NMR spectral analysis, the signal of methine carbon appeared at δ 144.2 ppm and at δ 169.76 ppm for the compounds **4j** and **4v**, respectively, as examples of the new methine carbon signal. Observing the molecular ion peaks of the mass spectra of all compounds, they conformed to the M + 1 or M − 1 peak of the appropriate molecular formula. For example, the $[M + 1]^+$ at *m*/*z* 249.96, 357.91, 366.20, and 397.12 were in accordance with the molecular formulae of **2**, **4b**, **4h**, and **4l** ($C_{13}H_{15}NO_4$, $C_{18}H_{16}ClN_3O_3$, $C_{21}H_{23}N_3O_3$, and $C_{20}H_{17}FN_4O_4$), respectively. Moreover, for $[M - 1]^-$ at *m*/*z* 363.34, 381.10, 367.21, 408.26, they were in accordance with the molecular formulae of **4k**, **4m**, **4n**, and **4o** ($C_{19}H_{16}N_4O_4$, $C_{19}H_{15}FN_4O_4$, $C_{18}H_{16}N_4O_5$, and $C_{19}H_{15}N_5O_6$), respectively.

2.2. In Vitro Studies

2.2.1. Efficacy of Novel Analogues on Cell Proliferation

U87-MG and U138-MG glioblastoma cell lines were used to evaluate the in vitro effects of the **4**(**a**–**w**) compounds on glioblastoma treatment. The effect of the compounds on cell proliferation was measured. Among them, the compounds **4t**, **4u**, and **4v** exhibited the highest antiproliferative activity on the U87-MG cell line, while the compounds **4q**, **4m**, and **4g** showed excellent results on the U138-MG cell line (Table 1, Figures 2 and 3).

Table 1. In vitro cell proliferation of compounds **4(a–w)** with their reagents against **U87-MG** and **U138-MG** cell lines.

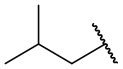
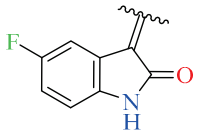
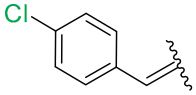
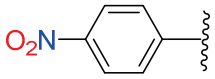
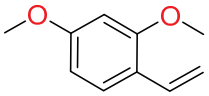
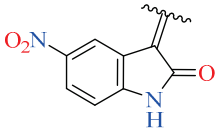
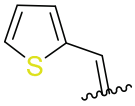
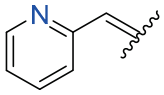
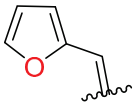
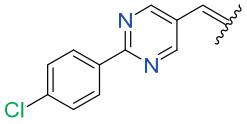
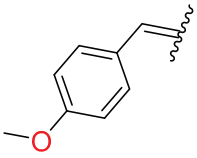
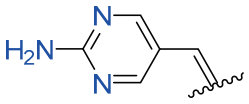
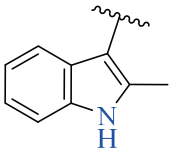
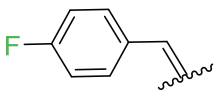
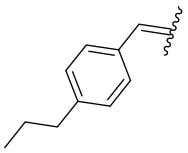
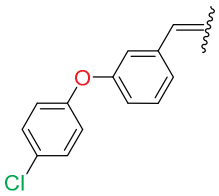
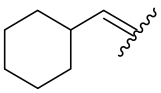
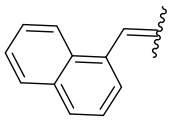
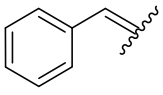
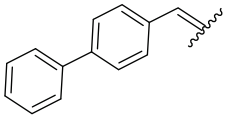
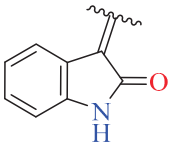
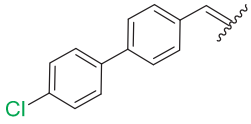
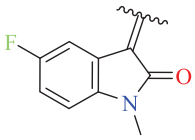
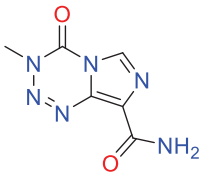
No	Code	R	Cell Proliferation ± Std dev. (%)		No	Code	R	Cell Proliferation ± Std dev. (%)	
			U87	U138				U87	U138
1	**4a**		79 ± 1.3	84 ± 5.2	**13**	**4m**		69 ± 2.1	54 ± 1.4
2	**4b**		65 ± 4.9	60 ± 4.0	**14**	**4n**		85 ± 1.2	92 ± 5.2
3	**4c**		98 ± 2.5	75 ± 3.0	**15**	**4o**		99 ± 7.0	83 ± 2.5
4	**4d**		80 ± 4.5	70 ± 5.0	**16**	**4p**		79 ± 1.4	81 ± 4.9
5	**4e**		75 ± 3.1	90 ± 5.0	**17**	**4q**		67 ± 4.9	39 ± 3.0
6	**4f**		92 ± 1.4	90 ± 2.6	**18**	**4r**		97 ± 1.3	92 ± 1.8
7	**4g**		66 ± 4.9	54 ± 3.0	**19**	**4s**		68 ± 5.0	68 ± 5.0
8	**4h**		64 ± 5.2	100 ± 2.3	**20**	**4t**		50 ± 3.2	65 ± 3.5
9	**4ı**		82 ± 5.5	68 ± 5.2	**21**	**4u**		53 ± 6.4	65 ± 5.5
10	**4j**		86 ± 2.6	74 ± 2.8	**22**	**4v**		49 ± 5.5	59 ± 7.0

Table 1. *Cont.*

No	Code	R	Cell Proliferation ± Std dev. (%)		No	Code	R	Cell Proliferation ± Std dev. (%)	
			U87	U138				U87	U138
11	**4k**		71± 2.1	95 ± 5	**23**	**4w**		63 ± 2.5	80 ± 6.9
12	**4l**		77 ± 7.0	73 ± 3.6	**24**	**TMZ**		80 ± 5	78 ± 5

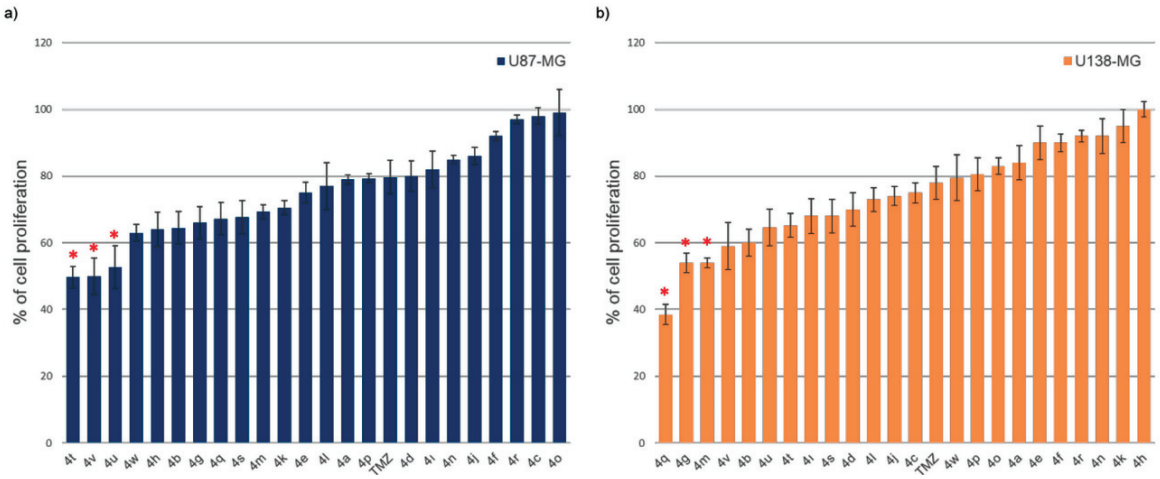


Figure 2. Cell proliferation of glioblastoma cell lines for 48 h of compound treatment: (**a**) U87-MG and (**b**) U138-MG. (*) indicates to most active compounds.

IC_{50} values were calculated depending on the increasing compound concentration (0–160 µM) for both glioblastoma cell lines. The most effective compounds for decreasing cell proliferation were **4u** (IC_{50} = 7.96 µM), **4t** (IC_{50} = 10.66 µM), and **4v** (IC_{50} = 11.66 µM) for U87-MG, and **4m** (IC_{50} = 4.20 µM), **4q** (IC_{50} = 8.00 µM), and **4g** (IC_{50} = 9.20 µM) for the U138-MG cell line (Figure 3). The decrease in cell proliferation was detected with the increase of the compound concentration.

2.2.2. Permeability of Novel Analogues Through Biological Membrane Mimics of the Blood–Brain Barrier (BBB) and Gastrointestinal (GI) Membrane

The permeability of the synthesised compounds was evaluated using the parallel artificial membrane permeability assay (PAMPA) to estimate their ability to cross biological membrane mimics of both the gastrointestinal (GI) membrane and the blood–brain barrier (BBB).

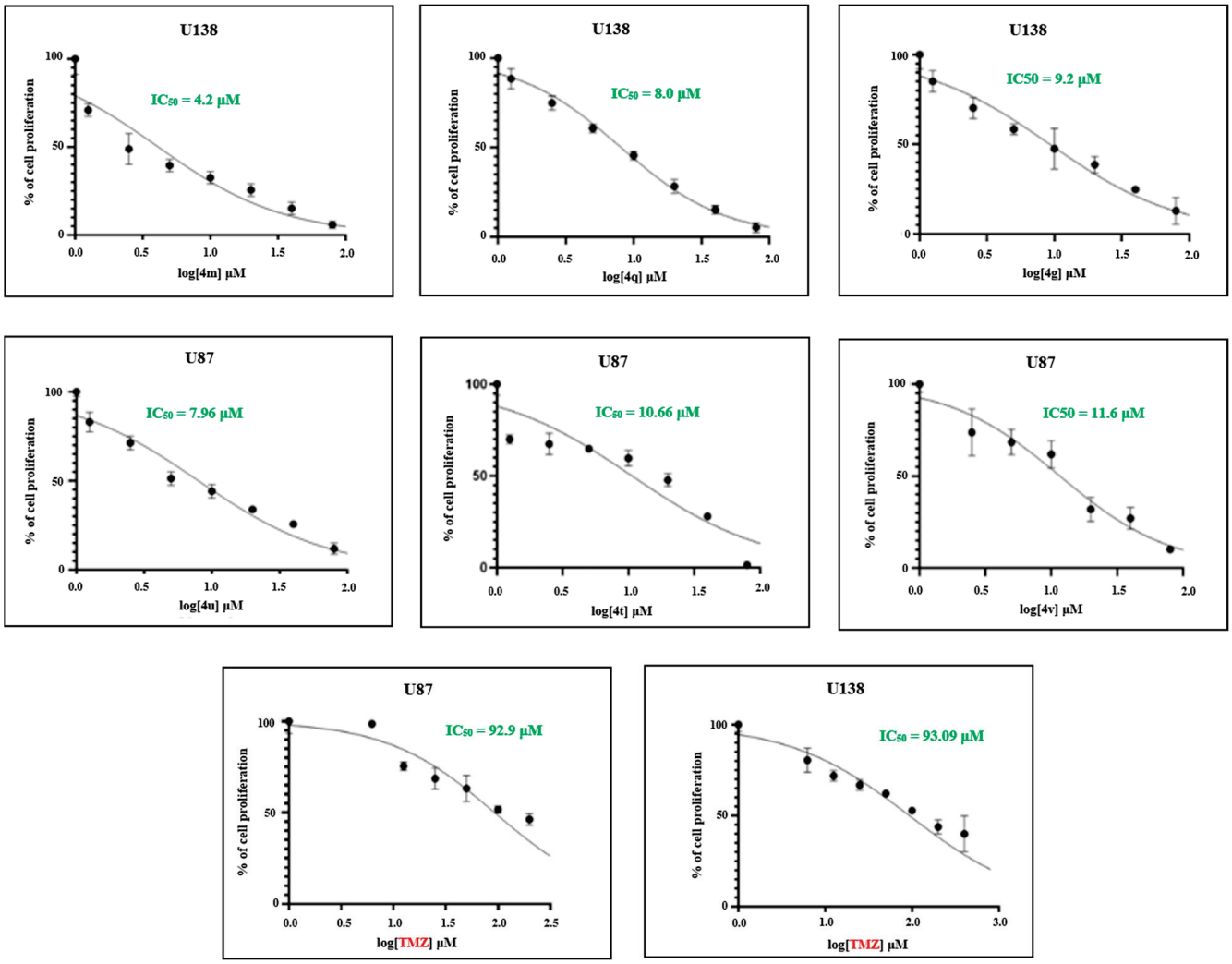


Figure 3. IC_{50} values of the best active molecules and TMZ for both glioblastoma cell lines.

The results showed that the compounds **4j**, **4t**, and **4h** exhibited high permeability across the BBB-mimicking membrane, proposing their potential to effectively penetrate the CNS. In contrast, the remaining compounds showed lower permeability across the BBB, suggesting that further structural optimisation may be necessary to enhance the penetration of the compounds in the CNS. However, for those demonstrating low PAMPA permeability, this could be attributed to the PAMPA model complexity. PAMPA uses a simple phospholipid-based artificial membrane that may not fully mimic the dynamic nature of the BBB in vivo, especially regarding active efflux mechanisms, tight junctions, and protein binding interactions.

Comparably, the GI permeability assay revealed that the compounds **4d** and **4s** showed high permeability across the gastrointestinal membrane, indicating their potential for efficient oral bioavailability. The remaining compounds demonstrated lower permeability, proposing possible limitations in their oral absorption profiles. These findings highlighted the importance of balancing hydrogen bonding, hydrophobicity, and molecular size to improve GI permeability. Remarkably, TMZ, used as a reference control, exhibited lower permeability in both the BBB and GI assays, consistent with its known pharmacokinetic profile (Figures 4 and 5).

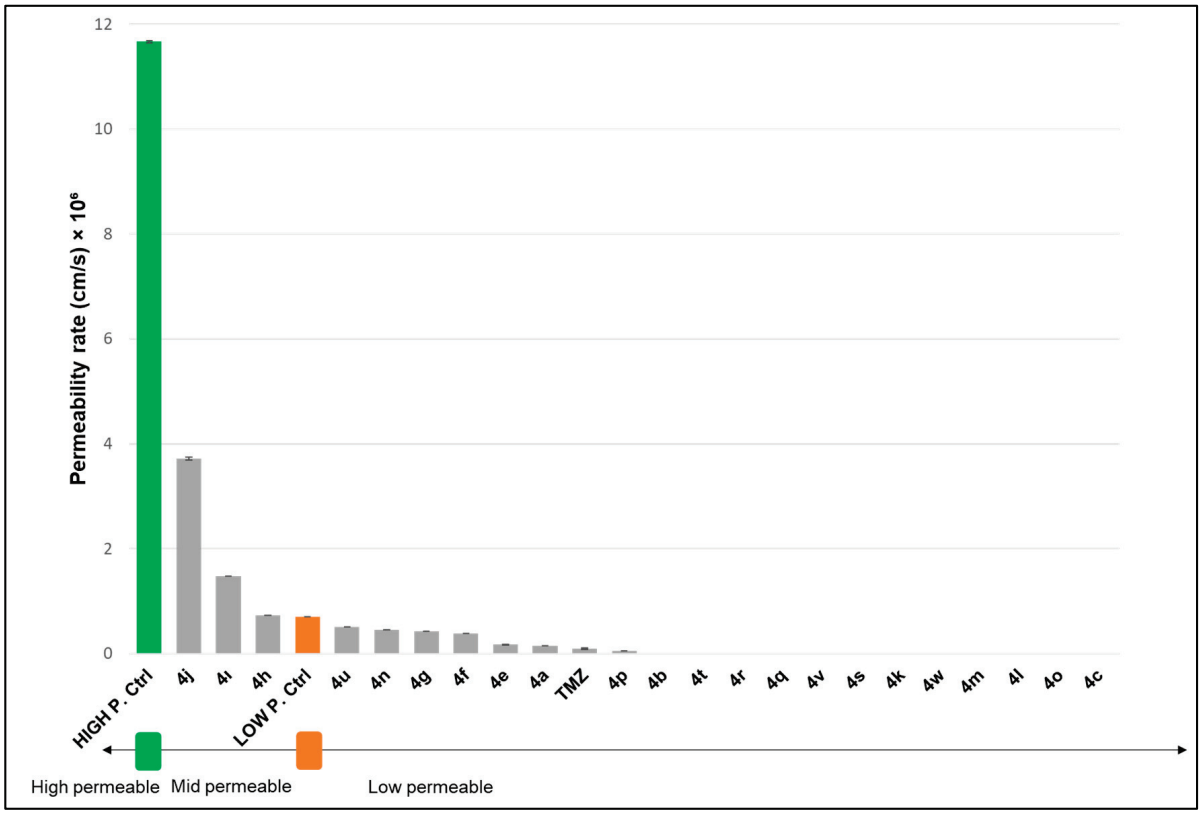


Figure 4. Permeability rates of compounds through the blood brain barrier (BBB)-mimicking membrane (HIGH P. Ctrl: high permeability control, LOW P. Ctrl: low permeability control, TMZ: temozolomide).

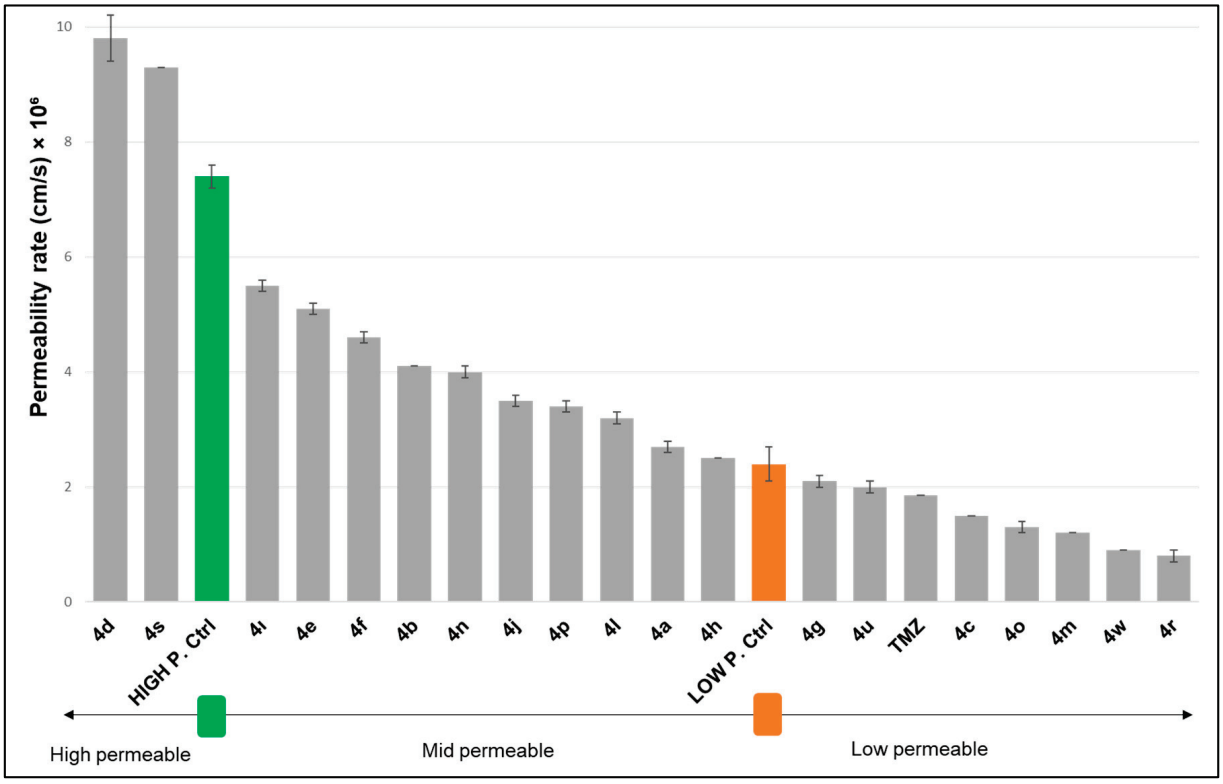


Figure 5. Permeability rates of compounds through the gastrointestinal (GI)-mimicking membrane (HIGH P. Ctrl: high permeability control, LOW P. Ctrl: low permeability control, TMZ: temozolomide).

2.3. In Silico Studies

2.3.1. Molecular Docking Studies

We conducted out molecular docking studies to have a thorough understanding of the binding method of the 3,4-dihydroquinolin-2(*1H*)-one derivatives, and TMZ as a reference, with the target protein VEGFR2, which plays a key role in angiogenesis—the formation of new blood vessels from existing ones. Targeting VEGFR2 with inhibitors can reduce angiogenesis and help restore blood–brain barrier (BBB) function. However, it may also have side effects by potentially increasing the permeability of the BBB to harmful substances [35,40,41]. The predicted docking scores for the binding interactions of the 3,4-dihydroquinolin-2(1H)-one derivatives and TMZ with VEGFR2 are presented in Table 2. The docking scores for the top compounds (**4w**, **4t**, **4u**, **4q**, and TMZ) were (−12.2, −11.4, −11.4, −11.1, and −6.2) kcal/mol, respectively, indicating moderate to strong interactions (Table 2). Additionally, it was observed that the top derivatives exhibited predominantly hydrophilic interactions, similar to the co-crystal ligand PF-00337210 (N,2-dimethyl-6-(7-(2-morpholinoethoxy)quinolin-4-yloxy)benzofuran-3-carboxamide), with the active site residues in the VEGFR2 kinase domain. These interactions involved key amino acids, such as Asp1046, Glu885, and Ile1025, alongside several important hydrophobic interactions (Figures 6 and 7). Meanwhile, compound **4m** displayed five hydrogen bonds with Asp1052, Leu840, Asn923, and Cys919, with a docking score of −9.9 kcal/mol, sharing common interacting residues with TMZ. Most of the synthesised 3,4-dihydroquinolin-2(1H)-one analogues demonstrated good binding affinity toward VEGFR2. Notably, the IC50 values observed for these compounds (**4u**: 7.96 μM, **4t**: 10.48 μM, **4m**: 4.20 μM, and **4q**: 8.00 μM) highlight their significant potency in reducing glioblastoma (GBM) cell viability, with compound **4m** exhibiting the lowest IC50, correlating with its strong docking score. In comparison, the standard treatment TMZ exhibited much higher IC50 values (92.90 μM for U87-MG and 93.09 μM for U138-MG), consistent with its weaker binding interactions in the docking studies.

Table 2. Molecular docking values of the synthesised analogues with TMZ as a reference drug.

Molecules	VEGF Receptor (PDB ID: 2XIR)		
	Binding Affinity (kcal/mol)	**Number of H-Bonds**	**H-Bonding Residues**
4a	−9.0	2	Asp1046
4b	−10.3	1	Glu885
4c	−9.9	2	Glu885, Arg1027
4d	−8.4	3	Leu840, Asn923
4e	−9.2	2	Asp1046, Glu885
4f	−8.8	2	Asp1046, Glu885,
4g	−10.3	2	Asp1046, Glu885
4h	−10.2	2	Glu885, Asp1046
4i	−9.2	5	Asp1046, Ile1025, Arg1027, Glu885
4j	−10.1	1	Glu885
4k	−10.7	5	Asp1046, Arg1027, Glu885
4l	−9.0	4	Ala881, His1026, Arg1027, Ile1025
4m	−9.9	5	Asp1052, Leu840, Asn923, Cys919
4n	−10.2	1	Asp1046

Table 2. *Cont.*

Molecules	VEGF Receptor (PDB ID: 2XIR)		
	Binding Affinity (kcal/mol)	Number of H-Bonds	H-Bonding Residues
4o	−9.3	1	Asp1046
4p	−9.9	2	Asp1046, Glu885
4q	−11.1	3	Asp1046, Glu885, Ile1025
4r	−9.4	2	Asp1046, Glu885
4s	−9.6	2	Asp1046, Glu885
4t	−11.4	4	Asp1046, Glu885, Ile1025
4u	−11.4	3	Asp1046, Glu885
4v	−10.4	2	Ser930, Cys919
4w	−12.2	1	Glu885
TMZ	−6.2	4	Leu840, Cys919

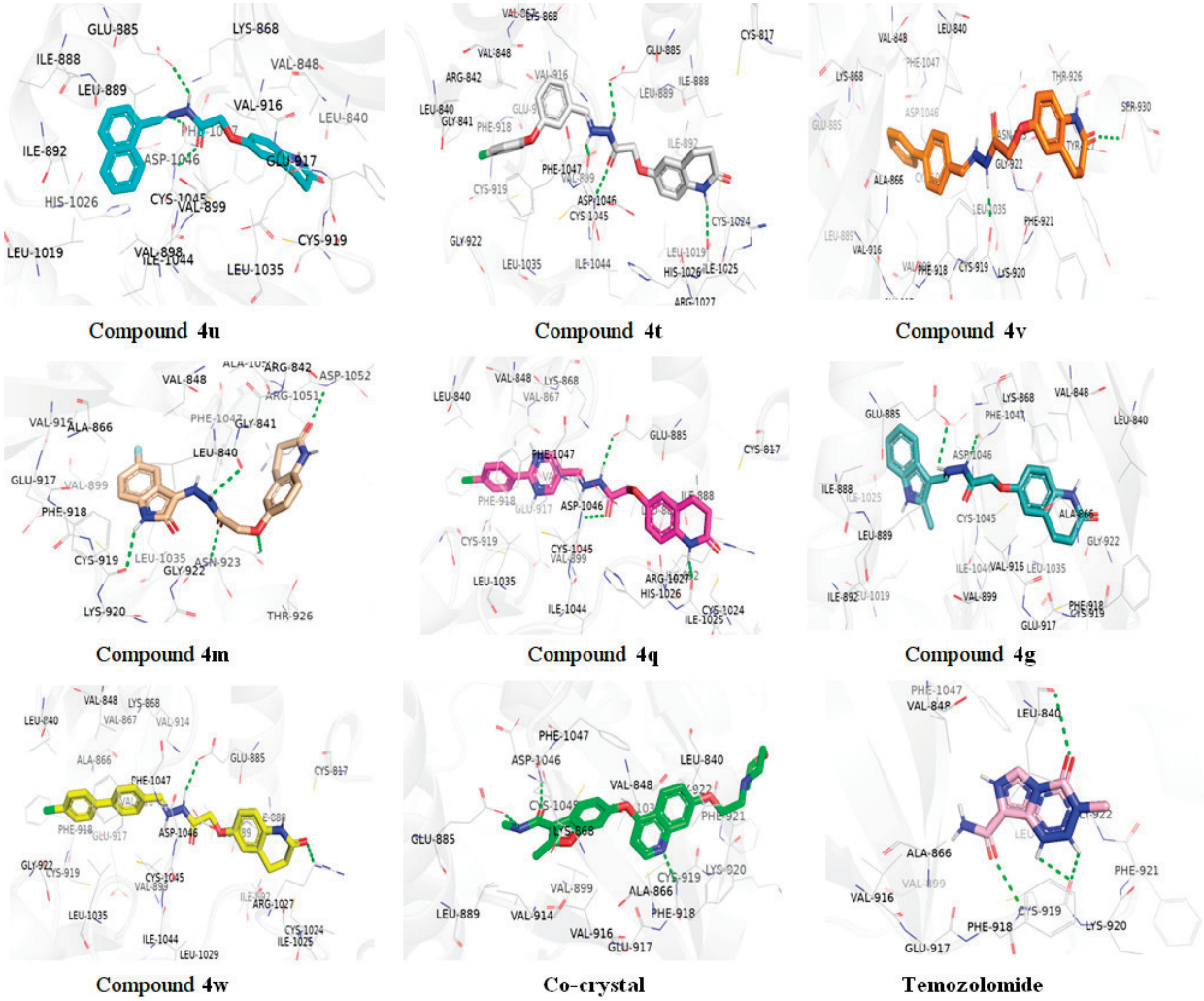


Figure 6. Molecular docking analysis of predicted drugs targeting VEGFR2. Putative binding modes of the best compounds **4u**, **4t**, **4v**, **4m**, **4q**, **4g**, **4w**, and temozolomide against VEGFR2 with the co-crystal PF-00337210 (*N*,2-dimethyl-6-(7-(2-morpholinoethoxy)quinolin-4-yloxy)benzofuran-3-carboxamide) as control. VEGFR2 is represented in a gray lines representation and hydrogen bonding residues highlighted by green dotted lines.

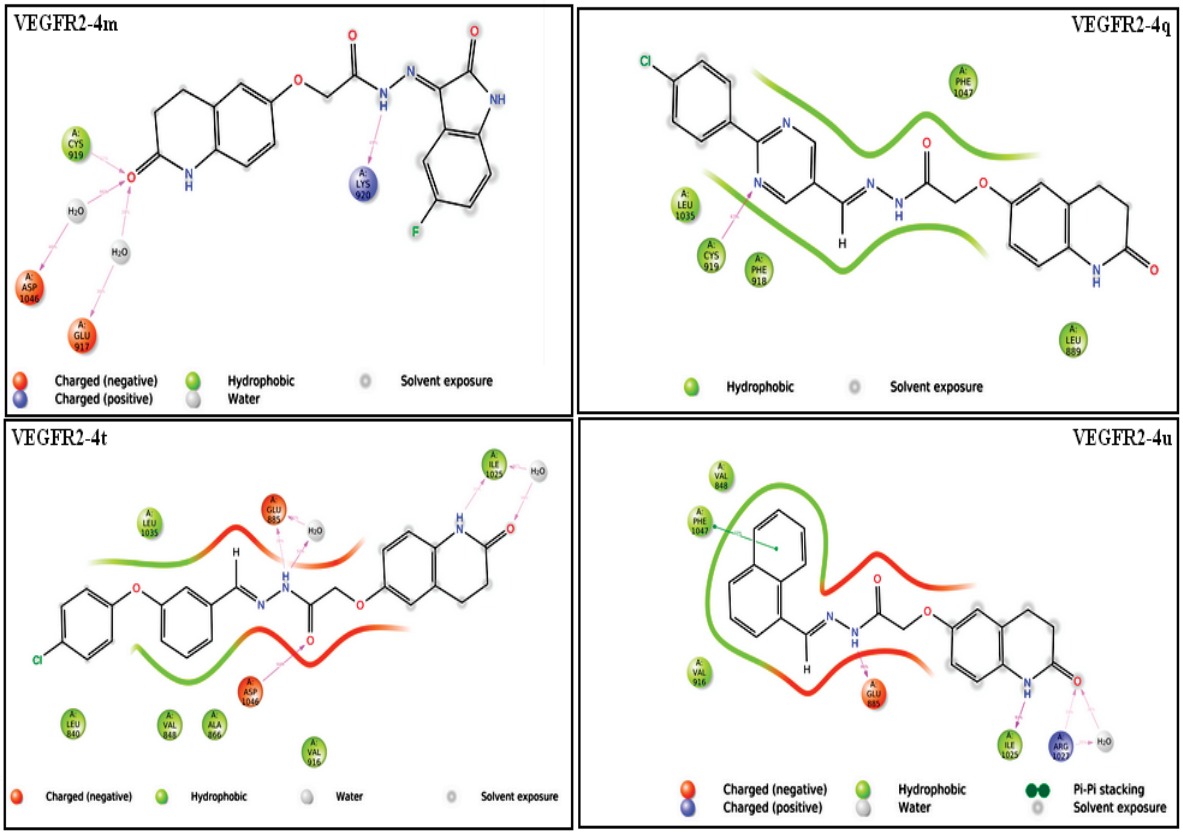


Figure 7. Diagram illustrating interactions between ligand atoms and protein residues of VEGFR2 complexed with the compounds **4m**, **4q**, **4t**, and **4u**.

Overall, the in silico findings and in vitro results underscore the promising anti-cancer potential of these 3,4-dihydroquinolin-2(1H)-one analogues, particularly the compounds **4m**, **4q**, **4t**, and **4u**.

2.3.2. Molecular Dynamics Simulation

To investigate the dynamic behaviour and stability of VEGFR2–3,4-dihydroquinolin-2(1H)-one derivative complexes, molecular dynamics (MD) simulations were conducted over a 100 ns period. This study focused on evaluating the structural changes induced by the best compounds with a good docking score and number of H-bond interactions, with insight into **4m**, **4q**, **4t**, and **4u**. The stability of each protein–ligand complex was monitored through RMSD analysis throughout the simulation. Figure 8 illustrates the RMSD values: the left *Y*-axis represents the RMSD of the protein, while the right *Y*-axis shows the ligand RMSD profile aligned with the protein backbone. Frames from the 100 ns trajectory were aligned with a reference frame to assess stability. The RMSD plot reveals that the VEGFR2–**4m** complex achieved stability after 20 ns, the VEGFR2–**4q** complex stabilised after 5 ns, although a minor increase in ligand RMSD was noted at 65 ns, the VEGFR2–**4t** complex became stable after 15 ns, and the VEGFR2–**4u** complex reached stability after 2 ns, with a slight rise in ligand RMSD observed at 40 ns. Such deviations could be linked to the high number ofthe ligand's rotatable bonds.

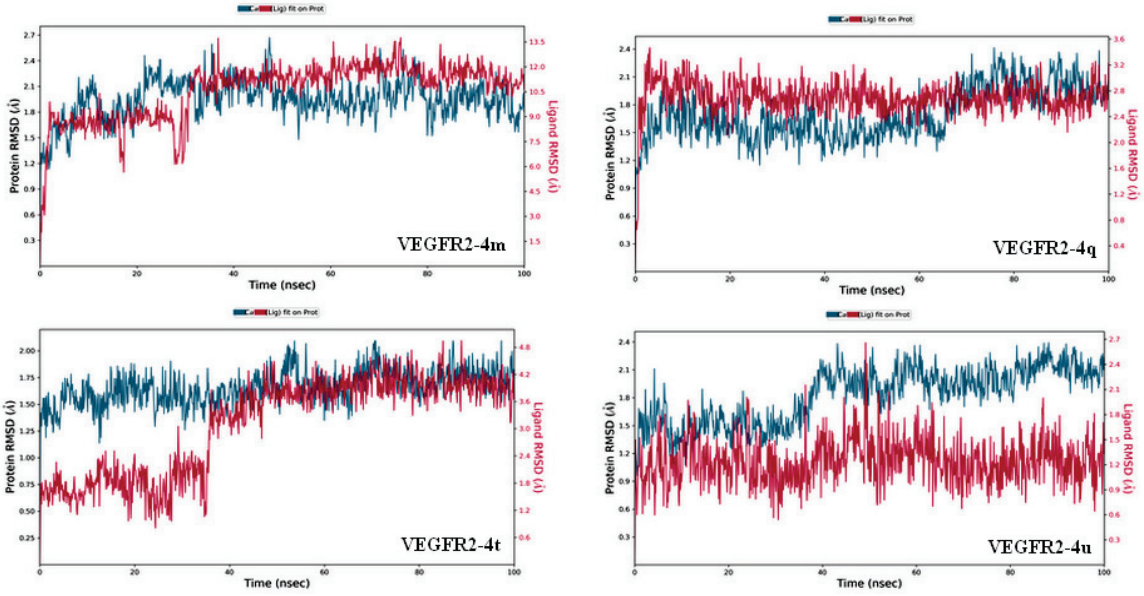


Figure 8. RMSD analysis of the protein and ligands over 100 ns of molecular dynamics simulation.

The root mean square fluctuation (RMSF) measures the average deviation of each atom's position from its mean position across a given simulation or set of structures. It provides valuable insights into the flexibility or mobility of residues, with higher RMSF values indicating greater flexibility or movement. In our study, some residues exhibited elevated RMSF values, indicating increased flexibility and notable impact. For instance, the residues at positions **860**, **873**, and **1064** showed the highest RMSF values in the dataset—**2.2**, **3.4**, and **4.0**, respectively—suggesting significant flexibility in the VEGFR2–**4q** complex compared to the other derivatives.

Conversely, residues with lower RMSF values, such as those between positions **840**–**848**, **868**, and **1028**, which correspond to the binding and active sites of VEGFR2, ranged from **0.5** to **1.0**, reflecting relatively low flexibility and greater rigidity.

Notably, key residues of the protein consistently maintained interactions with all 3,4-dihydroquinolin-2(*1H*)-one derivatives (**4m**, **4q**, **4t**, and **4u**). Residues involved in ligand binding were associated with lower RMSF values, indicating stable interactions, while residues not involved in binding exhibited higher RMSF values, highlighting their increased mobility.

Figure 8 illustrates the interactions between the compounds **4m**, **4q**, **4t**, and **4u** with the VEGFR2 active site pocket, observed for over 30% of the simulation duration. In the VEGFR2–4m complex, four hydrogen bonds were identified: one between **Cys919**, a water molecule, and the hydroxyl group, and another between **Lys920** and the amino group. Additionally, negative charge interactions occurred between **Glu917**, **Asp1046**, and the water molecule and hydroxyl group of **4m**. For the **VEGFR2–4q** complex, one hydrogen bond was formed between **Cys919** and the amino group, with hydrophobic interactions observed involving **Leu889**, **Cys919**, **Phe918**, **Leu1035**, and **Phe1047**. In the VEGFR2–4t complex, five hydrogen bonds were established: between **Glu885**, **Ile1025**, and the amino group, as well as between **Asp1046**, a water molecule, and the hydroxyl group. Hydrophobic interactions were observed with **Ile1025**, **Val916**, **Ala866**, **Val848**, **Leu840**, and **Leu1035**. For the VEGFR2–4u complex, four hydrogen bonds were identified: between **Glu885**, **Ile1025**, and the amino group, as well as between **Arg1027**, a water molecule, and the hydroxyl group. Additionally, a π–π interaction was noted between **Phe1047** and the aromatic group of **4u**, along with hydrophobic interactions involving **Leu840**, **Phe1047**,

Val916, and **Ile1025**. The underlined amino acids represent residues involved in the ATP interaction site of the VEGFR2 receptor (ID: **P35968 VGFR2_HUMAN**).

Figure 9 illustrates the energy components of the non-covalent interactions observed during the simulation. The *X*-axis represents the residues at the active site that interact with the ligand, while the *Y*-axis indicates the proportion of the simulation time during which these interactions occur. The stacked bar charts are normalised across the entire simulation trajectory. As shown in Figures 9 and 10, the ATP binding site residue Leu840 participated in hydrophobic interactions with the compounds **4m**, **4q**, and **4t** for approximately 42%, 58%, and 33% of the simulation time, respectively. Similarly, Leu848 engaged in hydrophobic interactions with **4m**, **4q**, **4t**, and **4u** for around 2%, 2%, 53%, and 78% of the simulation duration, respectively. The ATP binding site residue Leu868 exhibited hydrophobic interactions with **4m**, **4t**, and **4u** for about 1% of the simulation time each, and formed hydrogen bonds with **4q** for approximately 3%. Furthermore, Cys919 formed hydrogen bonds with **4m** and **4q** for over 65% of the simulation period, while Asp1046 showed hydrogen bonding with 4t and **4u** for 80% and 23% of the time, respectively. Ile1025 interacted with **4t** and **4u** for 33% and 100% of the simulation time, respectively. Other interacting residues displayed various types of interactions, including ionic and water-bridged interactions with the ligand. As a result, these residues were involved in multiple interactions throughout the simulation. Molecular dynamics simulations further confirmed that the compounds **4u**, **4t**, **4m**, and **4q** maintained stable interactions with VEGFR2 over time, supporting their potential to inhibit VEGFR2 activity and reduce GBM cell viability. Notably, compound 4m exhibited the lowest IC50, which correlated with its stable binding observed in the simulations.

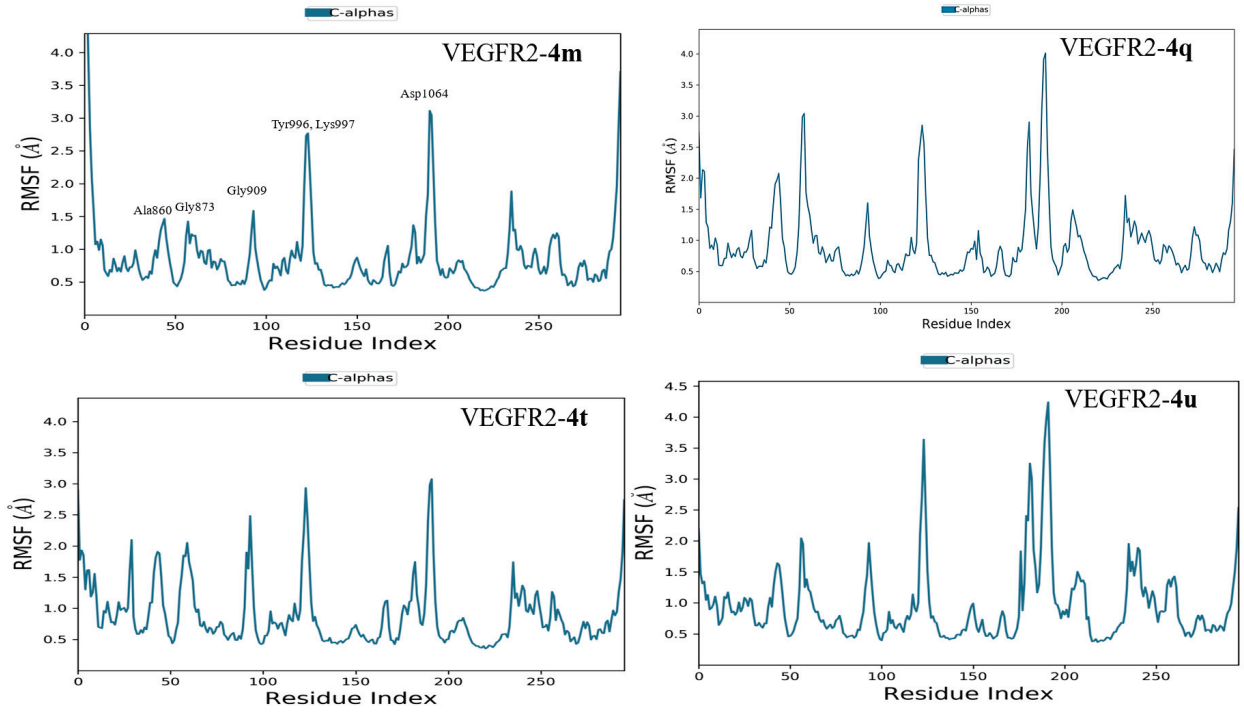


Figure 9. Residue-level fluctuations of VEGFR2 complexed with the compounds **4m**, **4q**, **4t**, and **4u**.

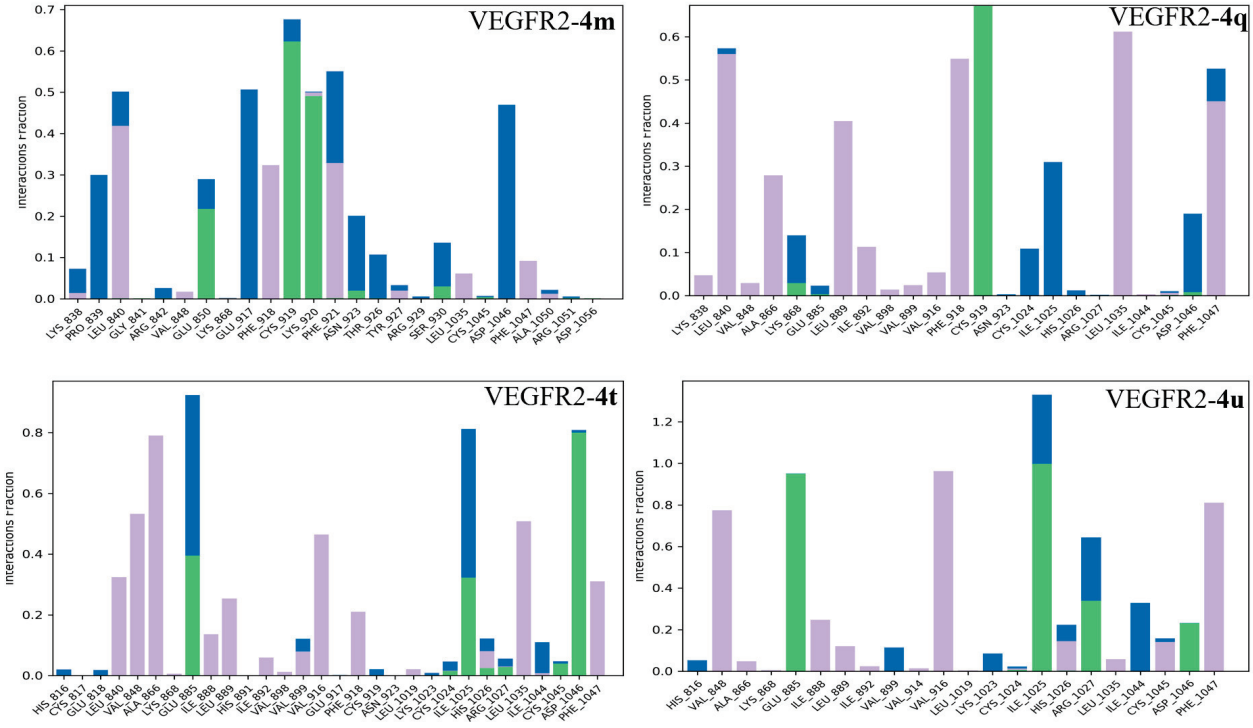


Figure 10. Protein–ligand interaction dynamics throughout the simulation period of VEGFR2 complexed with the compounds **4m**, **4q**, **4t**, and **4u**. Hydrogen bonds, hydrophobic contacts, ionic interactions, and water bridges are colored green, purple, pink, and blue, respectively.

2.3.3. In Silico ADME Studies

A computational study of all 23 molecules **4**(**a**–**w**) was conducted to predict their ADME properties using the QikProp3.2 tool that is available in Schrödinger and SwissADME, and the PreADME online server. Key parameters were selected, relevant to BBB permeability, using the QikProp3.2 tool available in Schrödinger (Table 2). logBB values for all molecules were found to be within a favourable range, indicating their potential for BBB penetration. MDCK cell permeability (QPPMDCK) is considered a good in vitro mimic for BBB permeability, with values < 25 indicating poor permeability and values > 500 indicating excellent permeability. Most of the tested molecules showed MDCK values within the optimal range, with the compounds **4b**, **4p**, **4t**, and **4w** exhibiting values > 300, demonstrating good BBB penetration. In contrast, only three compounds (**4n**, **4o**, and **4r**) displayed poor permeability, suggesting limited BBB transport potential.

Additionally, the drug-likeness prediction results for all derivatives were evaluated based on Lipinski's "Rule of Five", confirming that the majority of the compounds met the established criteria for oral bioavailability and drug-likeness [42].

All four principal parameters for the derivatives—H-bond donors (<5), H-bond acceptors (<10), LogP (<5), and molecular weight (<500 Da)—were within the acceptable ranges for drug-like compounds [43]. According to Jorgensen's "Rule of Three", the compounds should also meet the following criteria: QPlogS > −5.7, QPPCaco > 22 nm/s, and primary metabolites < 7. The blood–brain barrier (BBB) permeability of the compounds was calculated using SwissADME and PreADME, with the obtained data summarised in Tables 3 and 4. All parameter values were found to be within the specified limits, indicating that the 3,4-dihydroquinolin-2(1H)-one derivatives have the potential to effectively cross the BBB.

Table 3. In silico pharmacokinetic parameters obtained by the QikProp tool Schrödinger.

Comp.	CNS (−2 to +2)	Mw < 500	HBD (0–6)	HBA (2.0–20)	QPlogPo/w (−2.0–6.5)	QPlogS (−6.5–0.5)	QPPCaco <25 Poor >500 High	Metabsm (1–8)	%Human Oral Ab (>80% High <25% Poor)	N&O (<2 – 15>)	Rule of 5 (Max 4)	Rule of 3 (Max 3)
4a	−2	303.36	2	5.75	2.431	−4.886	209.18	5	82.715	6	0	0
4b	−2	357.796	2	5.75	3.319	−5.642	313.438	4	91.058	6	0	0
4c	−2	383.403	2	7.25	3.075	−5.791	296.37	6	89.191	8	0	1
4d	−2	329.373	2	5.75	2.765	−4.901	308.989	5	87.702	6	0	0
4e	−2	313.312	2	6.25	2.163	−4.157	301.102	5	83.972	7	0	0
4f	−2	353.377	2	6.5	2.938	−5.185	309.992	5	88.736	7	0	0
4g	−2	376.414	3	5.75	3.386	−5.97	228.303	5	88.981	7	0	1
4h	−2	365.431	2	5.75	3.849	−6.233	312.955	5	94.148	6	0	1
4i	−2	329.398	2	5.75	2.831	−5.617	203.198	5	84.828	6	0	0
4j	−2	323.351	2	5.75	2.831	−4.93	310.285	4	88.119	6	0	0
4k	−2	364.36	2	6.75	2.144	−5.335	49.12	4	69.77	8	0	0
4l	−2	396.377	1	7.25	2.775	−5.346	102.083	4	79.152	8	0	0
4m	−2	382.35	2	6.75	2.404	−5.603	54.475	4	72.098	8	0	0
4n	−2	368.348	2	6.75	2.135	−5.132	35.954	5	67.293	9	0	0
4o	−2	423.384	1	8.25	1.733	−4.265	27.701	5	49.954	11	1	0
4p	−2	323.351	2	5.75	2.828	−4.948	303.005	4	87.915	6	0	0
4q	−2	435.869	2	7.75	3.732	−6.758	171.738	6	88.794	8	0	1
4r	−2	340.341	4	8.25	0.686	−4.007	25.356	6	56.094	9	0	0
4s	−2	341.341	2	5.75	3.062	−5.289	310.103	4	89.47	6	0	0
4t	−2	449.893	2	6.25	4.882	−7.62	306.659	4	100	7	0	1
4u	−2	373.41	2	5.75	3.724	−6.051	304.844	4	93.21	6	0	1
4v	−2	399.448	2	5.75	4.409	−6.883	305.004	4	100	6	0	1
4w	−2	433.893	2	5.75	4.89	−7.596	303.939	4	100	6	0	1
TMZ	−2	194.152	2	7	−1.207	−1.379	59.807	1	51.677	8	0	0

Table 4. Blood–brain barrier (BBB) parameter obtained by the SwissADME, PreADME, and QikProp Schrödinger tools. HBD: Hydrogen bond doner, HBA: Hydrogen bond acceptor.

Molecules	BBB			
	PreADME	**SwissADME**	**QikProp Schrödinger**	**BBB**
4a	0.097/Low	No	Yes	Yes/Low
4b	0.149/Med	No	Yes	Yes/Low
4c	0.030/Low	No	Yes	Yes/Low
4d	0.0220/Low	No	Yes	Yes/Low
4e	0.0236/Low	No	Yes	Yes/Low
4f	0.0375/Low	No	Yes	Yes/Low
4g	0.314/Medl	No	Yes	Yes/Med
4h	0.394/Medl	No	Yes	Yes/Med
4i	0.258/Medl	No	Yes	Yes/Med
4j	0.062/Low	No	Yes	Yes/Low
4k	0.025/Low	No	Yes	Yes/Low
4l	0.0212/Low	No	Yes	Yes/Low
4m	0.024/Low	No	No	Yes/Low
4n	0.024/Low	No	No	Yes/Low
4o	1.061/Medl	Yes	No	Yes/Medl
4p	0.020/Low	No	Yes	Yes/Low
4q	0.015/Low	No	Yes	Yes/Low
4r	0.028/Low	No	No	Yes/Low
4s	0.082/Low	No	Yes	Yes/Low
4t	0.184/Medl	No	Yes	Yes/Medl
4u	0.090/Low	No	Yes	Yes/Low
4v	0.156/Medl	No	Yes	Yes/Medl
4w	0.450/Med	No	Yes	Yes/Medl
TMZ	0.017/Low	No	No	Yes/Low

*2.4. Structure–Activity Relationship (SAR) Exploration of Compounds (**4a–4w**) Against U138-MG Glioblastoma Cells*

Bioactivity analysis focused on explaining the description of relationships between molecular structures and the remarked biological activities against the U138-MG glioblastoma cell line, compared against **TMZ**. With understanding gathered from the molecular docking studies against the VEGF receptor, notable trends emerge to describe the flexibility in compound potency modulation. The compounds **4m** (IC_{50} = 4.20 μM), **4q** (IC_{50} = 8.00 μM), and **4g** (IC_{50} = 9.20 μM) exhibit significant cell inhibition percentages, underscored by their respective docking scores (−9.9 kcal/mol, −11.1 kcal/mol, and −10.3 kcal/mol). Molecular docking interprets key interactions, in which hydrogen bonding with critical residues such as Asp1046 and Glu885 confers favourable binding affinities. Noteworthy trends appear, highlighting the complexity that controls the compound potency. The compounds **4e** (10%) and **4d** (30%), which have a furan and thiophen moiety, respectively, display heightened bioactivity, underscoring the significance of specific structural motifs as heterocyclic structures. The *N′*-(3-methylbutylidene) group in **4a** contributes to moderate activity (16%). The presence of the 3-methylbutylidene moiety may lead to

the observed bioactivity. The *N*′-((2-methyl-1*H*-indol-3-yl)methylene) group in **4g** (46%) positively influences activity, suggesting the significance of aromatic functionalities. The *N*′-([1,1′-biphenyl]-4-ylmethylene) moiety in **4v** (41%) enhances activity, suggesting a role for extended aromatic systems. However, the presence of a chloro-substituent in **4w** that has a *N*′-((4′-chloro-[1,1′-biphenyl]-4-yl)methylene) group leads to a notable decrease in bioactivity (20%) compared to **4v**. The 4-chlorobenzylidene group in **4b** (40%) provides high activity, meaning that chlorine substitution might enhance interactions with the target, leading to increased bioactivity. Replacing chlorine by fluorine in the *N*′-(4-fluorobenzylidene) group in **4s** (32%) may enhance electronic effects and thus lead to the drop in its cytotoxic activity of 8%, and its docking score can relate to the same aspect, with docking scores of −10.3 and −9.6 for **4b** and **4s**, respectively. However, the replacement of halide with methoxy in the compound **4c** (25%), which has a *N*′-(2,4-dimethoxybenzylidene) group, provides moderate activity by emphasising the relevance of an electron-donating group (EDG). On the other hand, **4f** (10%) has the presence of one with the lowest docking score, −8.8, and comparing with **4b** and **4c**, its methoxy group on the para position in the *N*′-(4-methoxybenzylidene) group leads to reduced activity. Interestingly, compound **4h** (0%), lacking noticeable activity, suggests that 4-propylbenzylidene moiety may be unfavourable for bioactivity. The *N*′-benzylidene group in **4j** exhibits moderate activity (26%). The presence of the benzylidene group may contribute to favourable interactions. Replacing the phenyl group with cyclohexyl in the *N*′-(cyclohexylmethylene) group in **4i** shows moderate activity (32%). The cyclohexylmethylene substituent may contribute to favourable bioactivity. The presence of the 3-(4-chlorophenoxy) benzylidene group in **4t** exhibits moderate activity (35%), and may contribute to favourable interactions with the target. Replacing the benzylidene group by the naphthalen-1-ylmethylene substituent in **4u** exhibits moderate activity (35%). The *N*′-(4-nitrobenzylidene) group in **4n** shows limited activity (8%), suggesting that the nitro substituent as EWG may not boost bioactivity. Moreover, this group could sometimes be less favourable for cellular absorption. Seemingly, in 4r (8%), the 2-aminopyrimidin-5-yl group shows very low activity, meaning that this group may not be optimal for bioactivity. The presentation of the nitrobenzylidene group in **4o** resulted in moderate activity, proposing a potential role for electron-withdrawing substituents. On the contrary, the N′-(5-fluoro-2-oxoindolin-3-ylidene) group in **4m** contributes to high activity (46%). The presence of the fluoro and 2-oxoindolin-3-ylidene groups may enhance bioactivity through favourable interactions with five H-bonds with the important residues Asp1052, Leu840, Asn923, and Cys919. Contrarily, the combination of a pyridinylmethylene group in **4p** yielded analogous effects, indicating the relevance of deferent substituent effects. Compound **4q** (61%), containing a chlorophenylpyrimidinylmethylene group, exhibited a remarkable boost in activity, surpassing **TMZ**.

A computational and in vitro study was conducted on the most promising molecules—**4g**, **4m**, and **4q**—in comparison with temozolomide (TMZ) (Table 5). The key parameters, including molecular docking scores, IC50 values, and ADME properties, were found to be within optimal ranges, supporting the potential therapeutic value of these 3,4-dihydroquinolin-2(1H)-one derivatives and confirming their promise as effective candidates for further investigation. This highlights the importance of heterocyclic moieties with varied electronic properties. The findings from this bioactivity analysis provide a roadmap for targeted structural modifications, offering guiding principles for further optimisation and rational drug design efforts. These discoveries offer valuable insights into the development of novel therapeutics for glioblastoma, contributing to advancements in precision medicine.

Table 5. A summary comparing the key properties of the most promising compounds, including IC50 values, molecular docking scores, and ADME properties, as obtained using the QikProp tool from Schrödinger.

No.	Binding Affinity for VEGFR2 (kcal/mol)	IC_{50} Values (μM) U138	ADME						
			BBB	CNS (−2 to +2)	MolWt < 500	QPlogS (−6.5–0.5)	QPPCaco <25 Poor >500 High	Metabsm (1–8)	%Human OralAb >80% High <25% Poor
4g	−10.3	9.2	Yes	−2	376.41	−5.97	228.303	5	88.981
4m	−9.9	4.2	No	−2	382.35	−5.603	54.475	4	72.098
4q	−11.1	8.0	Yes	−2	435.86	−6.758	171.738	6	88.794
TMZ	−6.2	93.09	Yes	−2	194.15	−1.379	59.807	1	51.677

3. Materials and Methods

3.1. Materials

Reagents and solvents were obtained from Sigma-Aldrich and Ambeed and were used without further purification. Reactions were monitored via LC-MS (Thermo Fisher TSQ Series, Athena C18-WP column, Waltham, MA USA, water with 0.01% formic acid) or TLC (silica gel, Kieselgel 60, E. Merck, Germany). Proton NMR (^{1}H NMR) spectra were recorded on an Advance Bruker 500 MHz spectrometer (Bruker, Billerica, MA, USA) using $CDCl_3$ and DMSO-d_6, with data reported in δ/ppm relative to TMS, including multiplicity, hydrogen count, and coupling constants (J in Hz). Carbon-13 NMR (^{13}C NMR) data were also given in δ/ppm. Chemical shifts use the residual solvent peak as a reference, with standard abbreviations for signal patterns. Flash column chromatography was conducted on a PuriFlash XS 520 Plus system, Montluçon, France, using (Interchim) silica gel columns, with UV and ELSD detection and fraction collection.

3.2. General Procedures and Spectral Data

Acylation reactions were initiated as the primary step (**i**), wherein compounds containing hydroxy/amino groups (**1**) (1 eq.) and the base K_2CO_3 (1.5 eq.) or triethylamine TEA (1.5 eq.) were treated with acyl chloride derivatives in THF, acetone, or DMF. This process yielded acylated products [44]. Upon completion of the reaction, solvents were evaporated, residues were washed with water, dried, and subjected to purification to obtain the corresponding acyl compounds (**2**).

The second step (**ii**) involved the synthesis of hydrazide products (**3**). This was achieved by refluxing acetate-containing compounds (1 eq.) and hydrazines (2 eq.) in ethanol. After the reaction reached completion, the product was cooled to room temperature and filtered out of ethanol. The final products (**4**) were obtained in the third and last step (**ii**) through Schiff base reactions. Hydrazones were formed by reacting hydrazides and aldehydes/ketones in ethanol with drops of acetic acid under reflux conditions for 4–12 h. After cooling, the product was filtered and purified using column chromatography.

3.2.1. Synthesis of Ethyl 2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl) Oxy)Acetate (**2**)

A total of 250 mg of 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one (1.23 mmol) was stirred with 317.61 mg of the base K_2CO_3 (2.3 mmol) for 1 h in THF and drops of DMF. Then, 0.204 mL (1.51 g/cm^3) of ethyl 2-bromoacetate (1.5 mmol) was added after workup, and column chromatography (Hexane:EtOAc) (50:50) then afforded 313 mg (81%) as white crystals. m.p: 139.4, LC-MS: *m*/*z* 249.96 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d_6) δ 9.93 (s, 1H. N-H), 6.79–6.70 (m, $3H_{arom}$), 4.68 (s, 2H. O=C-C**H_2**-O), 4.15 (q, *J* = 7.1 Hz, 2H.

O-C**H$_2$**-CH$_3$), 2.82 (t, *J* = 7.5 Hz, 2H. O=C-CH$_2$-C**H$_2$**-), 2.39 (m, 2H. 2H. O=C-C**H$_2$**-), 1.20 (t, *J* = 7.1 Hz, 3H. C**H$_3$**). **^{13}C-NMR** (126 MHz, DMSO) δ 172.07, 169.05, 153.81, 131.93, 125.13, 116.47, 114.93, 113.41, 65.97, 61.44, 30.54, 25.66, 14.24.

3.2.2. Synthesis of 2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**3**)

A total of 300 mg of compound **2** (1.2 mmol) was heated with hydrazine hydrate 99% (3.6 mmol) in ethanol at 70 °C. After workup, the product was obtained as white powder with a yield of 90%. m.p: 210–211 °C, **^{1}H-NMR** (500 MHz, DMSO-*d*$_6$) δ 9.92 (s, 1H. N-**H**), 9.27 (s, 1H. O=C-N-**H**), 6.81–6.73 (m, 3H$_{arom}$), 4.40 (s, 1H. O-C**H$_2$**-C=O), 4.31 (s, 2H. N**H$_2$**), 2.83–2.80 (m, 2H. O=C-C**H$_2$**-), 2.41–2.38 (m, 2H. O=C-CH$_2$-C**H$_2$**). **^{13}C-NMR** (126 MHz, DMSO) δ 170.22, 167.19, 153.44, 132.75, 125.22, 116.13, 114.76, 113.72, 67.16, 30.77, 25.56.

3.2.3. Synthesis of (E)-N′-(3-Methylbutylidene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4a**)

According to the mentioned general procedure, starting from 6-hydroxy-3,4-dihydroquinolin-2(*1H*)-one, 300 mg of 6-hydroxy-3,4-dihydroquinolin-2(1H)-one (1.84 mmol) was stirred with 381 mg of the base K_2CO_3 (2.76 mmol) for 15 min in THF and drops of DMF, then 244 µL of ethyl 2-bromoacetate (2.21 mmol) was added slowly to the reaction mixture. After the completion of the reaction, the solvent was evaporated and the residue was washed with water, filtered, and air-dried to obtain ethyl 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetate (**2**), which was further refluxed with 241 µL hydrazine hydrate 99% (4.81 mmol) for 5 h. It was then cooled to room temperature, filtered out of ethanol, and dried to obtain 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide, compound **3**. Later, 300 mg of compound **3** (1.28 mmol) was refluxed with 168 µL of isovaleraldehyde (1.53 mmol) in ethanol and drops of acetic acid for 5 h. After that, the precipitate was filtered out of ethanol to obtain the final product as white powder with a yield of 85%. m.p: 197–199 °C, LC-MS: *m*/*z* 304.11 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d6) δ 11.08 (s, 1H. C-NH-N=), 9.93 (* 9.95, s, 1H. C-NH-C), 7.63–6.72 m, 3H$_{\mathbf{arom}}$), 6.67–6.64 (m, 1H. N=C**H**-), 4.87 (s, 1H. O-**CH$_2$**-C=O), 4.48 (s, 1H. O-C**H$_2$**-C=O), 2.84–2.80 (m, 2H), 2.43–2.36 (m, 2H), 2.12–2.06 (m, 2H), 1.84–1.78 (m, 1H), 0.91 (d, J = 2.9 Hz, 3H), 0.90 (d, *J* = 2.9 Hz, 3H). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.80, 163.86, 151.89, 147.77, 132.43, 131.99, 124.75, 114.41, 113.28, 66.90, 40.53, 30.32, 26.14, 25.12, 22.29, 22.26.

3.2.4. Synthesis of (E)-N′-(4-Chlorobenzylidene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl) Oxy) Acetohydrazide (**4b**)

A total of 150 mg of compound **3** (0.8 mmol) was heated with 134.45 mg of 4-chlorobenzaldehyde (0.956 mmol) in ethanol under reflux with the presence of AcOH (2 drops). The desired molecule was precipitated when the reaction finished (controlled by TLC) and the solvent was evaporated, and the obtained compound was washed by ethanol to obtain 210 mg (92%) as white powder. m.p:244–246 °C, LC-MS: *m*/*z* 357.91 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-*d*$_6$) δ 11.62 (* 11.59, s, 1H. C-**NH**-N=), 9.96 (* 9.92, s, 1H. C-N**H**-C), 8.33 (* 7.99, s, 1H. N=C**H**-) 7.74–6.72 (m, 7H$_{\mathbf{arom}}$), 5.07 (s, 1H. O-**CH$_2$**-C=O), 4.60 (s, 1H. O-**CH$_2$**-C=O), 2.86–2.81 (m, 2H. C**H$_2$**CH$_2$-C=O), 2.42–2.38 (m, 2H. C**H$_2$**C=O). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 170.23, 164.97, 153.86, 147.04, 142.89, 134.81, 133.39, 132.46, 129.32 (2C), 129.04, 125.21, 116.13, 114.59, 113.52, 65.50, 30.81, 25.55.

3.2.5. Synthesis of N′-(2,4-Dimethoxybenzylidene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy) Acetohydrazide (**4c**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, compound 4c was prepared through 3 steps, **i**, **ii** and **iii**, respectively. A total of 200 mg of compound **3** (0.8 mmol) was refluxed with 0.16 mL of 2,4-dimethoxybenzaldehyde (1 mmol) in ethanol and drops of acetic acid. After filtration, a pure white powder product was obtained. m.p: 243–245 °C, LC-MS: *m*/*z* 383.97 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d_6) δ 11.39 (* 11.36, s, 1H. C-**NH**-N=), 9.95 (* 9.91, s, 1H. C-N**H**-C), 8.58 (* 8.25, s, 1H. N=C**H**-) 7.77–6.59 (m, 6H$_{\mathbf{arom}}$), 5.02 (s, 1H. O-**CH_2**-C=O), 4.55 (s, 1H. O-**CH_2**-C=O), 3.85 (s, 3H. –O-C**H_3**), 3.81 (s, 3H. –O-C**H_3**), 2.86–2.81 (m, 2H. C**H_2**CH_2-C=O), 2.43–2.38 (m, 2H. C**H_2**C=O). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 170.22, 164.39, 162.74, 159.45, 153.93, 143.91 (-N=**C**H-), 139.95, 132.41, 127.16, 125.20, 116.13, 114.56, 113.51, 106.91, 98.56, 65.54, 56.23 (-O-**C**H_3), 55.90 (-O-**C**H_3), 30.81, 25.55.

3.2.6. Synthesis 2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)-N′-(Thiophen-2-Ylmethylene)Acetohydrazide (**4d**)

A total of 161 mg of compound **3** (0.684 mmol) was heated with 134.45 mg of thiophene-2-carbaldehyde (0.821 mmol) in ethanol under reflux with the presence of AcOH (3 drops). The desired molecule was precipitated when the reaction finished (controlled by TLC) and the solvent was evaporated, and the obtained compound was washed by ethanol to obtain 200 mg (88.7%) as white powder. m.p: 223–225 °C, LC-MS: *m*/*z* 329.97 $[M + H]^+$. **^{1}H-NMR** (500 MHz, Chloroform-*d*) δ 11.54 (* 11.48, s, 1H. C-**NH**-N=), 9.95 (* 9.92, s, 1H. C-N**H**-C), 8.55 (* 8.18, s, 1H. N=C**H**-) 7.67–6.70 (m, 6H$_{\mathrm{arom}}$), 4.97 (s, 1H. O-**CH_2**-C=O), 4.57 (s, 1H. O-**CH_2**-C=O), 2.85–2.80 (m, 2H), 2.42–2.38 (m, 2H). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, Chloroform) δ 169.78, 164.26, 152.88, 143.06, 138.90, 132.48, 131.15, 129.10, 127.88, 124.87, 115.75, 114.39, 113.33, 66.98, 30.30, 25.10.

3.2.7. Synthesis (E)-N′-(Furan-2-Ylmethylene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4e**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, compound **4e** was prepared through 3 steps, **i**, **ii**, and **iii**, respectively. A total of 210 mg of compound 3 (0.892 mmol) was refluxed with 0.89 mL of furan-2-carbaldehyde (1.07 mmol, d = 1.16 g/mL) in ethanol and drops of acetic acid. After filtration, a pure white powder product was obtained. m.p: 217–218 °C, LC-MS: *m*/*z* 314.04 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-*d*) δ 11.50 (* 11.47, s, 1H. C-**NH**-N=), 9.95 (* 9.91, s, 1H. C-N**H**-C), 8.24 (* 7.88, s, 1H. N=C**H**-) 7.84–6.62 (m, 6H$_{\mathrm{arom}}$), 4.98 (s, 1H. O-**CH_2**-C=O), 4.58 (s, 1H. O-**CH_2**-C=O), 2.85–2.82 (m, 2H. O=C-CH_2-C**H_2**-), 2.42–2.38 (m, 2H. O=C-CH_2-C**H_2**-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.79, 164.37, 153.37, 148.97, 145.05, 137.72, 133.91, 132.02, 124.78, 115.69, 113.79, 113.06, 112.16, 64.88, 30.35, 25.10.

3.2.8. Synthesis (E)-N′-(4-Methoxybenzylidene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4f)**

A total of 190 mg of compound **3** (0.807 mmol) was heated with 0.118 mL of 4-methoxybenzaldehyde (0.97 mmol, d = 1.12 g/mL) in ethanol under reflux with the presence of AcOH (3 drops). The desired molecule was precipitated when the reaction finished (controlled by TLC). After workup, 217 mg (76%) of **4f** was obtained as yellowish white powder. m.p: 203–205 °C, LC-MS: *m*/*z* 354.22 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-*d*) δ 11.44 (* 11.39, s, 1H. C-**NH**-N=), 9.96 (* 9.92, s, 1H. C-N**H**-C), 8.28 (* 7.95, s, 1H. N=C**H**-) 7.65–6.72 (m, 7H$_{\mathrm{arom}}$), 5.05 (s, 1H. O-**CH_2**-C=O), 4.58 (s, 1H. O-**CH_2**-C=O), 3.80 (s, 3H. -O-C**H_3**) 2.86–2.81 (m, 2H. O=C-CH_2-C**H_2**-), 2.43–2.38 (m, 2H. O=C-CH_2-C**H_2**-). (* refer to

the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.78 (O=**C**<), 164.12 (O=**C**-NH=N-), 160.69, 153.46, 143.61 (N=**C**H-), 131.98, 128.51, 126.57, 124.76, 115.69, 114.34, 114.28, 114.11, 113.30, 113.05, 65.04 (O-**C**H_2-C=O), 55.31 (-**C**H_3), 30.36, 25.10.

3.2.9. Synthesis of (E)-N′-((2-Methyl-1H-Indol-3-Yl)Methylene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4g**)

Starting from 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**), 4g was prepared through 3 steps, respectively following the mentioned general procedure. A total of 200 mg of compound **3** (0.8 mmol) was refluxed with 162.41 mg of 2-Methylindole-3-carboxaldehyde (1 mmol) in ethanol and drops of acetic acid. After workup and recrystallisation by EtOH, 110 mg (34.5%) was obtained as white powder. m.p: 265–267 °C, LC-MS: *m*/*z* 377.12 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d_6) δ 13.69 (s, 1H. N-**H**), 12.45 (O=C-N**H**-N), 11.25 (s, 1H. N=C**H**-), 9.97 (s, 1H. C-N**H**-), 7.58–6.81 (m, 7 **H_{arom}**), 4.78 (s, 2H O-C**H_2**-), 2.91–2.78 (m, 2H. O=C-CH_2-C**H_2**-), 2.50 (s, 3H. –C**H_3**), 2.41 (t, *J* = 6.9 Hz, 2H. O=C-C-C**H_2**-). **^{13}C-NMR** (125 MHz, DMSO) δ 169.80 (O=**C**), 162.54 (O=**C**-NH-N=), 152.19, 142.67, 135.57, 131.99, 132.95, 125.06, 122.70, 121.07, 119.71, 115.84 (2C), 114.48, 113.47, 111.22 (2C), 67.45 (O-C**H_2**-C=O), 30.26, 25.21, 25.06 (-**C**H_3).

3.2.10. Synthesis of (E)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)-N′-(4-Propylbenzylidene)Acetohydrazide (**4h**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, **4h** was prepared through 3 steps, i, ii, and iii, respectively. A total of 210 mg of compound **3** (0.893 mmol) was refluxed with 0.158 mL of 4-propylbenzaldehyde (1.07 mmol, d = 1.005 g/mL) in ethanol and drops of acetic acid. After filtration, a pure white powder product was yielded (285 mg, 87.3%). m.p: 189–190 °C, LC-MS: *m*/*z* 366.20 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d_6) δ 11.50 (* 11.45, s, 1H. C-**NH**-N=), 9.96 (* 9.92, s, 1H. C-N**H**-C), 8.31 (* 7.97, s, 1H. N=C**H**-) 7.61–6.73 (m, 7H_{arom}), 5.05 (s, 1H. O-**CH_2**-C=O), 4.58 (s, 1H. O-**CH_2**-C=O), 2.85–2.81 (m, 2H. O=C-C**H_2**-), 2.57 (t, *J* = 7.5 Hz, 2H. O=C-CH_2-C**H_2**-), 2.43–2.36 (m, 2H. =C-C**H_2**-CH_2), 1.62–1.56 (m, 2H. –C**H_2**-CH_3), 0.88 (t, *J* = 8.6 Hz, 3H. CH_3). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.76, 164.27, 153.44, 147.96, 144.63, 143.81, 132.00, 131.56, 128.80, 127.12, 126.90, 124.75, 115.68, 114.12, 113.05, 65.05, 37.10, 30.36, 25.09, 23.92, 13.59.

3.2.11. Synthesis of (E)-N′-(Cyclohexylmethylene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4i**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, compound 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) was prepared through a couple of steps, following procedures **i** and **ii**, respectively. After that, following procedure **iii**, 180 mg of compound **3** (0.765 mmol) was stirred under reflux with 0.112 mL of cyclohexanecarbaldehyde (0.918 mmol) in ethanol and 3 drops of acetic acid. After workup, by extracting with DCM and water, 200 mg (79.3%) was obtained as white powder. m.p: 161–163 °C, LC-MS: *m*/*z* 329.99 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d6) δ 11.08 (* 11.01, s, 1H. C-**NH**-N=), 9.94 (* 9.90, s, 1H. C-N**H**-C), 7.52 (* 7.21, s, 1H. N=C**H**-) 6.82–6.65 (m, 3H_{arom}), 4.86 (s, 1H. O-**CH_2**-C=O), 4.47 (s, 1H. O-**CH_2**-C=O), 2.84–2.80 (m, 2H. O=C-C**H_2**-), 2.41–2.37 (m, 2H. O=C-CH_2-C**H_2**-), 2.21–2.15 (m, 1H. =CH-C**H**<), 1.76–1.60 (m, 5$H_{CycloHexane}$), 1.31–1.13 (m, 5$H_{CycloHexane}$). (* r efer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.75, 168.62, 163.85, 155.78, 153.40, 151.61, 131.96, 124.73, 115.66, 114.04, 112.99, 64.89, 30.34, 30.29, 29.60, 29.53, 25.50, 25.08, 24.96, 24.95.

3.2.12. Synthesis of (E)-N′-Benzylidene-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6 -Yl)Oxy)Acetohydrazide (**4j**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, compound 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) was prepared through a couple of steps, following procedures **i** and **ii**, respectively. After that, following procedure **iii**, 200 mg of compound **3** (0.850 mmol) was refluxed with 95 µL of benzaldehyde (0.935 mmol) in ethanol and drops of acetic acid for 6 h. After completing the reaction, the precipitate was filtered out of the hot ethanol, washed with hot ethanol, and dried to obtain 220 mg (80%) of the final product as white powder. m.p: 161–163 °C, LC-MS: m/z 324.11 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d_6) δ 11.56 (* 11.53, s, 1H. C-**NH**-N=), 9.96 (* 9.92, s, 1H. C-N**H**-C), 8.34 (* 8.01, s, 1H. N=C**H**-) 7.70–6.74 (m, 8H$_{arom}$), 5.07 (s, 1H. O-**CH$_2$**-C=O), 4.60 (s, 1H. O-**CH$_2$**-C=O), 2.86–2.81 (m, 2H. O=C-CH$_2$-C**H$_2$**-), 2.43–2.38 (m, 2H. O=C-CH$_2$-C**H$_2$**-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 170.26 (O=**C**<), 164.89 (O=**C**-NH=N-), 153.88, 148.39, 144.21 (N=**C**H-), 134.41, 132.44, 130.41, 129.30, 127.37 (2C), 125.23, 116.15, 114.59, 113.51, 65.51, 30.80, 25.54.

3.2.13. Synthesis of (Z)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)-N′-(2-Oxoindolin-3-Ylidene)Acetohydrazide (**4k**)

Starting from hydroxy-3,4-dihydroquinolin-2*(1H)*-one, **4k** was prepared through three steps, following procedures **i**, **ii**, and **iii**, respectively. In the first and second steps, compounds **2** and **3** were obtained, as previously mentioned. In the third step, 250 mg of 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) (1.06 mmol) was refluxed with 156 mg of isatin (1.06 mmol) in ethanol and 3 drops of acetic acid for 5 h. After completing the reaction, the product was filtered out the hot ethanol to obtain 335 mg of the final product as yellow powder with a yield of 86%. m.p: >300 °C, LC-MS: m/z 363.34 $[M - H]^-$. **^{1}H-NMR** (500 MHz, DMSO) δ 13.69 (s, 1H .O=C-N**H**-), 11.25 (s, 1H .C=N-N**H**-), 9.97 (s, 1H.N=N=C**H**-), 7.58–6.81 (m, 6**H**$_{arom.}$), 4.78 (s, 2H. O-C**H$_2$**-), 2.89–2.84 (m, 2H. O=C-C**H$_2$**-), 2.41 (t, J = 3 Hz, 2H. O=CH2-C**H$_2$**-). **^{13}C-NMR** (125 MHz, DMSO) δ 169.80, 162.54, 152.19, 142.67, 135.57, 131.99, 125.06, 122.70, 121.07, 119.71, 115.84 (2C), 114.48, 113.47, 111.22 (2C), 67.45, 30.26, 25.06.

3.2.14. Synthesis of (Z)-N′-(5-Fluoro-1-Methyl-2-Oxoindolin-3-Ylidene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4l**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, compound 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) was prepared through a couple of steps, following procedures **i** and **ii**, respectively. Next, 5-fluoro-1-methyl-2-indolin-2,3-dione was prepared according to procedure **i**, by reacting 5-fluoruisatin with methyl iodide under the basic condition of K_2CO_3 at room temperature. After that, following procedure (**ii**), 200 mg of compound **3** (0.850 mmol) was refluxed with 152 mg of 5-fluoro-1-methyl-2-indolin-2,3-dione (0.850 mmol) in ethanol and drops of acetic acid for 10 h. After completing the reaction, the precipitate was filtered out of the hot ethanol, washed with hot ethanol, and dried to obtain 275 mg (81%) of the final product as white powder. m.p: >300 °C, LC-MS: m/z 397.12 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d6) δ 13.74 (s, 1H .O=C-N**H**-), 11.35 (s, 1H .C=N-N**H**-), 7.51–6.78 (m, 6**H**$_{arom.}$), 4.80 (s, 2H. O-C**H$_2$**-), 3.38 (s, 3H. -CH$_3$), 2.89–2.86 (m, 2H. O=C-C**H$_2$**-), 2.48 (t, J = 3 Hz, 2H. O=CH$_2$-C**H$_2$**-). **^{13}C-NMR** (125 MHz, DMSO) δ 170.24, 163.10, 159.74, 157.85, 139.35, 125.46, 121.46, 121.39, 118.64, 116.28, 114.88, 113.89, 112.81, 112.74, 108.72, 108.50, 56.49, 30.72, 27.48, 25.51.

3.2.15. Synthesis of (Z)-N′-(5-Fluoro-2-Oxoindolin-3-Ylidene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4m**)

Starting from hydroxy-3,4-dihydroquinolin-2*(1H)*-one, **4m** was prepared through three steps, following procedure **i**, **ii**, and **iii**, respectively. In the first and second steps, compounds **2** and **3** were obtained as previously mentioned. In the third step, 250 mg of 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide, compound **3** (1.06 mmol), was refluxed with 175 mg of 5-flouroisatin (1.06 mmol) in ethanol and 3 drops of acetic acid for 9 h. After completing the reaction, the product was filtered out the hot ethanol to obtain 321 mg of the final product as yellow powder with a yield of 79%. m.p: >300 °C, LC-MS: *m*/*z* 381.10 $[M - H]^-$. **^{1}H-NMR** (500 MHz, DMSO-d6) δ 13.70 (s, 1H. O=C-N**H**-), 11.28 (s, 1H .C=N-N**H**-), 9.97 (s, 1H. O=C-N$\mathbf{H_{Indol}}$), 7.40–6.81 (m, 6$\mathbf{H}_{arom.}$), 4.80 (s, 2H. O-C$\mathbf{H_2}$-), 2.87–2.84 (m, 2H. O=C-C$\mathbf{H_2}$-), 2.41 (t, *J* = 3 Hz, 2H. O=CH_2-C$\mathbf{H_2}$-). **^{13}C-NMR** (125 MHz, DMSO) δ 170.24, 163.10, 159.74, 157.85, 139.35, 125.46, 121.46, 121.39, 118.64, 116.28, 114.88, 113.89, 112.81, 112.74, 108.72, 108.50, 56.49, 30.72, 25.51.

3.2.16. Synthesis of (E)-N′-(4-Nitrobenzylidene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4n**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, compound 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) was prepared through a couple of steps, following procedures **i** and **ii**, respectively. After that, following procedure **iii**, 200 mg of compound **3** (0.850 mmol) was refluxed with 151 mg of 4-nitrobenzaldehyde (0.850 mmol) in ethanol and drops of acetic acid for 6 h. After completing the reaction, the precipitate was filtered out of the hot ethanol, washed with hot ethanol, and dried to obtain 271 mg (86%) of the final product as white powder. m.p: 249–251 °C, LC-MS: *m*/*z* 367.21 $[M-H]^-$. **^{1}H-NMR** (500 MHz, DMSO-d_6) δ 11.86 (* 11.84, s, 1H. C-**NH**-N=), 9.97 (* 9.93, s, 1H. C-N**H**-C), 8.45 (* 8.25, s, 1H. N=C**H**-) 8.30–8.25 (m, 2H$_{\mathbf{arom}}$), 7.98–6.77 (m, 5H$_{\mathbf{arom}}$), 5.12 (s, 1H. O-$\mathbf{CH_2}$-C=O), 4.65 (s, 1H. O-$\mathbf{CH_2}$-C=O), 2.85–2.82 (m, 2H. O=C-CH_2-C$\mathbf{H_2}$-), 2.42–2.39 (m, 2H. O=C-CH_2-C$\mathbf{H_2}$-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 170.23 (O=**C**<), 165.38 (O=**C**-NH=N-), 153.81, 148.17, 145.89, 141.80, 140.72, 132.51, 128.32 (2C), 124.43 (2C), 116.14, 114.61, 113.52, 65.52 (O-**C**H_2-C=O), 30.80, 25.54.

3.2.17. Synthesis of (Z)-N′-(5-Nitro-2-Oxoindolin-3-Ylidene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4o**)

Starting from hydroxy-3,4-dihydroquinolin-2*(1H)*-one, TR-G056 was prepared through three steps, following procedures **i**, **ii**, and iii, respectively. In the first and second steps, compounds **2** and **3** were obtained as previously mentioned. In the third step, 250 mg of 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide, compound **3** (1.06 mmol), was refluxed with 204 mg of 5-nitroisatin (1.06 mmol) in ethanol and 3 drops of acetic acid for 9 h. After completing the reaction, the product was filtered out of the hot ethanol to obtain 387 mg of the final product as yellow powder with a yield of 86%. m.p: 95–97 °C, LC-MS: *m*/*z* 408.26 $[M-H]^-$. **^{1}H-NMR** (500 MHz, DMSO-d6) δ 13.40 (s, 1H. O=C-N**H**-), 11.30 (s, 1H. C=N-N**H**-), 10.02 (s, 1H. O=C-N$\mathbf{H_{Indol}}$), 7.43–6.84 (m, 6$\mathbf{H}_{arom.}$), 4.81 (s, 2H. O-C$\mathbf{H_2}$-), 2.88–2.86 (m, 2H. O=C-C$\mathbf{H_2}$-), 2.42 (t, *J* = 3 Hz, 2H. O=CH_2-C$\mathbf{H_2}$-). **^{13}C-NMR** (125 MHz, DMSO) δ 172.61, 163.14, 160.03, 158.03, 140.39, 126.49, 121.46, 121.41, 118.68, 117.23, 117.88, 115.89, 114.81, 113.74, 109.72, 109.50, 57.49, 29.72, 26.51.

3.2.18. Synthesis of (E)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)-N′-(Pyridin-2-Ylmethylene)Acetohydrazide (**4p**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, compound 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) was prepared through a couple

of steps, following procedures **i** and **ii**, respectively. After that, following procedure **iii**, 200 mg of compound **3** (0.850 mmol) was refluxed with 88 µL of 2-pyridinecarboxaldehyde (0.935 mmol) in ethanol and drops of acetic acid for 6 h. After completing the reaction, the precipitate was filtered out of the hot ethanol, washed with hot ethanol, and dried to obtain 220 mg (80%) of the final product as beige powder. m.p: 134–136 °C, LC-MS: *m*/*z* 325.2 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d_6) δ 11.84 (s, 1H. C-**NH**-N=), 9.97 (* 9.93, s, 1H. C-N**H**-C), 8.59 (* 8.37, s, 1H. N=C**H**-) 8.05–6.74 (m, 7H_{arom}), 5.09 (s, 1H. O-**CH_2**-C=O), 4.64 (s, 1H. O-**CH_2**-C=O), 2.84–2.81 (m, 2H. O=C-CH_2-C**H_2**-), 2.41–2.38 (m, 2H. O=C-CH_2-C**H_2**-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.91, 153.44, 152.88, 149.52, 144.25, 136.88, 132.07, 124.85, 124.40, 119.91, 115.77, 114.46, 114.20, 113.13, 65.08, 30.34, 25.13.

3.2.19. Synthesis of (E)-N′-((2-(4-Chlorophenyl)Pyrimidin-5-Yl)Methylene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4q**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, **4q** was prepared through 3 steps, following procedures **i**, **ii**, and **iii**, respectively. Starting from 6-hydroxy-3,4-dihydroquinolin-2(1*H*)-one, firstly, 300 mg of 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one (1.84 mmol) was stirred with 381 mg of the base K_2CO_3 (2.76 mmol) for 15 min in THF and drops of DMF, then 244 µL of ethyl 2-bromoacetate (2.21 mmol, d = 1.51 g/mL) was added slowly to the reaction mixture. After the completion of the reaction, the solvent was evaporated and the residue was washed with water, filtered, and air-dried to obtain ethyl 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetate (compound **2**), which was further refluxed with 156 µL hydrazine hydrate 99% (4.81 mmol) for 5 h, then cooled to room temperature, filtered out of ethanol, and dried to obtain 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide. Later, 200 mg of 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) (0.850 mmol) was refluxed with 278.83 mg of 2-(4-chlorophenyl)pyrimidine-5-carbaldehyde (1.28 mmol) in ethanol and drops of acetic acid for 6 h. After filtration, the product was obtained as pure white powder. m.p: 259–261 °C, LC-MS: *m*/*z* 436.81 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d_6) δ 11.96 (* 11.88, s, 1H. C-**NH**-N=), 9.96 (* 9.92, s, 1H. C-N**H**-C), 9.20–9.13 (m, 2$H_{pyrimidine}$), 8.42–8.39 (m, 2H_{arom}), 8.07 (* 8.04, s, 1H. N=C**H**-) 7.60–6.74 (m, 5H_{arom}), 5.07 (s, 1H. O-**CH_2**-C=O), 4.60 (s, 1H. O-**CH_2**-C=O), 2.86–2.81 (m, 2H. O=C-CH_2-C**H_2**-), 2.43–2.38 (m, 2H. O=C-CH_2-C**H_2**-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.77, 169.58, 162.18, 155.66, 153.38, 152.87, 142.62, 138.08, 136.12, 135.52, 132.51, 132.02, 129.56, 128.96, 126.30, 124.76, 115.68, 114.15, 113.11, 65.10, 30.36, 25.10.

3.2.20. Synthesis of (E)-N′-((2-Aminopyrimidin-5-Yl)Methylene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4r**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, compound 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) was prepared through a couple of steps, following procedures **i** and **ii**, respectively. After that, following procedure **iii**, 200 mg of compound **3** (0.850 mmol) was refluxed with 105 mg of 2-pyridinecarboxaldehyde (0.850 mmol) in ethanol and drops of acetic acid for 6 h. After completing the reaction, the precipitate was filtered out of the hot ethanol, washed with hot ethanol, and dried to obtain 231 mg (79%) of the final product as white powder. m.p: 284–286 °C, LC-MS: *m*/*z* 339.21 $[M-H]^-$. **^{1}H-NMR** (500 MHz, DMSO-*d*) δ 11.42 (* 11,40, s, 1H. C-**NH**-N=), 9.95 (* 9.90, s, 1H. C-N**H**-C), 8.55 (s, 1H. -N**H_2**), 8.51 (s, 1H. -N**H_2**), 8.16 (* 7.82, s, 1H. N=C**H**-), 7.17–6.71 (m, 5**H_{arom}**), 5.02 (s, 1H. O-**CH_2**-C=O), 4.55 (s, 1H. O-**CH_2**-C=O), 2.84–2.81 (m, 2H. O=C-CH_2-C**H_2**-), 2.42–2.37 (m, 2H. O=C-CH_2-C**H_2**-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.79, 163.70, 156.98, 153.46, 144.36, 139.97, 132.44, 131.94, 124.87, 116.91, 115.67, 114.11, 113.08, 65.05, 30.31, 25.10.

3.2.21. Synthesis of (E)-N′-(4-Fluorobenzylidene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4s**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, compound 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) was prepared through a couple of steps, following procedures **i** and **ii**, respectively. After that, following procedure **iii**, 200 mg of compound **3** (0.850 mmol) was refluxed with 100 µL of 4-flourobenzaldehyde (0.935 mmol) in ethanol and drops of acetic acid for 6 h. After completing the reaction, the precipitate was filtered out of the hot ethanol, washed with hot ethanol, and dried to obtain 250 mg (86%) of the final product as white powder. m.p: 238–240 °C, LC-MS: *m*/*z* 342.22 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-d6) δ 11.58 (* 11.54, s, 1H. C-**NH**-N=), 9.97 (* 9.93, s, 1H. C-N**H**-C), 8.35 (* 8.01, s, 1H. N=C**H**-) 7.79–6.73 (m, 7H_{arom}), 5.08 (s, 1H. O-**CH_2**-C=O), 4.61 (s, 1H. O-**CH_2**-C=O), 2.87–2.82 (m, 2H. O=C-CH_2-C**H_2**-), 2.43–2.39 (m, 2H. O=C-CH_2-C**H_2**-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.82, 164.47, 153.45, 146.84, 142.61, 132.49, 132.02, 129.37, 129.10, 124.79, 115.95, 115.78, 114.41, 114.15, 113.08, 65.07, 30.38, 25.12.

3.2.22. Synthesis of (E)-N′-(3-(4-Chlorophenoxy)Benzylidene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4t**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2(1*H*)-one, compound 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) was prepared through a couple of steps, following procedures **i** and **ii**, respectively. After that, following procedure **iii**, 200 mg of compound **3** (0.850 mmol) was refluxed with 198 mg of 3-(4-chlorophenoxy)benzaldehyde (0.850 mmol) in ethanol and drops of acetic acid for 6 h. After completing the reaction, the precipitate was filtered out of the hot ethanol, washed with hot ethanol, and dried to obtain 320 mg (83%) of the final product. m.p: 205–207 °C, LC-MS: *m*/*z* 450.92 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-*d*) δ 11.61 (* 11.56, s, 1H. C-**NH**-N=), 9.96 (* 9.92, s, 1H. C-N**H**-C), 8.32 (* 7.98, s, 1H. N=C**H**-) 7.49–6.70 (m, 11H_{arom}), 5.03 (s, 1H. O-**CH_2**-C=O), 4.59 (s, 1H. O-**CH_2**-C=O), 2.83–2.80 (m, 2H. O=C-CH_2-C**H_2**-), 2.41–2.37 (m, 2H. O=C-CH_2-C**H_2**-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.77, 164.52, 156.57, 155.44, 153.39, 152.88, 147.08, 142.87, 136.19, 132.47, 132.00, 130.71, 129.94, 127.39, 124.74, 122.72, 120.28, 116.74, 115.74, 114.13, 113.03, 65.03, 30.36, 25.10.

3.2.23. Synthesis of (E)-N′-(Naphthalen-1-Ylmethylene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4u**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, compound 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**) was prepared through a couple of steps, following procedures **i** and **ii**, respectively. After that, following procedure **iii**, 200 mg of compound **3** (0.850 mmol) was refluxed with 125 µL of 1-naphthaldehyde (0.935 mmol) in ethanol and drops of acetic acid for 6 h. After completing the reaction, the precipitate was filtered out of the hot ethanol, washed with hot ethanol, and dried to obtain 231 mg (72%) of the final product as white powder. m.p: 259–261 °C, LC-MS: *m*/*z* 374.28 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-*d*) δ 11.63 (* 11.59, s, 1H. C-**NH**-N=), 9.97 (* 9.93, s, 1H. C-N**H**-C), 9.02 (* 8.68, s, 1H. N=C**H**-) 8.81–6.75 (m, 10**H_{arom}**), 5.15 (s, 1H. O-**CH_2**-C=O), 4.66 (s, 1H. O-**CH_2**-C=O), 2.87–2.82 (m, 2H. O=C-CH_2-C**H_2**-), 2.43–2.38 (m, 2H. O=C-CH_2-C**H_2**-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.78, 164.46, 153.44, 147.88, 143.50, 133.50, 132.02, 130.43, 130.05, 129.22, 128.85, 127.39, 126.30, 125.56, 124.78, 123.75, 115.72, 114.49, 113.07, 65.24, 54.94, 30.31, 25.11.

3.2.24. Synthesis of (E)-N′-([1,1′-Biphenyl]-4-Ylmethylene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4v**)

According to procedure **i**, **ii**, and **iii**, respectively, starting from 6-hydroxy-3,4-dihydroquinolin-2(1*H*)-one, 300 mg of 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one (1.84 mmol) was stirred with 381 mg of the base K_2CO_3 (2.76 mmol) for 15 min in THF and drops of DMF, then 244 µL of ethyl 2-bromoacetate (2.21 mmol, d = 1.51 g/mL) was added slowly to the reaction mixture. After the completion of the reaction, the solvent was evaporated and the residue was washed with water, filtered, and air-dried to obtain ethyl 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetate (compound **2**), which was further refluxed with 241 µL hydrazine hydrate 99% (4.81 mmol) for 5 h, then cooled to room temperature, filtered out of ethanol, and dried to obtain 2-((2-oxo-1,2,3,4-tetrahydroquinolin-6-yl)oxy)acetohydrazide (compound **3**). Later, 210 mg of compound **3** (0.892 mmol) was refluxed with 211.47 mg of [1,1′-biphenyl]-4-carbaldehyde (1.16 mmol) in ethanol and drops of acetic acid for 6 h. After that, the precipitate was filtered out of ethanol to obtain 283 mg of the final product, as white powder with a yield of 79%. m.p: 224–226 °C, LC-MS: *m*/*z* 400.59 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-*d*) δ 11.59 (s, 1H .C-**NH**-N=), 9.97 (* 9.93, s, 1H. C-N**H**-C), 8.39 (* 8.05, s, 1H. N=C**H**-) 7.79–6.76 (m, 12H$_{\mathbf{arom}}$), 5.10 (s, 1H. O-**CH_2**-C=O), 4.61 (s, 1H. O-**CH_2**-C=O), 2.84–2.81 (m, 2H. O=C-CH_2-C**H_2**-), 2.42–2.38 (m, 2H. O=C-CH_2-C**H_2**-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.79 (O=**C**<), 164.43 (O=**C**-NH=N-), 153.45, 147.49, 143.35, 141.42, 139.32, 133.09, 132.48, 132.02, 129.04 (2C), 127.75, 127.06, 126.68 (2C), 124.78, 115.77, 115.71, 114.15, 113.08, 65.09, 30.31, 25.11.

3.2.25. Synthesis of (E)-N′-((4′-Chloro-[1,1′-Biphenyl]-4-Yl)Methylene)-2-((2-Oxo-1,2,3,4-Tetrahydroquinolin-6-Yl)Oxy)Acetohydrazide (**4w**)

Starting from 6-hydroxy-3,4-dihydroquinolin-2*(1H)*-one, **4w** was prepared through 3 steps following procedures **i**, **ii**, and **iii**, respectively. A total of 200 mg of compound **3** (0.850 mmol) was refluxed with 221 mg of 4′-chloro-[1,1′-biphenyl]-4-carbaldehyde (1.02 mmol) in ethanol and drops of acetic acid. After filtration, the product was obtained as pure off-white powder with a yield of 84% (310 mg). m.p: 279–281 °C, LC-MS: *m*/*z* 434.97 $[M + H]^+$. **^{1}H-NMR** (500 MHz, DMSO-*d*) δ 11.61 (s, 1H. C-**NH**-N=), 9.96 (* 9.92, s, 1H. C-N**H**-C), 8.38 (* 8.04, s, 1H. N=C**H**-) 7.79–6.73 (m, 11H$_{\mathrm{arom}}$), 5.09 (s, 1H. O-**CH_2**-C=O), 4.61 (s, 1H. O-**CH_2**-C=O), 2.86–2.81 (m, 2H. O=C-CH_2-C**H_2**-), 2.42–2.38 (m, 2H. O=C-CH_2-C**H_2**-). (* refer to the rotameric peak). **^{13}C-NMR** (125 MHz, DMSO) δ 169.77 (O=**C**<), 164.45 (O=**C**-NH=N-), 153.43, 147.33, 143.20, 140.00, 138.11, 133.43, 132.75, 132.47, 132.01, 128.98, 128.45, 128.44 (2C), 127.58, 126.96, 124.76, 115.69, 114.14, 113.07, 65.08 (O-**C**H_2-C=O), 30.36, 25.10.

3.3. Molecular Docking

To understand the molecular interaction mechanisms between the target proteins and the suitably repositioned therapeutic candidates, molecular docking investigations were carried out. The protein data bank's 3D coordinates for the VEGFR2 kinase domain protein were obtained (PDB ID: 2XIR). A computational docking study was performed using the virtual screening tool PyRx. This software uses Lamarck's Genetic Algorithm (LGA) as a scoring function in AutoDock Vina. The PyRx-virtual screening tool was used in this investigation to perform molecular docking interactions via the AutoDock VINA module [45]. Schrödinger Maestro and Pymol software were used to get and visualise the docked compounds with better binding affinity (kcal/mol) results. The protein structure was prepared, and only polar hydrogens were added, using the AutoDock Tools 1.5.6 module in PyRx tool, and the pdbqt files of proteins were generated. ChemBioDraw Ultra

14.0 was used to build each compound, charged and minimised by the MMFF94x force field, along with the adjustment of hydrogens and lone pairs [46]. The compounds were saved in the SDF format. The target protein active site pocket was docked with the synthetic derivatives and reference molecules. Based on the interaction profile and docking scores, the best suited docked conformation was chosen.

3.4. Molecular Dynamics Simulation

Molecular dynamics (MD) simulations were performed using Desmond via the Schrödinger Maestro interface, 13.8 version. These simulations followed the docking phase and involved an orthorhombic simulation cell filled with TIP3 water model molecules. Chloride ions were added to neutralise the overall charge of the complex. The simulation ran for 100 ns within an NPT ensemble, which maintained a constant number of atoms, pressure at 1.01325 bar, and temperature at 300 K. The Nose–Hoover chain method with a 1.0 ps interval was employed as the thermostat, while the Martyna–Tobias–Klein method with a 2.0 ps interval was used as the barostat. The Maestro simulation interaction diagram was utilised to analyse the MD simulation results.

3.5. In Silico ADME Study

A computational analysis of 23 novel analogues **4**(**a**–**w**) was carried out using the Schrödinger QikProp tool for prediction of ADME properties, SwissADME, and the PreADME online server, a computational toolbox, to determine which substances have the best possible pharmacokinetic characteristics or BBB permeability.

3.6. In Vitro Experiments on GBM Models

3.6.1. Cell Cultures

U87-MG (HTB-16) and U138-MG (HTB-14) lines were purchased from ATCC. The cells were cultured in DMEM-HG Medium (41966-029, Gibco, Grand Island, NE, USA) and supplemented with 10% fetal bovine serum (FBS, 10500-064, Gibco) and 1% Penicillin–Streptomycin (P/S, 15140-122, Gibco). After 24 h of seeding cells, all compounds dissolved in 0.4% dimethyl sulfoxide (DMSO, A3672-0250, AppliChem, Darmstadt, Germany) were treated in the cells.

3.6.2. MTT Assay

The proliferation effects of the drug candidates on the U87-MG and U138-MG cell lines were tested using the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay. A total of 7×10^3 cells per well of both U87-MG and U138-MG cells were seeded into a 96-well plate in triplicate for the assay. After 24 h, each compound was added at a dose of 10 μM and incubated for 48 h. After the incubation period, 10 μL of MTT solution (5 mg/mL in PBS) was added to each well. The MTT solution was aspirated from the wells after 2 h. Then, 100 μL of DMSO was added and mixed to dissolve the formazan crystals. Cell proliferation was analysed by measuring the absorbance of the dissolved formazan at a wavelength of 570 nm using a microplate reader (Epoch, San Jose, CA, USA). IC50 values were later calculated using eight drug concentrations (0, 1.25, 2.5, 5, 10, 20, 40, 80, and 160 μM). Each experiment was carried out in triplicate to ensure reliability and statistical significance.

3.6.3. PAMPA Assay

The PAMPA assay (BioAssay System, Hayward, CA 94545, USA) has been used for testing the permeability rates of drugs through the BBB-mimicking membrane. The permeability assay and analyses were performed according to the manufacturer's instructions. Briefly,

high permeability and low permeability standards, compounds (500 μM), and equilibrium standards (200 μM) were prepared. After the loading of lipid solutions into the pores, all compounds' solutions and the permeability controls were put onto the lipid membrane in donor wells. The permeability assay was performed for 18 h in a 37 °C incubator. All samples were collected from acceptor wells after the incubation period was finished, and the absorbance values were measured with equilibrium standards for normalisation.

4. Conclusions

The most frequent malignant primary brain tumour is glioblastoma. Patients with this illness often have a dismal prognosis, and long-term survival is rare.

The common treatment upon initial diagnosis typically includes surgical resection, followed by concurrent radiation therapy and temozolomide treatment. However, biological factors, such as the blood–brain barrier, besides the unique tumour and immune microenvironment, present significant challenges to the development of novel therapeutics compared to other solid tumours. To improve outcomes for glioblastoma patients, a combination of innovative clinical trial designs, targeted drug therapies, and biomarker enrichment approaches is required.

This study focuses on the synthesis of 23 novel analogues of 3,4-dihydroquinolin-2(1H)-one, aiming to offer potential therapeutic solutions targeting the VEGFR2 gene for glioblastoma treatment. The synthesis and evaluation of these analogues were carried out using a deliberate strategy of design and structural modifications to optimise their interactions with the VEGFR2 protein.

This research includes extensive in vitro analyses to assess the therapeutic efficacy of these compounds in two glioblastoma cell lines, U87-MG and U138-MG. The findings demonstrated that all compounds, particularly **4u**, **4t**, **4m**, and **4q**, exhibited strong binding affinity for VEGFR2, with 4m showing the highest binding energy values.

To further assess the stability of the ligand–protein complexes, molecular dynamics (MD) simulations were performed, analysing their root mean square deviation (RMSD) and root mean square fluctuation (RMSF). Additionally, in silico ADMET predictions suggested that the synthesised derivatives possess favourable drug-like properties.

The significance of this study lies in its introduction of promising compounds—specifically **4u** (IC_{50} = 7.96 μM), **4t** (IC_{50} = 10.48 μM), **4m** (IC_{50} = 4.20 μM), and **4q** (IC_{50} = 8.00 μM)—that demonstrate specificity toward VEGFR2 kinase. This specificity highlights a promising route for the development of novel therapeutic agents for glioblastoma treatment. The findings and general indications of this study contribute to the ongoing efforts to address the critical need for more effective glioblastoma treatments, advancing the field of cancer therapy.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ph18020233/s1.

Author Contributions: S.B. and N.A.: writing—original draft, conceptualization, methodology, synthesis, data analysis, interpretation, and review. J.A.: writing—original draft, methodology, in-silico experiments, investigation, visualization, and data curation. S.S.: methodology, synthesis, and data interpretation. Y.K.: synthesis and data interpretation. S.D. and B.Y.: formal analysis, visualization, and cell culture experiments. G.B.-A.: writing—original draft, formal analysis, visualization, and cell culture experiments. E.H.: supervision, review, and project administration. H.T. and A.M.: supervision, review, project administration, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Molius Therapeutics and Trustlife Therapeutics and The Knut and Alice Wallenberg Foundation by grant number [No. 72254].

Institutional Review Board Statement: Not applicable. This study did not involve humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors gratefully acknowledge the financial support from Molius Therapeutics and Trustlife Therapeutics and The Knut and Alice Wallenberg Foundation (No. 72254).

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

1. Sucu, B.O.; Koc, E.B.; Ipek, O.S.; Mirat, A.; Almas, F.; Guzel, M.A.; Dogan, B.; Uludag, D.; Karakas, N.; Durdagi, S.; et al. Design and synthesis of novel caffeic acid phenethyl ester (CAPE) derivatives and their biological activity studies in glioblastoma multiforme (GBM) cancer cell lines. *J. Mol. Graph. Model.* **2022**, *113*, 108160. [CrossRef]
2. Stackhouse, C.T.; Gillespie, G.Y.; Willey, C.D. Exploring the Roles of lncRNAs in GBM Pathophysiology and Their Therapeutic Potential. *Cells* **2020**, *9*, 2369. [CrossRef]
3. Ajay Bemis, G.W.; Murcko, M.A. Designing libraries with CNS activity. *J. Med. Chem.* **1999**, *42*, 4942–4951. [CrossRef] [PubMed]
4. Pardridge, W.M. Drug transport across the blood-brain barrier. *J. Cereb. Blood Flow Metab.* **2012**, *32*, 1959–1972. [CrossRef]
5. Shergalis, A.; Bankhead, A.; Luesakul, U.; Muangsin, N.; Neamati, N. Current challenges and opportunities in treating glioblastomas. *Pharmacol. Rev.* **2018**, *70*, 412–445. [CrossRef]
6. Bellettato, C.M.; Scarpa, M. Possible strategies to cross the blood–brain barrier. *Ital. J. Pediatr.* **2018**, *44*, 127–133. [CrossRef] [PubMed]
7. Lipinski, C.A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235–249. [CrossRef]
8. Cornelissen, F.M.; Markert, G.; Deutsch, G.; Antonara, M.; Faaij, N.; Bartelink, I.; Noske, D.; Vandertop, W.P.; Bender, A.; Westerman, B.A. Explaining Blood–Brain Barrier Permeability of Small Molecules by Integrated Analysis of Different Transport Mechanisms. *J. Med. Chem.* **2023**, *66*, 7253–7267. [CrossRef]
9. Thakkar, J.P.; Dolecek, T.A.; Horbinski, C.; Ostrom, Q.T.; Lightner, D.D.; Barnholtz-Sloan, J.S.; Villano, J.L. Epidemiologic and Molecular Prognostic Review of Glioblastoma. *Cancer Epidemiol. Biomark. Prev.* **2014**, *23*, 1985–1996. [CrossRef] [PubMed]
10. Stupp, R.; Mason, W.P.; Van Den Bent, M.J.; Weller, M.; Fisher, B.; Taphoorn, M.J.; Belanger, K.; Brandes, A.A.; Marosi, C.; Bogdahn, U.; et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma in elderly patients. *J. Neurooncol.* **2005**, *352*, 987–996.
11. Brown, D.; Chen, Z.-P.; Guo, C.; Yang, Q.; Li, J.; Wu, S.; Deng, M.; Du, X.; Sai, K.; Jiang, X.; et al. Phase 2 clinical trial of VAL-083 as first-line treatment in newly-diagnosed MGMT-unmethylated glioblastoma multiforme (GBM): Halfway report. *Glioma* **2019**, *2*, 167. [CrossRef]
12. Hegi, M.E.; Diserens, A.C.; Gorlia, T.; Hamou, M.F.; De Tribolet, N.; Weller, M.; Kros, J.M.; Hainfellner, J.A.; Mason, W.; Mariani, L.; et al. MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. *N. Engl. J. Med.* **2005**, *352*, 997–1003. [CrossRef]
13. Fan, C.-H.; Liu, W.-L.; Cao, H.; Wen, C.; Chen, L.; Jiang, G. O6-methylguanine DNA methyltransferase as a promising target for the treatment of temozolomide-resistant gliomas. *Cell Death Dis.* **2013**, *4*, e876. [CrossRef] [PubMed]
14. Hanashalshahaby, E.H.A.; Unaleroglu, C. Mannich bases as enone precursors for water-mediated efficient synthesis of 2,3,6-trisubstituted pyridines and 5,6,7,8-tetrahydroquinolines. *ACS Comb. Sci.* **2015**, *17*, 374–380. [CrossRef] [PubMed]
15. Hanashalshahaby, E.H.A.; Ünaleroğlu, C.; Ak Can, A.; Özgün, A.; Garipcan, B. Design, synthesis, and antitumor evaluation of novel methylene moiety-tethered tetrahydroquinoline derivatives. *Turkish J. Chem.* **2019**, *43*, 1552–1569. [CrossRef]
16. Trawally, M. Beyond the heart—Exploring the therapeutic potential of PDE3 inhibitors. *J. Res. Pharm.* **2023**, *27*, 2218–2241. [CrossRef]
17. Mardinoglu, A.; Palsson, B.Ø. Genome-scale models in human metabologenomics. *Nat. Rev. Genet.* **2025**, *26*, 123–140. [CrossRef]
18. Xu, X.; Wang, J.; Han, K.; Li, S.; Xu, F.; Yang, Y. Antimalarial drug mefloquine inhibits nuclear factor kappa B signaling and induces apoptosis in colorectal cancer cells. *Cancer Sci.* **2018**, *109*, 1220–1229. [CrossRef]
19. Baudry, S.; Pham, Y.T.; Baune, B.; Vidrequin, S.; Crevoisier, C.; Gimenez, F.; Farinotti, R. Stereoselective passage of mefloquine through the blood-brain barrier in the rat. *J. Pharm. Pharmacol.* **1997**, *49*, 1086–1090. [CrossRef] [PubMed]
20. Golden, E.B.; Cho, H.-Y.; Hofman, F.M.; Louie, S.G.; Schönthal, A.H.; Chen, T.C. Quinoline-based antimalarial drugs: A novel class of autophagy inhibitors. *Neurosurg. Focus* **2015**, *38*, E12. [CrossRef]

21. Kim, M.S.; Yoo, B.C.; Yang, W.S.; Han, S.Y.; Jeong, D.; Song, J.M.; Kim, K.H.; Aravinthan, A.; Kim, J.H.; Kim, J.-H.; et al. Src is the primary target of aripiprazole, an atypical antipsychotic drug, in its anti-tumor action. *Oncotarget* **2018**, *9*, 5979–5992. [CrossRef]
22. Zhuo, C.; Xun, Z.; Hou, W.; Ji, F.; Lin, X.; Tian, H.; Zheng, W.; Chen, M.; Liu, C.; Wang, W.; et al. Surprising anticancer activities of psychiatric medications: Old drugs offer new hope for patients with brain cancer. *Front. Pharmacol.* **2019**, *10*, 1262. [CrossRef] [PubMed]
23. Tan, S.K.; Jermakowicz, A.; Mookhtiar, A.K.; Nemeroff, C.B.; Schürer, S.C.; Ayad, N.G. Drug repositioning in glioblastoma: A pathway perspective. *Front. Pharmacol.* **2018**, *9*, 218. [CrossRef] [PubMed]
24. Forno, F.; Maatuf, Y.; Boukeileh, S.; Dipta, P.; Mahameed, M.; Darawshi, O.; Ferreira, V.; Rada, P.; García-Martinez, I.; Gross, E.; et al. Aripiprazole cytotoxicity coincides with activation of the unfolded protein response in human hepatic cells. *J. Pharmacol. Exp. Ther.* **2020**, *374*, 452–461. [CrossRef]
25. Liu, C.-C.; Wu, C.-L.; Yeh, I.-C.; Wu, S.-N.; Sze, C.; Gean, P.-W. Cilostazol eliminates radiation-resistant glioblastoma by re-evoking big conductance calcium-activated potassium channel activity. *Am. J. Cancer Res.* **2021**, *11*, 1148–1169.
26. Farzaneh Behelgardi, M.; Zahri, S.; Mashayekhi, F.; Mansouri, K.; Asghari, S.M. A peptide mimicking the binding sites of VEGF-A and VEGF-B inhibits VEGFR-1/-2 driven angiogenesis, tumor growth and metastasis. *Sci. Rep.* **2018**, *8*, 17924. [CrossRef]
27. Batchelor, T.T.; Duda, D.G.; di Tomaso, E.; Ancukiewicz, M.; Plotkin, S.R.; Gerstner, E.; Eichler, A.F.; Drappatz, J.; Hochberg, F.H.; Benner, T.; et al. Phase II Study of Cediranib, an Oral Pan–Vascular Endothelial Growth Factor Receptor Tyrosine Kinase Inhibitor, in Patients With Recurrent Glioblastoma. *J. Clin. Oncol.* **2010**, *28*, 2817–2823. [CrossRef]
28. Bai, R.-Y.; Staedtke, V.; Riggins, G.J. Molecular targeting of glioblastoma: Drug discovery and therapies. *Trends Mol. Med.* **2011**, *17*, 301–312. [CrossRef]
29. Brandes, A.A.; Bartolotti, M.; Tosoni, A.; Poggi, R.; Franceschi, E. Practical Management of Bevacizumab-Related Toxicities in Glioblastoma. *Oncologist* **2015**, *20*, 166–175. [CrossRef]
30. Friedman, H.S.; Prados, M.D.; Wen, P.Y.; Mikkelsen, T.; Schiff, D.; Abrey, L.E.; Yung, W.A.; Paleologos, N.; Nicholas, M.K.; Jensen, R.; et al. Bevacizumab Alone and in Combination With Irinotecan in Recurrent Glioblastoma. *J. Clin. Oncol.* **2009**, *27*, 4733–4740. [CrossRef]
31. Keunen, O.; Johansson, M.; Oudin, A.; Sanzey, M.; Rahim, S.A.A.; Fack, F.; Thorsen, F.; Taxt, T.; Bartos, M.; Jirik, R.; et al. Anti-VEGF treatment reduces blood supply and increases tumor cell invasion in glioblastoma. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3749–3754. [CrossRef]
32. Ahmed, S.I.; Thomas, A.L.; Steward, W.P. Vascular endothelial growth factor (VEGF) inhibition by small molecules. *J. Chemother.* **2004**, *4*, 59–63. [CrossRef]
33. Wilhelm, S.M.; Adnane, L.; Newell, P.; Villanueva, A.; Llovet, J.M.; Lynch, M. Preclinical overview of sorafenib, a multikinase inhibitor that targets both Raf and VEGF and PDGF receptor tyrosine kinase signaling. *Mol. Cancer Ther.* **2008**, *7*, 3129–3140. [CrossRef] [PubMed]
34. Aly, R.M.; Serya, R.A.; Amira, M.; Al-Ansary, G.H.; Abou El Ella, D.A. Quinoline-based small molecules as effective protein kinases inhibitors (Review). *J. Am. Sci.* **2016**, *12*, 10–32.
35. Elkaeed, E.B.; Yousef, R.G.; Khalifa, M.M.; Ibrahim, A.; Mehany, A.B.M.; Gobaara, I.M.M.; Alsfouk, B.A.; Eldehna, W.M.; Metwaly, A.M.; Eissa, I.H.; et al. Discovery of New VEGFR-2 Inhibitors: Design, Synthesis, Anti-Proliferative Evaluation, Docking, and MD Simulation Studies. *Molecules* **2022**, *27*, 6203. [CrossRef] [PubMed]
36. Alafeefy, A.M.; Ashour, A.E.; Prasad, O.; Sinha, L.; Pathak, S.; Alasmari, F.A.; Rishi, A.K.; Abdel-Aziz, H.A. Development of certain novel N-(2-(2-(2-oxoindolin-3-ylidene) hydrazinecarbonyl)phenyl)-benzamides and 3-(2-oxoindolin-3-ylideneamino)-2-substituted quinazolin-4(3H)-ones as CFM-1 analogs: Design, synthesis, QSAR analysis and anticancer activity. *Eur. J. Med. Chem.* **2015**, *92*, 191–201. [CrossRef]
37. Liao, W.; Xu, C.; Ji, X.; Hu, G.; Ren, L.; Liu, Y.; Li, R.; Gong, P.; Sun, T. Design and optimization of novel 4-(2-fluorophenoxy)quinoline derivatives bearing a hydrazone moiety as c-Met kinase inhibitors. *Eur. J. Med. Chem.* **2014**, *87*, 508–518. [CrossRef]
38. McTigue, M.; Wickersham, J.; Pinko, C.; Hong, Y.; Marrone, T. Crystal Structure of the VEGFR2 Kinase Domain in Complex with PF-00337210 (N,2-dimethyl-6-(7-(2-morpholinoethoxy)quinolin-4-yloxy) Benzofuran-3-carb Oxamide). *wwpdb 2023*. Available online: https://www.wwpdb.org/pdb?id=pdb_00002xir (accessed on 20 December 2023).
39. Yao, X.; Ping, Y.; Liu, Y.; Chen, K.; Yoshimura, T.; Liu, M.; Gong, W.; Chen, C.; Niu, Q.; Guo, D.; et al. Vascular Endothelial Growth Factor Receptor 2 (VEGFR-2) Plays a Key Role in Vasculogenic Mimicry Formation, Neovascularization and Tumor Initiation by Glioma Stem-like Cells. *PLoS ONE* **2013**, *8*, e57188. [CrossRef]
40. Abdelgawad, M.A.; Hayallah, A.M.; Bukhari, S.N.A.; Musa, A.; Elmowafy, M.; Abdel-Rahman, H.M.; El-Gaber, M.K.A. Design, Synthesis, Molecular Modeling, and Anticancer Evaluation of New VEGFR-2 Inhibitors Based on the Indolin-2-One Scaffold. *Pharmaceuticals* **2022**, *15*, 1416. [CrossRef]

41. Nguyen, H.-M.; Guz-Montgomery, K.; Lowe, D.B.; Saha, D. Pathogenetic features and current management of glioblastoma. *Cancers* **2021**, *13*, 856. [CrossRef] [PubMed]
42. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2012**, *64*, 4–17. [CrossRef]
43. Athanasiou, C.; Cournia, Z. From Computers to Bedside: Computational Chemistry Contributing to FDA Approval. In *Biomolecular Simulations in Structure-Based Drug Discovery*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2018; pp. 163–203. [CrossRef]
44. Abad, N.; Sallam, H.H.; Al-Ostoot, F.H.; Khamees, H.A.; Al-Horaibi, S.A.; Khanum, S.A.; Madegowda, M.; El Hafi, M.; Mague, J.T.; Essassi, E.M.; et al. Synthesis, crystal structure, DFT calculations, Hirshfeld surface analysis, energy frameworks, molecular dynamics and docking studies of novel isoxazolequinoxaline derivative (IZQ) as anti-cancer drug. *J. Mol. Struct.* **2021**, *1232*, 130004. [CrossRef]
45. Dallakyan, S.; Olson, A.J. Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* **2015**, *1263*, 243–250. [PubMed]
46. Halgren, T.A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519. [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.

pharmaceuticals

MDPI

Article

From Gene Networks to Therapeutics: A Causal Inference and Deep Learning Approach for Drug Discovery

Sudhir Ghandikota [1] and Anil G. Jegga [1,2,*]

[1] Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA; sudhir.ghandikota@cchmc.org
[2] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA
* Correspondence: anil.jegga@cchmc.org; Tel.: +1-513-636-0261

Abstract: Background/Objectives: Drug discovery is a lengthy and expensive process, taking an average of 10 years and more than USD 2 billion from target discovery to drug approval. It is even more challenging in complex diseases due to disease heterogeneity and limited knowledge about the underlying mechanisms. We present a novel computational framework that integrates network analysis, statistical mediation, and deep learning to identify causal target genes and repurposable small-molecule candidates. **Methods**: We applied weighted gene co-expression network analysis (WGCNA) and bidirectional mediation analysis (causal WGCNA) to transcriptomic data from idiopathic pulmonary fibrosis (IPF) patients to identify genes causally linked to the disease phenotype. These genes were used as a phenotypic signature for deep learning-based compound screening using the DeepCE model. **Results**: Using RNA-seq data from 103 IPF patients and 103 controls, we identified seven significantly correlated modules and 145 causal genes. Five of these genes (*ITM2C, PRTFDC1, CRABP2, CPNE7,* and *NMNAT2*) were predictive of disease severity in IPF. Our compound screening identified several promising candidates, such as Telaglenastat (GLS1 inhibitor), Merestinib (MET kinase inhibitor), and Cilostazol (PDE3 inhibitor), with significant inverse correlation with the IPF-specific gene signature. **Conclusions**: This study demonstrates the utility of combining causal inference and deep learning for drug discovery. Our framework identified novel gene targets and therapeutic candidates for IPF, offering a scalable strategy for phenotype-driven drug discovery and repurposing.

Keywords: drug repurposing; drug repositioning; drug discovery; deep learning; mediation analysis; network analysis; idiopathic pulmonary fibrosis

1. Introduction

Target selection and engagement is a crucial first step in drug discovery, and a significant percentage of the clinical trial failures are linked to incorrect target selection and poor target engagement [1]. It is even more challenging in complex diseases because of the multifactorial nature and heterogeneity of these diseases; they are often driven by multiple genes, biological pathways, and environmental factors. The interconnected nature of these biological mechanisms makes it even more challenging to identify robust druggable targets. Additionally, disease progression can alter the relevance and importance of the selected target over time, making the target selection step even more critical in complex diseases.

Conventional drug discovery approaches are mostly based on a one-drug–one-gene paradigm [2,3], where compound screening is performed against a single, specific target [4].

Understandably, these approaches have limited success in complex diseases [3], as target-based screening methods often yield a poor correlation between modulating a single protein by a chemical compound and organism-level responses [2]. Furthermore, compound screening using single-gene targets often fails to recognize off-target interactions that can lead to unintended side effects [5]. On the other hand, phenotype-based compound screening represents a systematic approach by possibly capturing the full spectrum of a compound's biological activity [6]. They often begin with a phenotype-specific gene signature, which is queried against a reference database of small-molecule perturbations, typically using correlation measures. Moreover, deep learning (DL)-based frameworks can help mitigate the typical issues of noise and incompleteness associated with large-scale transcriptomics data [6].

We present a bioinformatic methodology that combines network analysis, statistical mediation models, and DL-based small-molecule compound screening frameworks for therapeutic drug discovery (See Materials and Methods). Our methodology starts with analyzing one (or more) transcriptomic datasets, converting them into gene co-expression networks, and applying a network analysis framework [7,8] to identify significantly correlated gene modules. Next, we employ statistical mediation techniques to examine the relationship between the phenotype and the candidate modules and identify candidate causal genes [9]. Instead of modeling the direct gene-phenotype associations, our approach enables us to capture high-level relationships between genes, network modules, and the phenotype, while adjusting for any clinical confounders. Finally, we use the significant mediator genes as the phenotypic signature and apply a mechanism-driven compound screening method to identify potential therapeutic candidates that can modulate this causal signature [2] (Figure 1). Compared to a list of dysregulated genes that are often used as the input signature in connectivity mapping approaches, causal genes make for more effective drug targets [10] as they represent high-level biological mechanisms driving phenotype progression. In contrast, differentially expressed gene (DEG) sets reflect downstream responses. The causal regulators provide potentially actionable insights for drug discovery.

To demonstrate the utility of our methodology, we used idiopathic pulmonary fibrosis (IPF) as a disease case study. IPF is a severe fibrotic lung disease characterized by chronic, progressive scarring, and destruction of the lung parenchyma, often requiring a lung transplant in advanced cases [11]. Two FDA-approved antifibrotic drugs, nintedanib and pirfenidone, are currently used as the first-line therapy for IPF. Even though both of them are effective in managing the symptoms of lung function decline in IPF patients, these drugs do not cure the disease and do not reverse the lung fibrosis [12,13]. Owing to the complexity, heterogeneity, and multifactorial nature of IPF pathogenesis, combination therapies are hypothesized to be more effective than monotherapies [14].

In this study, we applied our methodology and identified several novel candidate causal genes potentially important in the progression of IPF. Several of these were found to be significantly correlated with the forced vital capacity (FVC) and diffusing capacity of lungs for carbon monoxide (DLCO) lung function traits. Furthermore, we ran machine learning models to evaluate whether these mediator genes can predict the phenotype status using independent study cohorts. Finally, we ran a DL-based compound screening method to identify small-molecule candidates that can potentially modulate the IPF causal signature.

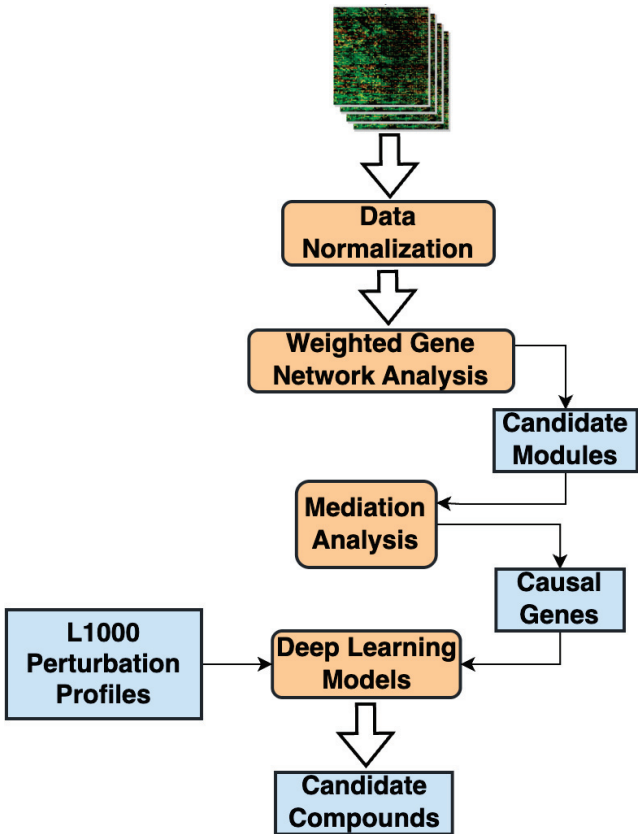


Figure 1. Overall workflow of our proposed compound screening methodology. Starting with a matrix of gene expression profiles as the input, the network analysis step identifies all gene modules co-expressed with respect to a phenotype. From the modules significantly associated with a phenotype, causal genes are identified by implementing statistical mediation analysis models. Finally, a multimodal deep learning model is used to identify candidate small-molecule compounds that potentially target the causal genes.

2. Results

2.1. Transcriptomic Datasets

In our IPF case study, we implemented a network analysis framework on a gene co-expression network, layered with a statistical mediation analysis. For constructing the gene network, we used RNA-seq data from whole lung tissues of 103 IPF samples and 103 healthy controls [15]. Additionally, we used two other IPF RNA-seq datasets, GSE124685 [16] and GSE213001 [17], to validate the biomarker candidates identified in the network analysis step (Table 1). All three publicly available datasets were downloaded from NCBI's Gene Expression Omnibus (GEO) repository [18]. For compound screening, we used the Library of Integrated Network-Based Cellular Signatures (LINCS) [19], a small-molecule gene expression database of drug perturbation profiles.

Table 1. List of IPF transcriptomic datasets used.

GEO Accession ID	# IPF	# Controls	Reference
GSE150910	103	103	[15]
GSE124685	49 (19 = mild (IPF1); 16 = moderate (IPF2); 14 = advanced (IPF3))	35	[16]
GSE213001	61 (22 = Advanced, 27 = Severe)	40	[17]

2.2. Correlated Gene Modules

Gene expression profiles from whole lung tissues (GSE150910) from both IPF and control samples were used for constructing gene co-expression networks and running the WGCNA algorithm [7]. Before this, we normalized the RNA-seq counts using the *voom* (variance modeling at the observational level) method [20]. We identified sixteen non-overlapping gene modules (Figure 2), out of which seven modules were significantly (Student asymptotic p-value < 0.05) correlated/upregulated in IPF (File S1). The *greenyellow* module (486 genes) was the most significant module (Figure 2) that contained genes predominantly involved in extracellular matrix organization (ECM) (GO:0030198) and collagen fibril organization (GO:0030199). Furthermore, five out of the seven correlated modules (*greenyellow, brown, yellow, red,* and *purple)* were strongly associated with lung function (FVC and DLCO) (Figure 2). Finally, we used these seven significantly correlated modules to run the mediation analysis to identify potential candidate causal genes.

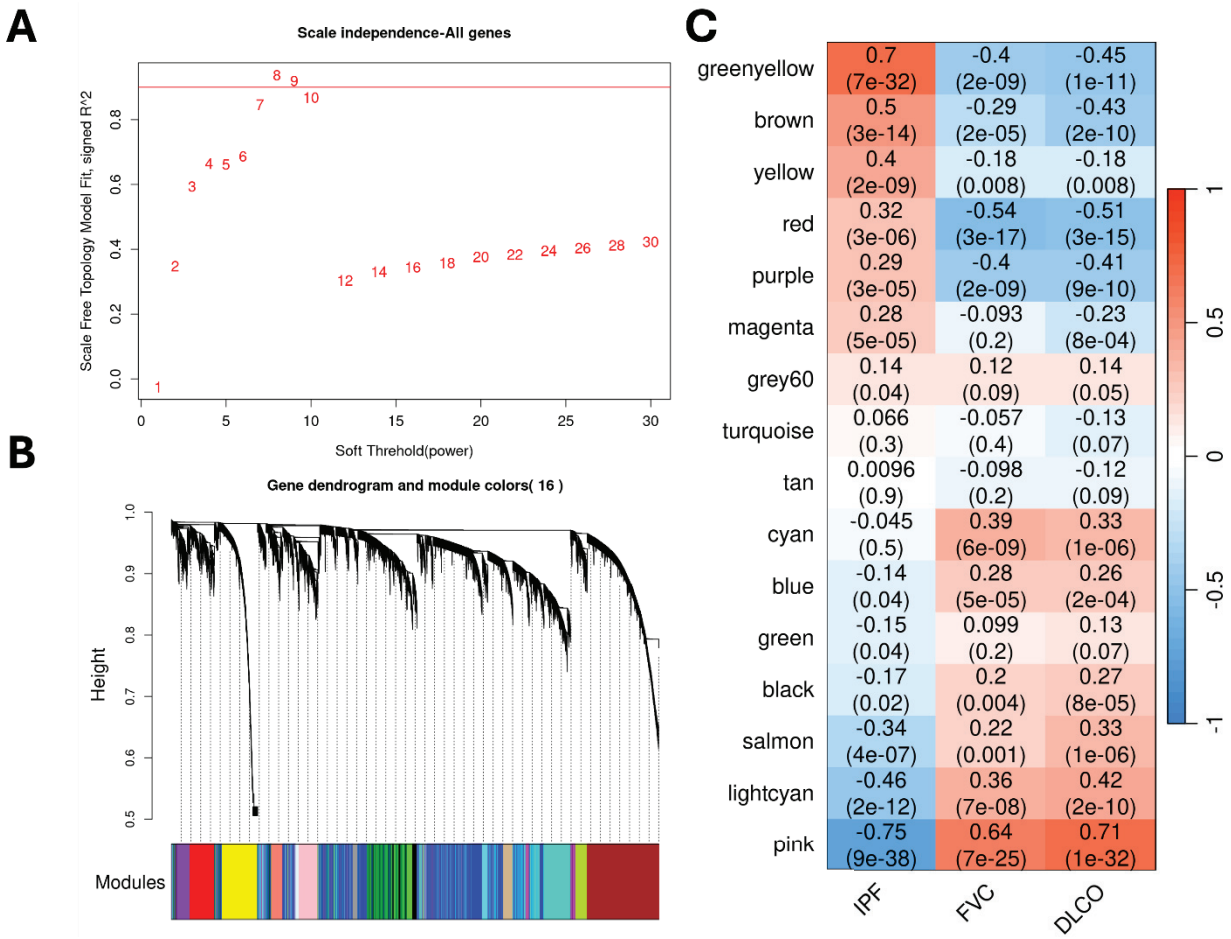


Figure 2. WGCNA Steps: (**A**) Scatter plot of the WGCNA scale-free analysis experiments on the IPF transcriptomic data. (**B**) Hierarchical dendrogram of gene modules constructed based on the topological overlap matrix (TOM). (**C**) Correlation heatmap of gene modules with the phenotypic status and the lung-function traits (FVC and DLCO) from the transcriptomic cohort.

2.3. Candidate Causal Genes in IPF

We implemented mediation analysis among the phenotype, the module, and the module genes using the CWGCNA framework [9]. Before this, using type-III ANOVA models, we tested for any potential confounding effects of several different clinical variables, including age, gender, and smoking status, and all mediation analysis models were adjusted

for statistically significant (age and smoking status) effects (See Materials and Methods). We applied bidirectional mediation models for each of the seven candidate WGCNA modules and identified 145 unique mediator genes with significant mediation that are potentially drivers (forward direction) of the disease progression in IPF (File S2). The majority of them (114 out of 145 genes) were found to be significantly upregulated ($\log_2$FC > 0.58 and adjusted p-value < 0.05) in IPF (Figure 3). Similarly, 101 out of the 145 genes were significantly associated (adjusted p-value < 0.05) with both FVC and DLCO lung function traits (File S2). We also found that 35/145 genes are part of the druggable genome collection, indicating their potential as therapeutic targets. This gene list represents the set of targets that are known or predicted to interact with drug compounds and was manually compiled from multiple sources, including DrugBank [21], DGIDb (v5.0) [22], and the CLUE-Drug repurposing hub [23].

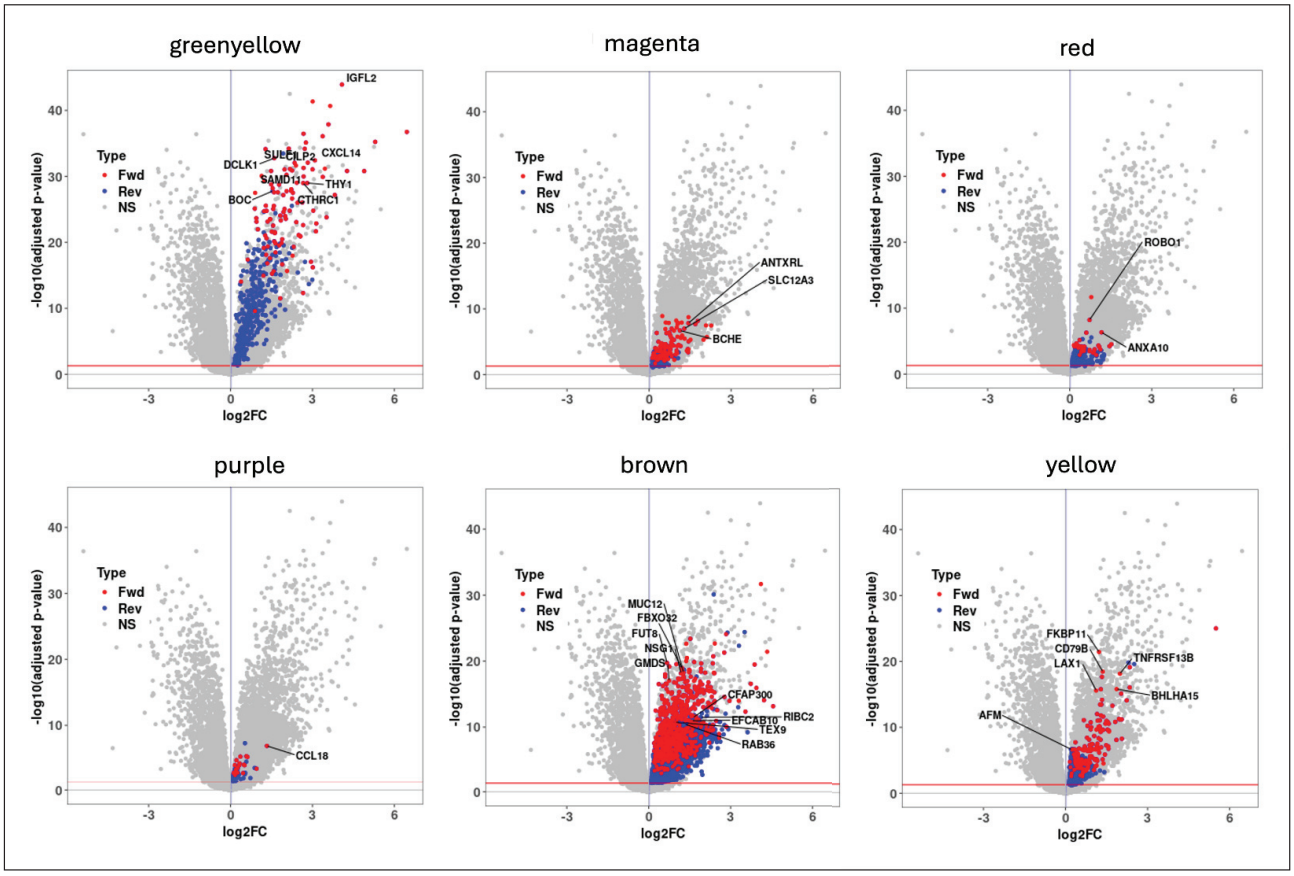


Figure 3. Volcano plots of the six candidate modules with significant mediator genes (red—forward; blue—reverse; gray—nonsignificant) highlighted in each of them.

Furthermore, we identified a significant (p-value < 0.05) number of mediator genes (37 out of 145 genes) that were previously known to be associated with the IPF phenotype. To do this, we compiled a list of known IPF-associated genes (EFO_0000768; 2910 genes) from the Open Targets Platform that mines and scores disease target genes from multiple evidence data sources, including GWAS associations, literature curation, and clinical evidence [24] (File S2). Additionally, we found several IPF lung single-cell markers (RAD51, CDKN3, TROAP, TAFA3, SGO1, NCAPH, CEP55, UTS2, ANLN, BUB1B, DGKI, and SKA3) among the remaining 108 causal genes from our mediation models.

2.4. Candidate Genes Enriched in Pro-Fibrotic Niches

To identify potential disease-driving mediator candidates localized in the lung tissue, we utilized a spatial transcriptomics study in IPF [25]. In this study, Mayr et al. identified three disease-associated niches (the fibrotic niche, the airway macrophage niche, and the immune niche) with unique cellular compositions and localizations. Intersecting the niche markers ($\log_2$FC > 0.25 and adjusted p-value < 0.05) with the significant mediator candidates, we identified four genes (CRABP2, MKI67, PRDX4, PLPP5) found in all three IPF niches (File S3). In one of our earlier studies, we demonstrated that CRABP2, a retinol-binding protein, is dysregulated in multiple lung cells of IPF and is strongly correlated with decline in lung function [26]. In that study, immunostaining of IPF lung sections demonstrated increased levels of CRABP2 in both spindle-shaped fibroblasts and airway epithelial cell types in the distal fibrotic lung lesions of IPF compared to control lungs. Similarly, studies analyzing human serum samples obtained from IPF patients have shown that PRDX4 (peroxiredoxin-4) is associated with the aggravation of inflammatory changes and fibrotic effects in the pathogenesis of IPF [27]. Furthermore, elevated levels of serum PRDX4 were associated with poor prognosis in murine models, demonstrating its potential clinical significance.

Additionally, we identified 24 genes (p-value < 0.05) belonging to the fibrotic niche, consisting of myofibroblasts and aberrant basaloid cells, located around the airways [25]. Similarly, we identified 20 genes (not statistically significant; p-value $= 0.3$) from the airway macrophage niche in the lumen containing $SPP1^+$ (Osteopontin) macrophages (File S3). Our mediator gene list also included 36 genes (p-value < 0.05) from the immune niche within IPF tissue, predominantly characterized by lymphoid cells surrounded by remodeled endothelial cells. In total, our mediation analysis identified 54 unique candidates that are marker genes in at least one of the three pro-fibrotic niches (File S3).

2.5. Mediator Genes Associated with IPF Severity

Next, to identify candidate causal genes associated with severe IPF phenotypes, we intersected the significant mediator candidates with the upregulated genes ($\log_2$FC > 0.58 and adjusted p-value < 0.05) in progressive-stage and end-stage IPF tissue from a lung RNA-seq dataset (GSE124685) [16]. The study used quantitative micro-CT imaging and tissue histology on IPF lung samples to categorize three distinct stages of fibrotic remodeling (early-stage or IPF1, progressive-stage or IPF2, and end-stage or IPF3). We identified 40 mediator genes (p-value $= 4.69 \times 10^{-13}$) that were upregulated in end-stage IPF and 38 candidate genes (p-value $= 8.07 \times 10^{-14}$) overexpressed in progressive-stage IPF samples. There were 30 mediator genes upregulated in both IPF2 and IPF3 samples. Additionally, we also found 40 candidate genes (p-value $= 2.88 \times 10^{-16}$) that were upregulated in early-stage IPF (File S4).

2.6. Biomarker Candidates in IPF

To assess the predictive ability of the mediator candidate genes and identify potential biomarker candidates in IPF, we ran generalized linear models (GLMs) via penalized maximum likelihood [28,29]. We used an independent IPF cohort (Table 1) consisting of 61 IPF cases (including 22 advanced and 27 severe cases) and 40 control samples to train the GLM models. These models were evaluated using the area under the precision-recall curves (AUPRC) and receiver operating characteristic (AUROC) curves (See Materials and Methods). We identified several candidate genes (ITM2C, PRTFDC1, CRABP2, CPNE7, FAM83D, NMNAT2, P4HA3, PDGFD, and PAPP2) with high area under the curve (AUC > 0.9) scores (Table 2). Among them, the integral membrane protein 2C (ITM2C) gene performed the best with high AUPRC and AUROC scores (Figure 4). The ITM2C gene is a known marker of plasmablast cells in the lung [30], which are implicated in the progression of IPF [31].

Table 2. Best-performing causal mediator genes in categorizing the IPF samples (including severe and advanced IPF) from healthy controls in an independent validation cohort (GSE213001). The table lists the AUPRC and AUROC scores averaged across 100 different experiments (both mean and standard deviation). The full list of AUC scores can be found in File S5.

	Gene	AUPRC	AUROC
IPF vs. Controls	ITM2C	0.993 ± 0.006	0.987 ± 0.01
	PRTFDC1	0.989 ± 0.007	0.983 ± 0.011
	CRABP2	0.984 ± 0.01	0.971 ± 0.018
	CPNE7	0.969 ± 0.018	0.955 ± 0.024
	FAM83D	0.965 ± 0.018	0.943 ± 0.029
	NMNAT2	0.964 ± 0.006	0.945 ± 0.026
	P4HA3	0.95 ± 0.03	0.933 ± 0.034
	PDGFD	0.945 ± 0.031	0.932 ± 0.036
	PAPPA2	0.943 ± 0.023	0.902 ± 0.04
Severe IPF vs. Controls	ITM2C	0.996 ± 0.004	0.997 ± 0.003
	CPNE7	0.97 ± 0.036	0.98 ± 0.026
	PRTFDC1	0.969 ± 0.024	0.975 ± 0.02
	CRABP2	0.96 ± 0.022	0.958 ± 0.025
	NMNAT2	0.943 ± 0.018	0.956 ± 0.014
	LAX1	0.943 ± 0.039	0.957 ± 0.029
	PAPPA2	0.92 ± 0.042	0.927 ± 0.045
Advanced IPF vs. Controls	ITM2C	0.996 ± 0.006	0.998 ± 0.004
	CRABP2	0.991 ± 0.013	0.994 ± 0.008
	PRTFDC1	0.983 ± 0.017	0.99 ± 0.01
	FAM83D	0.976 ± 0.02	0.986 ± 0.011
	NMNAT2	0.947 ± 0.033	0.968 ± 0.021
	MYOF	0.936 ± 0.033	0.955 ± 0.026
	P4HA3	0.927 ± 0.058	0.965 ± 0.022
	CDH3	0.923 ± 0.047	0.949 ± 0.029
	CPNE7	0.903 ± 0.058	0.936 ± 0.038

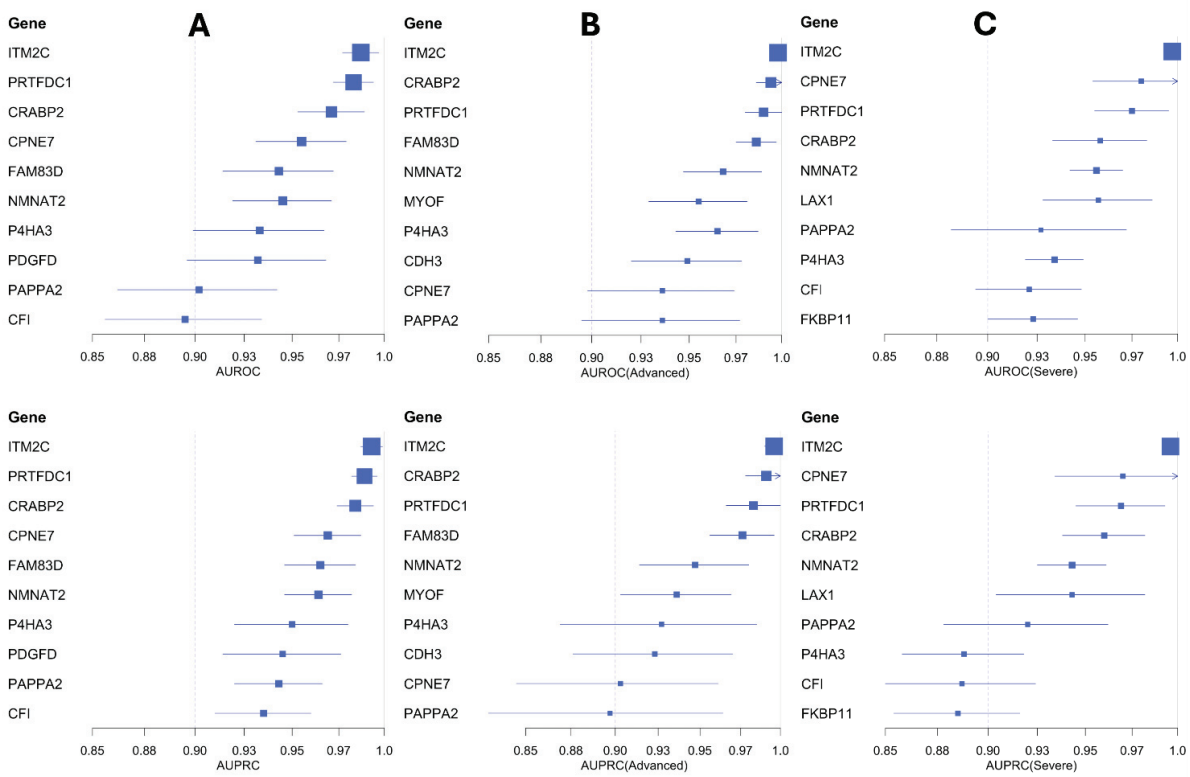


Figure 4. Forest plots of the top-performing biomarker candidates in classifying IPF samples. Panel (**A**) highlights the top biomarkers in classifying all IPF samples from the validation cohort (GSE213001) while Panels (**B**,**C**) include the best-performing biomarkers in categorizing advanced IPF and severe IPF samples, respectively. These biomarker genes are filtered based on the AUPRC (**bottom row**) and AUROC (**top row**) scores.

We also trained GLM models to distinguish severe and advanced IPF samples from controls (Figure 4). Based on the AUC scores, we identified seven gene candidates (ITM2C, CPNE7, PRTFDC1, CRABP2, NMNAT2, LAX1, and PAPPA2) with strong predictive performance in classifying severe IPF samples (Table 2). Similarly, we found nine additional gene candidates (ITM2C, CRABP2, PRTFDC1, FAM83D, NMNAT2, MYOF, P4HA3, CDH3, and CPNE7) highly capable of distinguishing advanced IPF samples from the healthy controls (Table 2 and File S5). Five of the mediator candidates (ITM2C, PRTFDC1, CRABP2, CPNE7, and NMNAT2) were strong predictive biomarkers across both sets of models (severe and advanced).

2.7. Small-Molecule Compounds Targeting the Causal Genes

Finally, we used a modified version of the DeepCE [2] model (See Materials and Methods) to identify compounds targeting the potentially causal genes in IPF. Specifically, we used 145 significant mediator genes (Section 2.3) as the query signature and implemented compound screening by computing Spearman's rank-order correlation between the drug perturbation gene expression profiles and the IPF-specific signature. We identified 126 small-molecule compounds with significant negative correlation scores (p-value < 0.05) in any of the seven different cell lines used in the model (File S6). Of these, 23 compounds (Table 3) exhibited significant negative correlations in multiple (two or more) cell lines.

Table 3. List of top compounds with significant negative correlation (Spearman rank-order) with the causal IPF signature. The correlations in each of the seven distinct cell lines, along with the significance p-values, are listed here (statistically significant correlations are highlighted in bold). File S6 has the full list of drugs.

Drug	A375	HA1E	HELA	HT29	MCF7	PC3	YAPC
1,4-Bis((3,4-dimethoxyphenyl)sulfonyl)-1,4-diazepane	−0.25 (0.051)	**−0.27 (0.036)**	**−0.34 (0.006)**	**−0.28 (0.027)**	**−0.33 (0.009)**	−0.25 (0.053)	**−0.28 (0.03)**
CB-839 (Telaglenastat)	**−0.29 (0.023)**	0.04 (0.76)	**−0.29 (0.02)**	**−0.26 (0.044)**	**−0.25 (0.049)**	**−0.28 (0.03)**	−0.18 (0.16)
[2-(4-Amino-1,2,5-oxadiazol-3-yl)-1-ethylimidazo[4,5-c]pyridin-7-yl]-[(3S)-3-aminopyrrolidin-1-yl]methanone	**−0.31 (0.014)**	−0.18 (0.17)	**−0.26 (0.041)**	**−0.32 (0.014)**	**−0.31 (0.015)**	**−0.29 (0.02)**	−0.19 (0.13)
Aminofurazanyl-azabenzimidazole 6n	**−0.35 (0.005)**	**−0.27 (0.03)**	**−0.3 (0.02)**	−0.23 (0.07)	**−0.29 (0.023)**	−0.25 (0.053)	−0.19 (0.15)
[2-(4-Amino-furazan-3-yl)-1-ethyl-1H-imidazo[4,5-c]pyridin-7-ylmethyl]-piperidin-4-yl-amine	**−0.28 (0.025)**	−0.1 (0.42)	−0.25 (0.052)	**−0.3 (0.02)**	**−0.28 (0.026)**	**−0.3 (0.02)**	−0.09 (0.51)
Pentamidine	**−0.31 (0.015)**	−0.16 (0.24)	−0.21 (0.098)	−0.21 (0.11)	**−0.26 (0.046)**	**−0.26 (0.044)**	**−0.27 (0.034)**
RHC-80267	−0.22 (0.089)	0.08 (0.52)	−0.18 (0.16)	**−0.35 (0.006)**	**−0.29 (0.023)**	−0.23 (0.075)	**−0.26 (0.039)**
RK-682	**−0.27 (0.033)**	0.12 (0.35)	0.019 (0.88)	0.22 (0.097)	−0.18 (0.17)	−0.22 (0.091)	**−0.29 (0.025)**
Merestinib	−0.08 (0.54)	−0.09 (0.48)	**−0.33 (0.008)**	−0.06 (0.64)	**−0.28 (0.03)**	−0.07 (0.57)	−0.05 (0.73)
LY-255283	−0.25 (0.053)	0.17 (0.2)	−0.18 (0.16)	**−0.26 (0.047)**	−0.2 (0.12)	**−0.28 (0.028)**	−0.07 (0.58)
Cilostazol	**−0.31 (0.015)**	0.018 (0.89)	−0.15 (0.25)	−0.16 (0.22)	−0.11 (0.41)	−0.06 (0.67)	**−0.28 (0.033)**
M2-PK-activator	−0.19 (0.15)	**−0.26 (0.04)**	−0.21 (0.11)	0.25 (0.05)	−0.22 (0.089)	−0.23 (0.07)	**−0.29 (0.02)**
NNC-711	−0.11 (0.4)	−0.03 (0.79)	−0.2 (0.11)	**−0.3 (0.018)**	**−0.27 (0.038)**	0.1 (0.44)	0.009 (0.95)
CID 11973736	−0.23 (0.077)	−0.2 (0.12)	**−0.33 (0.009)**	−0.24 (0.06)	**−0.32 (0.014)**	−0.18 (0.16)	−0.07 (0.6)

Table 3. *Cont.*

Drug	A375	HA1E	HELA	HT29	MCF7	PC3	YAPC
Azeloyl diethyl salicylate	−0.15 (0.26)	0.1 (0.45)	**−0.34 (0.008)**	**−0.32 (0.011)**	−0.16 (0.22)	−0.19 (0.13)	0.02 (0.88)
Cetrimonium	**−0.31 (0.015)**	−0.04 (0.78)	−0.15 (0.26)	0.23 (0.07)	−0.2 (0.13)	**−0.28 (0.027)**	−0.24 (0.06)
Decamethonium	**−0.26 (0.043)**	−0.04 (0.75)	−0.06 (0.65)	**−0.26 (0.047)**	−0.23 (0.072)	−0.23 (0.07)	−0.24 (0.056)
Gemcadiol	**−0.29 (0.021)**	−0.004 (0.98)	−0.06 (0.63)	**−0.25 (0.049)**	−0.14 (0.29)	−0.25 (0.056)	0.24 (0.066)
Clofilium	**−0.28 (0.026)**	0.23 (0.079)	−0.19 (0.14)	−0.25 (0.051)	−0.14 (0.29)	**−0.27 (0.03)**	−0.06 (0.64)
1-[[4,5-Bis(4-methoxyphenyl)-2-thiazolyl]carbonyl]-4-methylpiperazine	−0.13 (0.33)	−0.15 (0.24)	**−0.29 (0.025)**	−0.17 (0.19)	**−0.28 (0.029)**	−0.17 (0.19)	0.04 (0.78)
Zindotrine	−0.03 (0.82)	0.17 (0.18)	0.05 (0.68)	**−0.28 (0.03)**	−0.19 (0.15)	0.04 (0.74)	**−0.31 (0.015)**
Eliglustat	−0.04 (0.78)	−0.01 (0.94)	**−0.26 (0.04)**	**−0.27 (0.036)**	−0.1 (0.44)	−0.16 (0.22)	−0.01 (0.93)
3-(Azepan-1-ylsulfonyl)-N-(3-bromophenyl)benzamide	−0.02 (0.87)	−0.25 (0.054)	**−0.26 (0.04)**	−0.08 (0.54)	−0.11 (0.43)	−0.08 (0.53)	**−0.25 (0.047)**

Among them, the compound Telaglenastat (CB-839), a glutaminase 1 (GLS1) inhibitor, was significantly negatively correlated with our causal signature in five out of the seven cell lines. This compound has been shown to be therapeutically efficacious in treating both bleomycin- and transforming growth factor-β1-induced pulmonary fibrosis in mice [32]. GLS1 inhibition is also shown to regulate collagen production in human lung myofibroblasts by promoting its stabilization [32,33]. Collagen production plays a central role in the pathogenesis of IPF with excessive collagen deposition leading to lung tissue stiffening and impaired gas exchange. Similarly, the kinase inhibitor Merestinib targets the tyrosine kinase receptor MET (hepatocyte growth factor) and is being investigated as a potential anti-cancer therapy candidate [34]. The MET-signaling pathway has been implicated in the epithelial-mesenchymal transition (EMT) and fibroblast activation in IPF [35,36]. Consequently, MET has been proposed as a functional marker and an actionable target in IPF [37]. Our analysis also highlighted Cilostazol, a Phosphodiesterase 3 (PDE 3) inhibitor approved for use as an antiplatelet agent and vasodilator, for the symptomatic relief of intermittent claudication. Other studies have demonstrated its antifibrotic effects in a rat model of liver fibrosis [38] and cardiac fibrosis in mice [39]. We also identified the ROCK inhibitor Fasudil, which is known to ameliorate experimental pulmonary fibrosis by inhibiting mechanosensitive signaling in myofibroblasts [40] (File S6).

2.8. Causal Candidate Genes and Small-Molecule Targets

To understand the potential mechanism of actions (MOAs) associated with some of the top small-molecule compound hits explicitly to IPF, we computed functional enrichment analysis by combining the 145 mediator candidate genes and drug targets (Figure 5). Specifically, we obtained targets genes for Cilostazol (68 genes), Telaglenastat (5 genes), and Merestinib (16 genes) from the PubChem [41] database (File S7). In the combined gene set several IPF-related processes were enriched including fibroblast proliferation (12 genes; FDR B&H = 7.03×10^{-6}), fibroblast apoptotic process (7 genes; FDR B&H = 3.32×10^{-6}), and extracellular matrix organization (17 genes; FDR B&H = 2.82×10^{-5}). Additionally, we also observed enrichment of human lung cell type marker genes [42] from proliferating basal cells (21 genes; FDR B&H = 2.14×10^{-3}), adventitial fibroblasts (10 genes; FDR B&H = 8.49×10^{-3}), and differentiating basal cells (7 genes; FDR B&H = 0.011). Also,

we found 10 genes (FDR B&H = 4.56×10^{-8}) associated with IPF from the DisGeNET database [43].

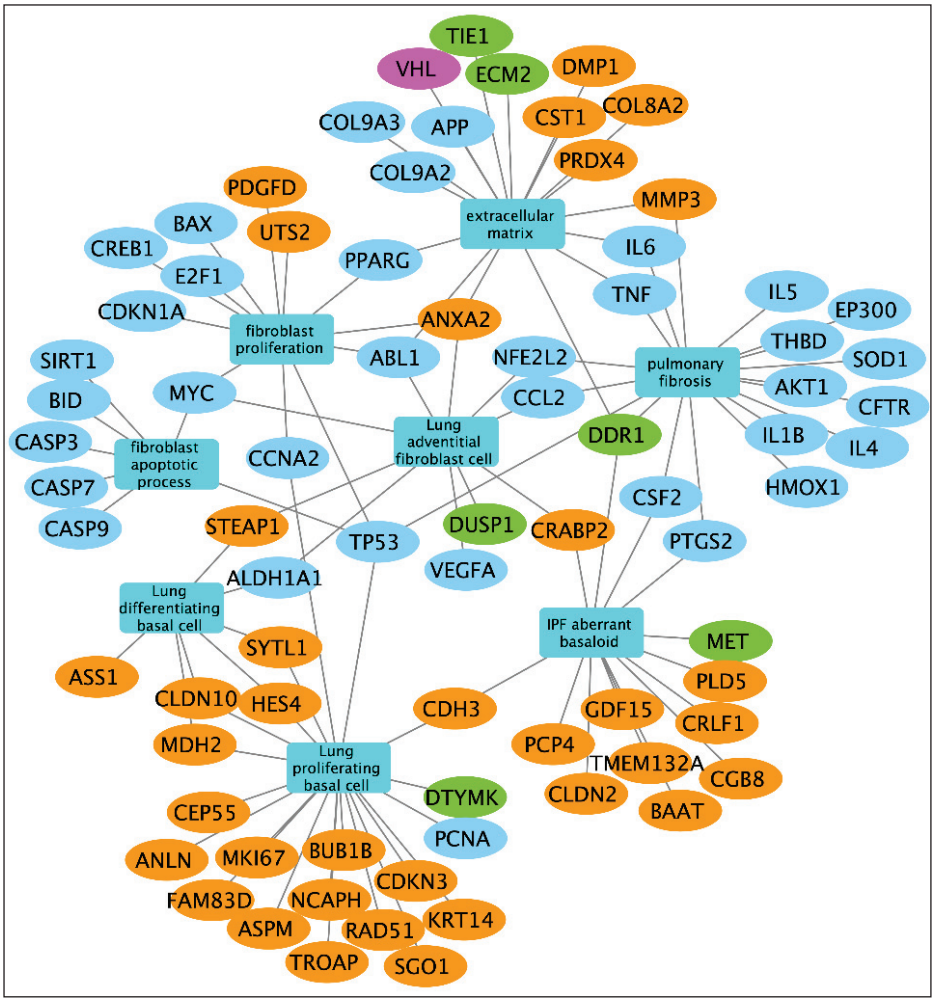


Figure 5. Functional enrichment network of significant mediator genes and drug targets of top hit small molecules. The mediator genes are represented as yellow elliptical nodes while the drug targets are represented in blue, green, or magenta colors. Targets genes of Cilostazol are highlighted in blue, while the green and magenta nodes represent the target genes of Merestinib and Telaglenastat, respectively. Enriched functional terms relevant to IPF are shown in turquoise-colored rectangular nodes. The network was generated using Cytoscape application (v3.10.3) and the functional enrichment was carried out using ToppFun application of the ToppGene Suite (https://toppgene.cchmc.org/; last accessed on 25 August 2025).

Furthermore, to illustrate the connectivity (protein–protein interactions—PPIs) between the mediator genes and the small-molecule targets, we also constructed a joint interactome (Figure S2). In total, we found 67 mediator candidates directly interacting with at least one drug target gene. We also observed a significant number of interactions between the targets genes of the three small-molecule compounds (Figure S2). We used the ToppGene tool suite (https://toppgene.cchmc.org/; last accessed on 25 August 2025) [44] to run the functional enrichment analysis and to obtain the interaction network (File S7). The network visualization was carried out using the Cytoscape application (v3.10.3) [45].

3. Discussion

In this study, we presented a therapeutic target and compound discovery approach by combining unsupervised network analysis, statistical mediation analysis, and Deep Learning (DL) models. We applied this framework to IPF and identified 16 non-overlapping gene modules, including 7 candidate modules significantly upregulated in the phenotype. From these modules, we identified 145 mediator gene candidates with significant mediation effects, after adjusting for confounders. These genes are hypothesized to *cause* the sample group differences by being involved in high-level pathways or biological processes (modules) [9]. Indeed, several of them were found to be localized in pro-fibrotic niches in the lung tissue. We also identified 35 known drug targets among these mediator genes. Additionally, machine learning models using these mediator genes demonstrated their predictive ability to categorize IPF samples from healthy controls and also for classifying severe and advanced IPF cases from controls with strong performance metrics.

The significant mediator genes were used as the phenotype signature for compound screening to identify small-molecule candidates that can potentially reverse the IPF phenotypic effects. By using Spearman rank-order correlations between the IPF signature and drug perturbation profiles, we identified more than 120 compounds with significant negative correlations with the IPF signature across different cell lines. Several of our identified candidate drugs have been reported to have therapeutic (antifibrotic) effects in in vitro and animal models. Functional enrichment analysis has shown several IPF-relevant pathways and processes shared between the causal mediator genes and the targets of the top hit small molecules. Further validation studies centered around some of the other small-molecule candidates are warranted to identify new therapeutic strategies for attenuating or reversing lung fibrosis.

Phenotype-based compound screening is generally considered systematic and comprehensive as it utilizes drug-induced gene expression profiles as readouts [2]. In comparison to previous methods, our methodology has the advantage of combining network analysis with causal inference, using bidirectional mediation models, to generate phenotype-specific signatures. This gives us the ability to potentially encode high-level causal mechanisms in these signatures. Consequently, the screened molecules are hypothesized to be safe and efficacious. Moreover, we used a customized DL model [2] for predicting robust drug perturbation profiles across multiple cell lines from otherwise noisy expression data (See Materials and Methods). Our customized model offers higher coverage by predicting expression profiles in the best inferred genes (BING) space instead of just the 978 landmark genes [19]. This enables capturing more comprehensive and relevant phenotypic signatures for a robust compound screening process. Our framework, including the experiments in the IPF use case, can be freely accessed from https://doi.org/10.5281/zenodo.16935145, accessed on 25 August 2025.

Our approach does hold certain limitations. Firstly, our mechanism-driven compound screening approach is highly dependent on the input signature and the quality of the gene expression profiles used in the analysis. Irrelevant target genes within the signature can lead to false-positive small-molecule hits. Secondly, our mediation models assume linear relationships between the phenotype and the modules and therefore cannot identify non-linear effects. Also, unmeasured confounding effects can impact the statistical mediation models, which in turn can influence the causal phenotype signatures used in the compound screening step. Hence, proper attention should be given to test and adjust for all possible confounding variables during the mediation analysis step. Additionally, noisy and unreliable perturbation profiles can lead to unsuccessful compound hits. To manage this issue, we utilized a DL model that employs an effective data augmentation technique, which can extract useful information from noisy or unreliable expression profiles [2]. Finally, the LINCS data is based on cancer cell lines and may not accurately represent the

physiology of non-cancerous tissues. Therefore, additional targeted validation experiments with relevant model systems are required to test the therapeutic hypotheses generated from implementing our framework.

As a future direction, we plan to develop compound screening approaches that can leverage single-cell expression profiles and spatially resolved cellular niches towards drug discovery applications. Using such frameworks, we can also include cell–cell interaction information to potentially capture and adjust for off-target activity of the identified drug compounds. With the help of interpretable DL models, non-linear relationships between drugs, targets, and phenotypes can be efficiently captured and better understood.

4. Materials and Methods

4.1. WGCNA and Candidate Correlated Modules

The weighted gene co-expression network analysis (WGCNA) framework is useful for identifying modules of highly co-expressed genes. Briefly, WGCNA uses pairwise Pearson correlations for constructing gene networks and converts them into topological overlap matrices that encode relative gene interconnectedness within the network [8]. Then, average linkage hierarchical clustering with adaptive branch pruning is implemented to identify modules of strongly co-expressed genes (Figure 2B) [46]. We used signed co-expression networks with the power parameter $(\beta = 8)$ chosen based on the scale-free topology criterion (Figure 2A) [8]. Additionally, the cluster sensitivity parameter (deepSplit) was set to the default value of 2 to achieve balanced clusters with a minimum module size set to 100 genes. Then, module eigengenes were identified to compute the module-level correlations with the phenotype status and any available clinical variables. These eigengenes are generally considered to be a module representative summarizing the expression patterns within a module and are identified through principal component analysis (PCA). Additionally, to assess the significance of the module-trait correlations, we calculated the Student asymptotic p-values for each correlation distribution (Figure 2C). Candidate modules were filtered using a significance threshold of p-value < 0.05.

4.2. Mediation Analysis and Candidate Causal Genes

We implemented statistical mediation analysis to reveal the causal relationships between the modules and the phenotype. Specifically, we used the CWGCNA package [9] to construct mediation models using the WGCNA module, the phenotype status, and the module genes. CWGCNA uses bidirectional models to test two causal directions: (a) the forward direction of the module driving the phenotype via the gene (module $\rightarrow$ gene $\rightarrow$ phenotype), and (b) the reverse direction of phenotype causing the module changes via the gene (phenotype $\rightarrow$ gene $\rightarrow$ module). These tests, conducted using the inverse probability weighting (IPW) method [47], help clarify the cause and the mediators of a given phenotypic condition. In both models, the module genes are used as the potential mediator and tested for their mediating effect. For each gene, the proportion mediated was computed by combining both the natural and the indirect effects, and 95% confidence intervals (CIs) were estimated via bootstrapping (100 times). Next, consistent mediator genes were chosen based on the significance and magnitude of the direct and indirect effect terms. Additionally, CWGCNA applied a bagging-SMOTE (synthetic minority oversampling technique) framework to prevent the sample size differences from influencing the mediation analysis step [9].

We applied mediation analysis models using the seven modules that showed positive correlations with the IPF phenotype (Figure 3). Before this, type-III ANOVA analysis was implemented to test for any significant confounding effects. Both gender and smoking status were determined to be significant confounding variables for the phenotype status. Next, we

tested the mediation effects of each module gene from the seven correlated modules, in both the forward and reverse directions. In total, we identified 145 genes with significant mediation effects uniquely in the forward direction (causal to phenotype). These genes were used in all the downstream analyses, including the phenotype-based compound screening step.

4.3. Machine Learning Models for Biomarker Analysis

For identifying a subset of biomarker candidate genes, we used generalized linear models (GLMs) that were trained with the elastic net penalty [28]. These models use the individual gene expression profiles as the input and are trained to predict the sample phenotype status. We implemented a leave-one-out cross-validation (LOOCV) approach to perform parameter tuning. In each training experiment, 70% of all samples were randomly picked for training the model and performing parameter tuning, and the remaining 30% of the test samples were used for model evaluation. The area under the curve (AUC) values of both precision-recall curves and the receiver operating characteristic (ROC) curves were computed to quantify the performance of these trained models. Furthermore, we conducted 100 distinct randomized experiments with separate training and validation data splits to compute robust evaluation metrics by averaging them across all the runs. Finally, in our IPF case study, we followed the same framework for models in both the training dataset (GSE150910) and the independent transcriptomic cohort (GSE213001).

4.4. DeepCE Model for Small-Molecule Screening

For screening small-molecule candidates, we followed the mechanism-driven approach used in [2]. Specifically, we used a modified version of the DeepCE model trained to predict the small-molecule perturbation profiles. It uses a graph neural network (GNN) to model the chemical sub-structure information [48] encoded in the form of a compound graph. Additionally, DeepCE also employs a network embedding model [49] for encoding gene–gene interaction networks [50]. It is trained using the expression profiles from the level 5 LINCS L1000 dataset (based on 978 landmark genes) [19]. DeepCE also includes experimental information, such as the cell line information and the drug dosage, in the model training step. Furthermore, a data augmentation technique is used, wherein a high-quality training set is first generated to train the model. The trained model is then used to predict and filter the gene expression profiles associated with the unreliable experiments in the L1000 data [2].

To increase the coverage of the small-molecule perturbation profiles, we trained a new DeepCE model using the level 5 L1000 data of 10,174 BING genes (978 landmark + 9196 well-inferred genes with high fidelity). We reused the original hyperparameter values and followed the same procedure to train the new model by selecting the gene expression profiles from the seven most popular cell lines (MCF7, A375, HT29, PC3, HA1E, YAPC, HELA) and the six most popular dosages (0.04 μm, 0.12 μm, 0.37 μm, 1.11 μm, 3.33 μm, 10.0 μm). We reused the high-quality signatures to train the new model and predicted the perturbation levels of approximately 2000 small molecules and nearly 6000 compounds from the BROAD Drug Repurposing Hub [23] across all seven cell lines. The model training was conducted on a HP Z4 workstation with Intel(R) Xeon(R) W-2133 CPU 3.60 GHz, 64 GB RAM, and Nvidia Quadro RTX 8000 GPU. For screening the candidate compounds, we use Spearman rank-order correlations between the causal phenotype signature (Section 4.2) and the predicted perturbation profiles. We also computed the p-values for these correlations using a hypothesis test whose null hypothesis is that the drug and the signature have no ordinal correlation. Finally, compounds that demonstrated significant negative correlations with the phenotype signature across multiple cell lines were shortlisted as likely drug repurposing candidates.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ph18091304/s1. Figure S1: Venn diagrams representing the overlaps between the significant mediators and the marker genes of three pro-fibrotic niches (Panel A) and the upregulated genes in IPF subtypes (Panel B). Figure S2: Interactome network of the significant mediators and the target genes of the top small-molecule candidates (Cilostazol, Telaglenastat, Merestinib). File S1: Network analysis results after implementing the WGCNA algorithm on the IPF RNA-seq data. File S2: Output file from the statistical mediation analysis step centered around the candidate modules chosen using a network analysis framework. File S3: Lists of significant mediator genes enriched in pro-fibrotic niches from a spatial transcriptomics study in IPF. File S4: Lists of mediator genes significantly upregulated in three distinct IPF subtypes (severities). File S5: Results from running generalized linear models (GLM) on an independent validation dataset (GSE213001) for identifying candidate biomarkers. File S6: List of small-molecule compounds (and their correlations with the IPF phenotype signature) shortlisted using the DeepCE DL model. File S7: Enriched functional terms and the joint interactome of significant mediator genes and the targets of the top small-molecule candidates shortlisted by the DL model.

Author Contributions: Conceptualization, S.G. and A.G.J.; methodology, S.G.; formal analysis, S.G.; investigation, S.G. and A.G.J.; resources, S.G. and A.G.J.; data curation, S.G. and A.G.J.; writing—original draft preparation, S.G.; writing—review and editing, S.G. and A.G.J.; visualization, S.G.; supervision, A.G.J. 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: All data supporting reported results can be found in the manuscript and supplementary material.

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

References

1. Sun, D.; Gao, W.; Hu, H.; Zhou, S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B* **2022**, *12*, 3049–3062. [CrossRef] [PubMed]
2. Pham, T.H.; Qiu, Y.; Zeng, J.; Xie, L.; Zhang, P. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. *Nat. Mach. Intell.* **2021**, *3*, 247–257. [CrossRef] [PubMed]
3. Xie, L.; Ge, X.; Tan, H.; Xie, L.; Zhang, Y.; Hart, T.; Yang, X.; Bourne, P.E. Towards Structural Systems Pharmacology to Study Complex Diseases and Personalized Medicine. *PLoS Comput. Biol.* **2014**, *10*, e1003554. [CrossRef] [PubMed]
4. Zhavoronkov, A.; Ivanenkov, Y.A.; Aliper, A.; Veselov, M.S.; Aladinskiy, V.A.; Aladinskaya, A.V.; Terentiev, V.A.; Polykovskiy, D.A.; Kuznetsov, M.D.; Asadulaev, A.; et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **2019**, *37*, 1038–1040. [CrossRef]
5. Lounkine, E.; Keiser, M.J.; Whitebread, S.; Mikhailov, D.; Hamon, J.; Jenkins, J.L.; Lavan, P.; Weber, E.; Doak, A.K.; Côté, S.; et al. Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **2012**, *486*, 361–367. [CrossRef]
6. Ghandikota, S.K.; Jegga, A.G. Chapter Seven—Application of artificial intelligence and machine learning in drug repurposing. In *Progress in Molecular Biology and Translational Science*; Singh, V., Ed.; Academic Press: Cambridge, MA, USA, 2024; pp. 171–211.
7. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef]
8. Zhang, B.; Horvath, S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.* **2005**, *4*, 17. [CrossRef]
9. Liu, Y. CWGCNA: An R package to perform causal inference from the WGCNA framework. *NAR Genom. Bioinform.* **2024**, *6*, lqae042. [CrossRef]
10. Porcu, E.; Sadler, M.C.; Lepik, K.; Auwerx, C.; Wood, A.R.; Weihs, A.; Sleiman, M.S.B.; Ribeiro, D.M.; Bandinelli, S.; Tanaka, T.; et al. Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nat. Commun.* **2021**, *12*, 5647. [CrossRef]

11. Kaur, A.; Mathai, S.K.; Schwartz, D.A. Genetics in Idiopathic Pulmonary Fibrosis Pathogenesis, Prognosis, and Treatment. *Front. Med.* **2017**, *4*, 154. [CrossRef]
12. Maher, T.M. Interstitial Lung Disease: A Review. *JAMA* **2024**, *331*, 1655–1665. [CrossRef] [PubMed]
13. Wang, Y.; Yella, J.K.; Ghandikota, S.; Cherukuri, T.C.; Ediga, H.H.; Madala, S.K.; Jegga, A.G. Pan-transcriptome-based candidate therapeutic discovery for idiopathic pulmonary fibrosis. *Ther. Adv. Respir. Dis.* **2020**, *14*, 1753466620971143. [CrossRef] [PubMed]
14. Kolb, M.; Bonella, F.; Wollin, L. Therapeutic targets in idiopathic pulmonary fibrosis. *Respir. Med.* **2017**, *131*, 49–57. [CrossRef] [PubMed]
15. Furusawa, H.; Cardwell, J.H.; Okamoto, T.; Walts, A.D.; Konigsberg, I.R.; Kurche, J.S.; Bang, T.J.; Schwarz, M.I.; Brown, K.K.; Kropski, J.A.; et al. Chronic Hypersensitivity Pneumonitis, an Interstitial Lung Disease with Distinct Molecular Signatures. *Am. J. Respir. Crit. Care Med.* **2020**, *202*, 1430–1444. [CrossRef]
16. McDonough, J.E.; Ahangari, F.; Li, Q.; Jain, S.; Verleden, S.E.; Herazo-Maya, J.; Vukmirovic, M.; DeIuliis, G.; Tzouvelekis, A.; Tanabe, N.; et al. Transcriptional regulatory model of fibrosis progression in the human lung. *JCI Insight* **2019**, *4*, e131597. [CrossRef]
17. Jaffar, J.; Wong, M.; Fishbein, G.A.; Alhamdoosh, M.; McMillan, L.; Gamell-Fulla, C.; Ng, M.; Wilson, N.; Symons, K.; Glaspole, I.; et al. Matrix metalloproteinase-7 is increased in lung bases but not apices in idiopathic pulmonary fibrosis. *ERJ Open Res.* **2022**, *8*, 00191–2022. [CrossRef]
18. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2013**, *41*, D991–D995. [CrossRef]
19. Subramanian, A.; Narayan, R.; Corsello, S.M.; Peck, D.D.; Natoli, T.E.; Lu, X.; Gould, J.; Davis, J.F.; Tubelli, A.A.; Asiedu, J.K.; et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **2017**, *171*, 1437–1452.e17. [CrossRef]
20. Law, C.W.; Chen, Y.; Shi, W.; Smyth, G.K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **2014**, *15*, R29. [CrossRef]
21. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672. [CrossRef]
22. Cannon, M.; Stevenson, J.; Stahl, K.; Basu, R.; Coffman, A.; Kiwala, S.; McMichael, J.F.; Kuzma, K.; Morrissey, D.; Cotto, K.; et al. DGIdb 5.0: Rebuilding the drug-gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Res.* **2024**, *52*, D1227–D1235. [CrossRef]
23. Corsello, S.M.; Bittker, J.A.; Liu, Z.; Gould, J.; McCarren, P.; Hirschman, J.E.; Johnston, S.E.; Vrcic, A.; Wong, B.; Khan, M.; et al. The Drug Repurposing Hub: A next-generation drug library and information resource. *Nat. Med.* **2017**, *23*, 405–408. [CrossRef] [PubMed]
24. Buniello, A.; Suveges, D.; Cruz-Castillo, C.; Llinares, M.B.; Cornu, H.; Lopez, I.; Tsukanov, K.; Roldan-Romero, J.M.; Mehta, C.; Fumis, L.; et al. Open Targets Platform: Facilitating therapeutic hypotheses building in drug discovery. *Nucleic Acids Res.* **2025**, *53*, D1467–D1475. [CrossRef] [PubMed]
25. Mayr, C.H.; Santacruz, D.; Jarosch, S.; Bleck, M.; Dalton, J.; McNabola, A.; Lempp, C.; Neubert, L.; Rath, B.; Kamp, J.C.; et al. Spatial transcriptomic characterization of pathologic niches in IPF. *Sci. Adv.* **2024**, *10*, eadl5473. [CrossRef] [PubMed]
26. Ghandikota, S.; Sharma, M.; Ediga, H.H.; Madala, S.K.; Jegga, A.G. Consensus Gene Co-Expression Network Analysis Identifies Novel Genes Associated with Severity of Fibrotic Lung Disease. *Int. J. Mol. Sci.* **2022**, *23*, 5447. [CrossRef]
27. Hanaka, T.; Kido, T.; Noguchi, S.; Yamada, S.; Noguchi, H.; Guo, X.; Nawata, A.; Wang, K.Y.; Oda, K.; Takaki, T.; et al. The overexpression of peroxiredoxin-4 affects the progression of idiopathic pulmonary fibrosis. *BMC Pulm. Med.* **2019**, *19*, 265. [CrossRef]
28. Friedman, J.H.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef]
29. Tay, J.K.; Narasimhan, B.; Hastie, T. Elastic Net Regularization Paths for All Generalized Linear Models. *J. Stat. Softw.* **2023**, *106*, 1–31. [CrossRef]
30. Zhang, T.; Hou, Z.; Ding, Z.; Wang, P.; Pan, X.; Li, X. Single Cell RNA-Seq Identifies Cell Subpopulations Contributing to Idiopathic Pulmonary Fibrosis in Humans. *J. Cell. Mol. Med.* **2025**, *29*, e70402. [CrossRef]
31. Prele, C.M.; Miles, T.; Pearce, D.R.; O'Donoghue, R.J.; Grainge, C.; Barrett, L.; Birnie, K.; Lucas, A.D.; Baltic, S.; Ernst, M.; et al. Plasma cell but not CD20-mediated B-cell depletion protects from bleomycin-induced lung fibrosis. *Eur. Respir. J.* **2022**, *60*, 2101469. [CrossRef]
32. Cui, H.; Xie, N.; Jiang, D.; Banerjee, S.; Ge, J.; Sanders, Y.Y.; Liu, G. Inhibition of Glutaminase 1 Attenuates Experimental Pulmonary Fibrosis. *Am. J. Respir. Cell Mol. Biol.* **2019**, *61*, 492–500. [CrossRef]
33. Ge, J.; Cui, H.; Xie, N.; Banerjee, S.; Guo, S.; Dubey, S.; Barnes, S.; Liu, G. Glutaminolysis Promotes Collagen Translation and Stability via alpha-Ketoglutarate-mediated mTOR Activation and Proline Hydroxylation. *Am. J. Respir. Cell Mol. Biol.* **2018**, *58*, 378–390. [CrossRef]

34. He, A.R.; Cohen, R.B.; Denlinger, C.S.; Sama, A.; Birnbaum, A.; Hwang, J.; Sato, T.; Lewis, N.; Mynderse, M.; Niland, M.; et al. First-in-Human Phase I Study of Merestinib, an Oral Multikinase Inhibitor, in Patients with Advanced Cancer. *Oncologist* **2019**, *24*, e930–e942. [CrossRef]
35. Crestani, B.; Marchand-Adam, S.; Quesnel, C.; Plantier, L.; Borensztajn, K.; Marchal, J.; Mailleux, A.; Soler, P.; Dehoux, M. Hepatocyte growth factor and lung fibrosis. *Proc. Am. Thorac. Soc.* **2012**, *9*, 158–163. [CrossRef] [PubMed]
36. Stella, G.M.; Gentile, A.; Balderacchi, A.; Meloni, F.; Milan, M.; Benvenuti, S. Ockham's razor for the MET-driven invasive growth linking idiopathic pulmonary fibrosis and cancer. *J. Transl. Med.* **2016**, *14*, 256. [CrossRef] [PubMed]
37. Stella, G.M.; D'Agnano, V.; Piloni, D.; Saracino, L.; Lettieri, S.; Mariani, F.; Lancia, A.; Bortolotto, C.; Rinaldi, P.; Falanga, F.; et al. The oncogenic landscape of the idiopathic pulmonary fibrosis: A narrative review. *Transl. Lung Cancer Res.* **2022**, *11*, 472–496. [CrossRef] [PubMed]
38. El Awdan, S.A.; Abdel Rahman, R.F.; Ibrahim, H.M.; Hegazy, R.R.; El Marasy, S.A.; Badawi, M.; Arbid, M.S. Regression of fibrosis by cilostazol in a rat model of thioacetamide-induced liver fibrosis: Up regulation of hepatic cAMP, and modulation of inflammatory, oxidative stress and apoptotic biomarkers. *PLoS ONE* **2019**, *14*, e0216301. [CrossRef]
39. Hada, Y.; Uchida, H.A.; Umebayashi, R.; Yoshida, M.; Wada, J. Cilostazol Attenuates AngII-Induced Cardiac Fibrosis in apoE Deficient Mice. *Int. J. Mol. Sci.* **2022**, *23*, 9065. [CrossRef]
40. Zhou, Y.; Huang, X.; Hecker, L.; Kurundkar, D.; Kurundkar, A.; Liu, H.; Jin, T.-H.; Desai, L.; Bernard, K.; Thannickal, V.J. Inhibition of mechanosensitive signaling in myofibroblasts ameliorates experimental pulmonary fibrosis. *J. Clin. Investig.* **2013**, *123*, 1096–1108. [CrossRef]
41. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2025 update. *Nucleic Acids Res.* **2025**, *53*, D1516–D1525. [CrossRef]
42. Travaglini, K.J.; Nabhan, A.N.; Penland, L.; Sinha, R.; Gillich, A.; Sit, R.V.; Chang, S.; Conley, S.D.; Mori, Y.; Seita, J.; et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **2020**, *587*, 619–625. [CrossRef] [PubMed]
43. Piñero, J.; Bravo, À.; Queralt-Rosinach, N.; Gutiérrez-Sacristán, A.; Deu-Pons, J.; Centeno, E.; García-García, J.; Sanz, F.; Furlong, L.I. DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **2017**, *45*, D833–D839. [CrossRef] [PubMed]
44. Chen, J.; Bardes, E.E.; Aronow, B.J.; Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **2009**, *37* (Suppl. 2), W305–W311. [CrossRef] [PubMed]
45. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]
46. Langfelder, P.; Zhang, B.; Horvath, S. Defining clusters from A hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* **2008**, *24*, 719–720. [CrossRef]
47. VanderWeele, T.J. Mediation Analysis: A Practitioner's Guide. *Annu. Rev. Public Health* **2016**, *37*, 17–32. [CrossRef]
48. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36. [CrossRef]
49. Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.
50. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S.; et al. The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646. [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.

pharmaceuticals

MDPI

Article

Resveratrol Alleviates Inflammatory Response Through P2X7/NLRP3 Signaling Pathway: In Silico and In Vitro Evidence from Activated Microglia

Bianca Fagan Bissacotti [1], Marcylene Vieira da Silveira [1], Charles Elias Assmann [1], Priscila Marquezan Copetti [1], André Flores dos Santos [2], Solange Binotto Fagan [2], João Augusto Pereira da Rocha [3], Maria Rosa Chitolina Schetinger [1], Vera Maria Melchiors Morsch [1], Nathieli Bianchin Bottari [4], Alencar Kolinski Machado [2] and Aleksandro Schafer da Silva [1,5,*]

[1] Graduate Program in Biological Sciences: Toxicological Biochemistry, Department of Biochemistry and Molecular Biology, Federal University of Santa Maria (UFSM), Santa Maria 97105-900, RS, Brazil; bianca_fbissacotti@hotmail.com (B.F.B.); marcyvieirass@outlook.com (M.V.d.S.); charles.ufsm@gmail.com (C.E.A.); priscilaa_marquezan@hotmail.com (P.M.C.); mariachitolina@gmail.com (M.R.C.S.); veramorsch@gmail.com (V.M.M.M.)

[2] Graduate Program in Nanosciences, Franciscan University (UFN), Santa Maria 97010-032, RS, Brazil; andreflores2009@gmail.com (A.F.d.S.); solange.fagan@gmail.com (S.B.F.); alencarkolinski@gmail.com (A.K.M.)

[3] Laboratory of Modeling and Computational Chemistry, Federal Institute of Education, Science and Technology of Pará (IFPA), Bragança Campus, Bragança 68600-000, PA, Brazil; joao.rocha@ifpa.edu.br

[4] Department of Microbiology and Parasitology, Federal University of Pelotas (UFPel), Pelotas 96010-610, RS, Brazil; nathieli_bb@hotmail.com

[5] Graduate Program in Animal Science, Santa Catarina State University (UDESC), Chapecó 89815-630, SC, Brazil

* Correspondence: dasilva.aleksandro@gmail.com

Abstract: Background/Objectives: Chronic inflammation and inappropriate NLRP3 inflammasome regulation are related to many brain diseases. Purinergic mediators may play an important role in inflammation regulation and could be targeted for effective therapies for these illnesses. We evaluated resveratrol's anti-neuroinflammatory potential in BV-2 microglia cells using an innovative in vitro method of NLRP3 inflammasome activation, correlating with the P2X7 purinergic receptor. **Methods**: In silico analyses were used to estimate resveratrol's interaction with NLRP3, and its cytotoxicity was measured for 24, 48, and 72 h. Moreover, microglia were exposed to lipopolysaccharide and nigericin to activate the NLRP3 inflammasome and treated with resveratrol between these inflammatory agents. **Results**: It was found that resveratrol has binding compatible with modulating NLRP3. Specifically, 0.1–25 μM of resveratrol presented a favorable safety profile in BV-2 cells. Microglia exposed to the inflammatory agents had increased levels of oxidative species, the P2X7 receptor, and pro-inflammatory cytokines. However, resveratrol decreased the NLRP3, caspase-1, IL-1β, IL-6, and TNF-α mRNA levels and protein density; on the other hand, IL-10 was increased, acting as a protector, preventing exacerbated inflammation. Under resveratrol exposure, P2X7 was negatively expressed, regulating inflammation to establish homeostasis and microglial proliferation. Additionally, resveratrol activates the A1 adenosine receptor, possibly correlated with neuroprotective effects. **Conclusions**: We confirmed the anti-neuroinflammatory action of resveratrol via the P2X7 receptor and NLRP3's combined modulation, regulating the cell cycle and reducing pro-inflammatory and oxidant agents. Considering this pathway, resveratrol could be a candidate for further investigations as a potential treatment against neuroinflammatory diseases.

Keywords: brain inflammation; BV-2 cells; immunomodulation; polyphenol; purinergicreceptors

1. Introduction

Inflammation is a natural organism's protective response against pathogens and stressors [1]. In the central nervous system (CNS), microglia and astrocytes are the primary immune defense cells, which can express the proteins that form the NOD-like receptor family pyrin domain-containing 3 (NLRP3) inflammasome and activate it [1–3].

The NLRP3 inflammasome is a multiprotein complex formed by the proteins NLRP3 and pro-caspase-1 and the apoptosis-associated adapter protein containing the CARD domain (ASC) [2]. The union and activation of these molecules involve the presence of pathogens (viruses, bacteria, parasites, fungi) or tissue damage (reactive oxygen or nitrogen species release, ROS and RNS, respectively). Immediately after NLRP3 activation, active caspase-1 (CASP-1) is released, followed by interleukin-1 beta (IL-1β) and interleukin-18 (IL-18), resulting in the subsequent pro-inflammatory cascade [2–5]. The priming and activation mechanisms are the primary regulators of NLRP3. Priming occurs through recognizing pathogen-associated molecular patterns (PAMPs), such as lipopolysaccharide (LPS) found in some bacteria, or damage-associated patterns (DAMPs), which increase the mRNA transcription of inflammatory components [6,7]. The entire NLRP3 activation process comprises assembly and activation steps. Decreases in the intracellular potassium and calcium ion levels and the production of ROS can act as activation signals. Nigericin (NIG) leads to the induction of potassium efflux, a potent activator of the NLRP3 inflammasome [7,8].

It is essential to mention that the purinergic system is an expressive contributor to the immune response and the NLRP3 inflammasome's activation. The purinergic pathway involves receptors and molecules that act as extracellular signals to modulate many processes in the body [9,10]. When activated/expressed, the purinergic receptors P2X7 (which binds adenosine triphosphate) and A1 (which binds adenosine) are responsible for developing inflammatory and neuroprotective responses, respectively [11–13]. Interestingly, there are some studies involving psychiatric and neurodegenerative diseases, as well as neurotoxoplasmosis, that show a strong relationship between the P2X7-NLRP3-IL-1β pathway and the illness's pathophysiology, mainly in terms of the inflammatory response [14–18]. Despite the A1 receptor being able to attenuate the inflammatory response and inhibit the activation of immune cells [13,19], there is still a lack of knowledge of the A1 receptor's involvement within the NLRP3 inflammasome and P2X7 receptor context.

Chronic inflammation and inappropriate NLRP3 inflammasome regulation in the brain have been associated with degenerative, psychiatric, and parasitic diseases, such as Alzheimer's disease, Parkinson's disease, schizophrenia, bipolar disorder, major depression, and neurotoxoplasmosis [20–24]. These diseases have complex pathophysiologies and are considered increasingly common public health problems worldwide [25,26]. Additionally, treating these illnesses is very difficult, and many research groups look forward to the development of scientific investigations into new possible therapies for these diseases. MCC950 is an example of a molecule that has been studied to attenuate inflammation; however, this agent can present toxicity by inhibiting NLRP3 completely [27].

On the other hand, natural health products are being studied, considering their medicinal effects, associated with antioxidant and anti-inflammatory actions [28,29]. Resveratrol (RSV) is a natural polyphenol, found in the skins and seeds of grapes, blueberries, blackberries, and peanuts, that has gained scientific attention [30–32]. This polyphenol has

pleiotropic action, mediated by multiple molecular pathways, showing remarkable biological properties and potential therapeutic benefits in several experimental models, as demonstrated in a review study [30]. These multiple properties of RSV are of interest in conditions where oxidative stress, dysregulated inflammation, and excessive cell death play a central role in the pathophysiology, as observed in neurodegenerative diseases and inflammation in microglial cells, for example [30–33].

RSV has demonstrated the ability to act as a potent antioxidant, both through the direct scavenging of ROS and through the modulation and induction of antioxidant enzymes, such as superoxide dismutase and catalase, often via the activation of the nuclear factor erythroid 2-related factor 2 pathway, in 3D neuronal culture in pro-inflammatory conditions [30,34]. Its anti-apoptotic action was observed in cellular protection against programmed death, mediating the expression of Bcl-2 proteins and caspase activity in a study with rats exposed to a pesticide and treated with RSV [35], promoting cellular survival and homeostasis [30,31]. In addition, RSV exerts significant anti-inflammatory actions, inhibiting the activation of pro-inflammatory pathways such as nuclear factor kappa B and the consequent reduction in pro-inflammatory cytokine production and modulating the expression of enzymes such as COX-2 and iNOS [30,33]. Notably, RSV has also been associated with NLRP3 inflammasome and purinergic signaling modulation (the targets of this study), suggesting its therapeutic potential through controlling chronic inflammation and the consequent complications [33,34,36–39].

Despite this, the proof of these modulations using a methodology considering the NLRP3 inflammasome's activation has not been considered. Therefore, we aimed to evaluate the anti-neuroinflammatory potential of RSV through in silico methods and via an innovative in vitro method of NLRP3 inflammasome activation (examining the preventive effect of RSV on the inhibition of inflammasome protein oligomerization) in microglia cells (BV-2 cell line), with a particular focus on the pathway involving the A1 and P2X7 purinergic receptors.

2. Results

2.1. RSV Has Binding Affinity for the PYD Domain of NLRP3

An in silico analysis was performed to evaluate the interaction of RSV with the PYD region of NLRP3 and compare it with the MCC950 inhibitor. As expected, with data obtained by molecular docking, the best target affinity result was for MCC950 (−8.4 kcal/mol). For RSV, the calculated value was −6.7 kcal/mol. The concentration of 12.27 μM of RSV was obtained in our results as capable of inhibiting half of the NLRP3 activity. However, for MCC950, the estimated Ki was 696.25 nM. The root mean square deviation (RMSD) values were 0.480 Å for RSV and 0.541 Å for MCC950, compatible with target binding (Table 1).

Table 1. Molecular docking between the PYD domain of NLRP3 and RSV or MCC950 ligands.

Ligand	Affinity (kcal/mol)	Estimated Ki	RMSD (Å)
RSV	−6.7	12.27 μM	0.480
MCC950	−8.4	696.25 nM	0.541

Figure 1 demonstrates the primary conformation possibility and the bond types between the RSV (Figure 1A,B) or MCC950 (Figure 1C,D) atoms and the PYD target domain of NLRP3. Highlighting RSV, the existing bonds with the NLRP3 inflammasome are hydrogen bonds, Pi–Sigma, stacked Pi–Pi, and Pi–alkyl, with the amino acids leucine (A:39), asparagine (A:60), leucine (A:55), phenylalanine (A:59), and proline (A:40) (Figure 1B).

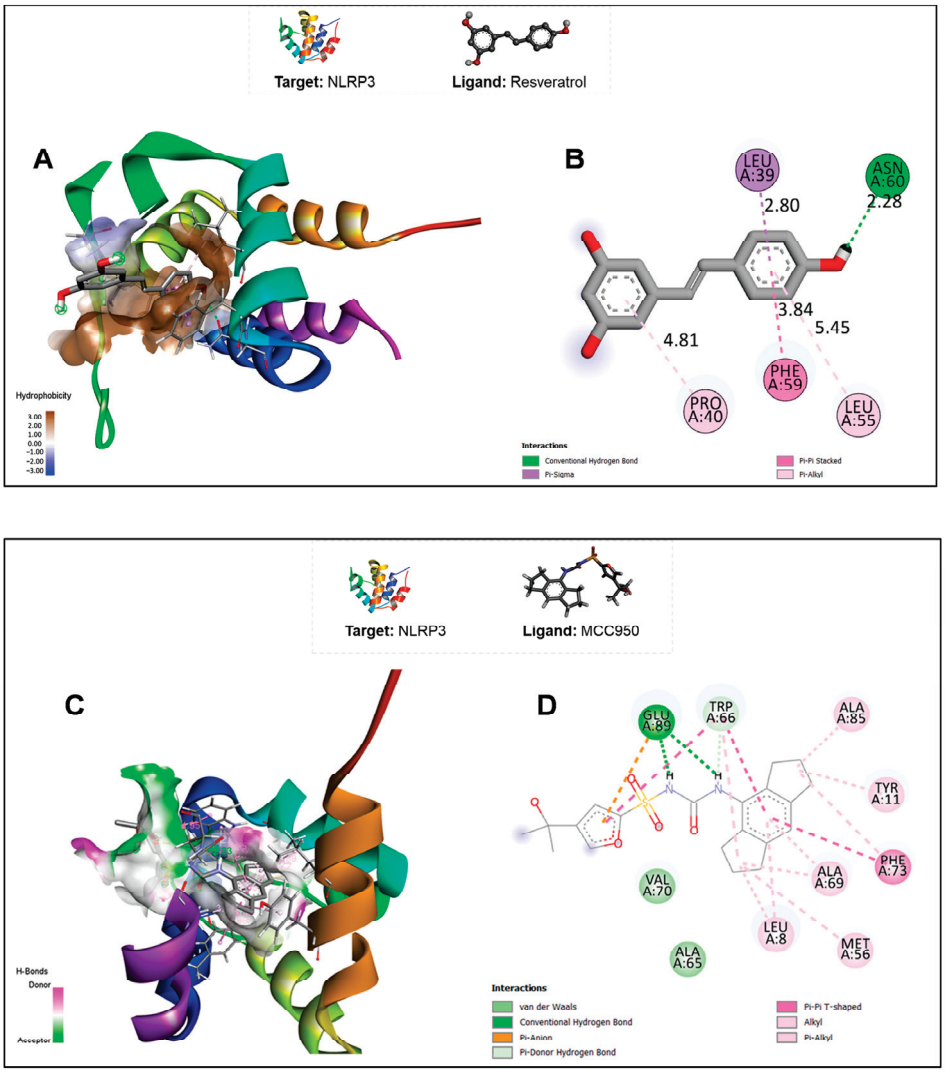


Figure 1. (**A**) Interactions between the PYD domain of the NLRP3 protein and RSV, highlighting its hydrophobicity; (**B**) 2D map of interactions between the PYD domain of the NLRP3 protein and RSV, highlighting the amino acid residues that participate in the bonds and the types of bonds; (**C**) interactions between the PYD domain of the NLRP3 protein and MCC950, highlighting the regions of hydrogen bond formation; (**D**) 2D map of interactions between the PYD domain of the NLRP3 protein and MCC950, highlighting the amino acids that participate in the bonds and the types of bonds.

Molecular dynamics simulations were carried out to evaluate the structural stability of the complexes formed between the PYD domain of NLRP3 (PDB ID: 2NAQ) and the ligands MCC950 and RSV. The APO form was also simulated as a comparative control. An analysis of the RMSD revealed that the APO system maintained average stability of 2.26 Å (±0.33 Å), while the RSV complex exhibited the lowest average RMSD of 2.17 Å (±0.33 Å), indicating the favorable accommodation of the ligand within the binding site. In contrast, the MCC950 complex showed a higher average RMSD of 3.25 Å (±0.60 Å), suggesting greater conformational flexibility throughout the simulation (Figure 2A; Table 2).

Root mean square fluctuation (RMSF) analysis revealed that terminal residues exhibited higher mobility, as expected. However, differences were also observed in internal regions: the MCC950 complex displayed slightly higher fluctuation peaks compared to the APO form, whereas the RSV complex showed a marked reduction in the mobility of the central region of the PYD domain, suggesting a stabilizing effect (Figure 2B).

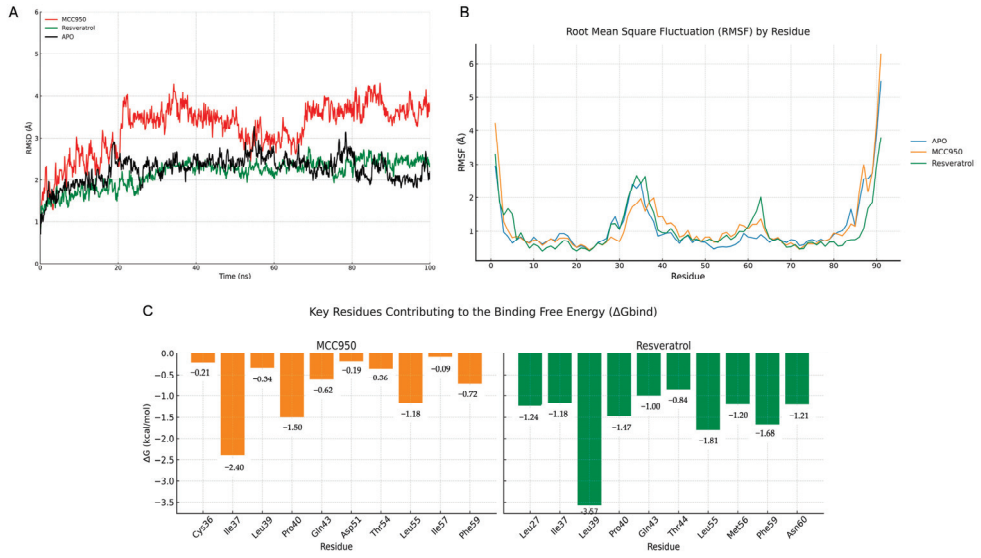


Figure 2. (**A**) Root mean square deviation (RMSD) of the simulations for the APO, MCC950, and RSV systems; (**B**) root mean square fluctuation (RMSF) per residue for the APO, MCC950, and RSV systems; (**C**) per-residue binding free energy contributions of MCC950 and RSV to the PYD domain (PDB ID: 2NAQ).

Table 2. Root mean square deviation (RMSD), radius of gyration (Rg), and solvent-accessible surface area (SASA) statistics for the APO, MCC950, and RSV systems, obtained throughout 100 ns molecular dynamics simulations.

Ligand	RMSD		SASA		Rg	
	Average (Å)	Standard Deviation (Å)	Average (Å^2)	Standard Deviation (Å^2)	Average (Å)	Standard Deviation (Å)
APO	2.26	0.33	5975.17	226.85	22.59	1.87
RSV	2.17	0.33	5751.66	231.06	21.55	1.22
MCC950	3.25	0.60	5802.91	209.53	27.03	2.18

The radius of gyration (Rg) indicated distinct structural compactness among the systems: the MCC950 complex showed the most significant expansion (27.03 Å $\pm$ 2.18), followed by the APO form (22.59 Å $\pm$ 1.87), while the RSV complex exhibited the highest degree of compactness (21.55 Å $\pm$ 1.22), reinforcing its ability to restrict protein flexibility (Table 2). Binding free energy analysis using the MM/GBSA approach confirmed the higher binding affinity for RSV (-22.44 ± 5.91 kcal/mol) compared to MCC950 (-13.63 ± 11.08 kcal/mol), with significant contributions from both electrostatic and van der Waals interactions in the case of RSV (Table 3).

Table 3. Binding free energy (ΔG_bind) components calculated by MM/GBSA for the complexes between the NLRP3 PYD domain and the ligands MCC950 and RSV. Mean values and standard deviations were obtained from 100 snapshots extracted from the last 10 ns of the simulation.

Molecule	ΔE_{vdW}	ΔE_{ele}	ΔG_{GB}	ΔG_{nonpol}	$\Delta G_{MM/GBSA}$
RSV	-31.9652 ± 2.9491	-23.0579 ± 4.0545	37.3798 ± 3.1234	-4.7990 ± 0.2427	-22.4423 ± 5.9119
MCC950	-22.4536 ± 7.2380	-0.6858 ± 5.3272	12.4086 ± 6.4026	-2.9039 ± 1.0302	-13.6347 ± 11.0825

Finally, the energy decomposition per residue highlighted Leu39, Pro40, Leu55, and Phe59 residues as the main contributors to the binding of both ligands. RSV showed a

more favorable interaction with Leu39 (ΔG = −3.57 kcal/mol), while MCC950 interacted more intensely with Ile37 (ΔG = −2.40 kcal/mol) (Figure 2C).

2.2. RSV Decreases Cell Viability Without Increasing the Oxidative Profile

To evaluate its safety, the murine microglia BV-2 cell line was initially exposed to a concentration curve of RSV (0.01–100 μM) for 24, 48, and 72 h (Figure 3). Figure 3 represents a concentration–response curve according to the exposure times. The statistical differences can be observed in Supplementary Material S1, were the same results are demonstrated in a column graph to facilitate their visualization.

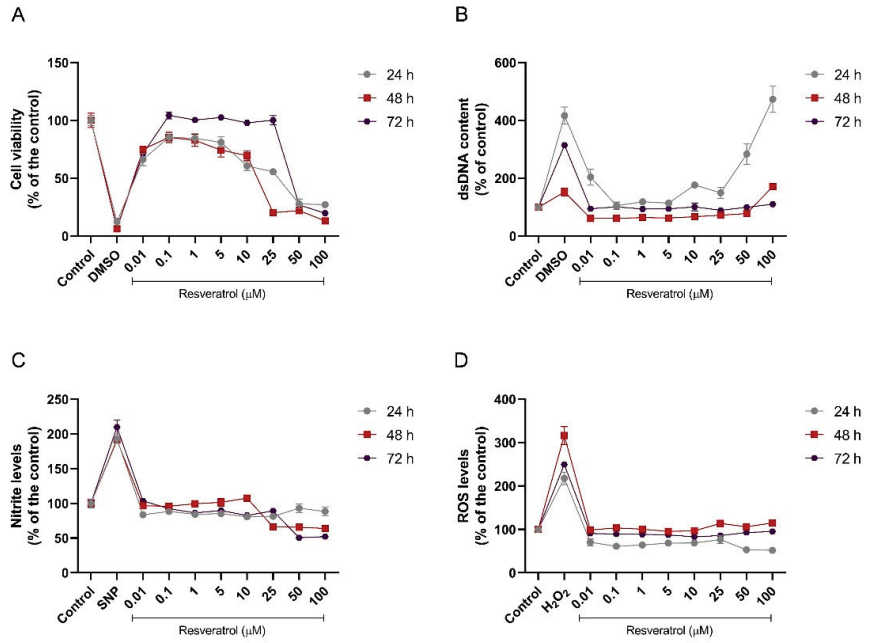


Figure 3. Concentration–response curve of RSV in the microglial cells (BV-2 cell line)—in vitro safety profile. (**A**) Cell viability for 24, 48, and 72 h exposure to RSV; (**B**) dsDNA free content for 24, 48, and 72 h exposure to RSV; (**C**) nitrite levels for 24, 48, and 72 h exposure to RSV; (**D**) ROS levels for 24, 48, and 72 h exposure to RSV. SNP—sodium nitroprusside. The experiments were performed in triplicate. The statistical analysis is demonstrated in Supplementary Material S1 (with the same results in a column graph) and was conducted by one-way ANOVA followed by the Tukey post hoc test. Data are expressed as mean values ± SEM.

During the 24 h incubation period, BV-2 cells showed a decrease in cell viability when exposed to concentrations of 0.01 μM (33.67% reduction, $p < 0.0001$), 10 μM (39.27% reduction, $p < 0.0001$), 25 μM (44.28% reduction, $p < 0.0001$), 50 μM (72.22% reduction, $p < 0.0001$), and 100 μM (72.75% reduction, $p < 0.0001$) when compared to the control (Figure 3A). In addition, membrane damage could be observed when testing the free dsDNA content in the supernatant. Only concentrations of 0.01 μM ($p = 0.0065$), 50 μM ($p < 0.0001$), and 100 μM ($p < 0.0001$) significantly increased this parameter (104.4%, 183.5%, 373.1%, respectively) compared to the control (Figure 3B). Increases in the oxidative profile were not observed; only a decrease in ROS was found under 0.1 μM ($p = 0.0102$), 1 μM ($p = 0.0235$), 50 μM ($p = 0.0009$), and 100 μM ($p = 0.0005$) concentrations (38.93%, 35.98%, 46.86%, and 48.28% reductions, respectively) (Figure 3C,D).

After 48 h of incubation, the decreases in cell viability were maintained under 0.01 μM of RSV (25% reduction compared to the control, $p = 0.0011$). Moreover, 5 and 10 μM of RSV decreased the cell viability to 25.65% ($p = 0.0007$) and 30.32% ($p < 0.0001$) when compared to the control, respectively. The most significant reduction was observed at 25,

50, and 100 μM of RSV (79.74%, 77.89%, and 86.80% reduction compared to the control, $p < 0.0001$). Only 100 μM of RSV increased the dsDNA levels (71.83% compared to the control, $p < 0.0001$) (Figure 3A,B). Unlike what was observed in 24 h, changes in the ROS levels were not observed; however, there was a decrease in the nitrite levels when the cells were exposed to 25, 50, and 100 μM concentrations of RSV (33.82%, 34.28%, and 36.10% reductions, respectively, when compared to the control group, $p < 0.0001$) (Figure 3C,D).

At 72 h of incubation, cell viability was reduced at 0.01 μM of RSV, repeating the other results (28.68% reduction compared to the control, $p < 0.0001$). The two highest concentrations of RSV reduced the cell viability more harmfully, presenting significance of $p < 0.0001$ (50 μM—73.11% and 100 μM—80.15% reduction in cell viability when compared to the control) (Figure 3A). Decreases in oxidative parameters were also observed at this time, at concentrations of 50 and 100 μM for nitrite (49.38%; 47.83%, respectively, $p < 0.0001$) and 10 and 25 μM for ROS (16.86%—$p = 0.0003$; 14.13%—$p = 0.0037$, respectively) (Figure 3C,D). Changes were not observed for the dsDNA release test (Figure 3B).

2.3. Microglia with Activated NLRP3 and Pre-Treated with RSV Arrest Cell Cycle

BV-2 microglial cells were exposed to signals to activate the NLRP3 inflammasome. As expected, cellular oxidative stress was generated for the activation control (LPS + NIG), with an increase in ROS ($p < 0.0001$) and nitrite ($p = 0.0124$) levels. Moreover, the decrease in cell viability was confirmed by the MTT test when compared to the control group ($p = 0.013$) (Figure 4A–C). MCC950, used as an inflammasome inhibitor control, reduced the ROS and nitrite levels and increased the cell viability compared to the activation control ($p < 0.0001$; $p = 0.0001$; $p = 0.0007$, respectively) (Figure 4A–C).

RSV was added to this protocol to evaluate its potential preventive effect in terms of inhibiting inflammasome protein oligomerization. RSV preventive treatment increased the cell viability when compared to the activation control at 0.1, 1, 25, and 100 μM concentrations ($p = 0.0130$; $p < 0.0001$; $p = 0.0183$; $p < 0.0001$, respectively) (Figure 4A). Additionally, all concentrations of RSV tested were able to reduce nitrite significantly, as well as the ROS levels generated in the in vitro NLRP3 inflammasome activation model, when compared to the activation control (LPS + NIG) ($p < 0.0001$, only 5 μM, $p = 0.0002$, for nitrite levels) ($p < 0.0001$ for all RSV concentrations in ROS levels) (Figure 4B,C).

Cell cycle measurement was performed to verify the cell proliferation index (Figure 4D–F). The activation control ($p = 0.0012$), the MCC950 inhibitor ($p = 0.0021$), and all concentrations of RSV ($p = 0.0051$; $p = 0.008$; $p = 0.0004$, respectively) increased the G0/G1 phase compared to the control. MCC950 ($p = 0.0482$) and 10 μM ($p = 0.0114$) and 25 μM ($p = 0.0043$) of RSV also increased this phase when compared to the activation control (LPS + NIG) (Figure 4D). However, only the activation control increased the S phase ($p = 0.0412$), while MCC950 and RSV had the opposite effect, with a decrease compared to the control group (MCC—$p = 0.0439$; 10 μM—$p = 0.0025$; 25 μM—$p = 0.0035$) and activated cell group (MCC—$p = 0.0013$; 10 μM—$p = 0.0029$; 10 μM—$p = 0.0002$; 25 μM—$p = 0.0003$) (Figure 4E). In the G2/M phase, only a decrease was observed (activation control—$p = 0.0019$; MCC—$p = 0.0029$, 1 μM—$p = 0.0053$; 10 μM—$p = 0.0046$; 25 μM—$p = 0.0010$; compared to the control group), with no difference between the activation control (LPS + NIG) and the inhibition control (MCC) and RSV preventive treatments (Figure 4F).

Figure 4G illustrates the microglia exposed to inflammasome activation conditions and RSV preventive treatment at 1, 10, and 25 μM concentrations. It is possible to observe that the group exposed to LPS + NIG presented fewer cells compared to the control. These cells also showed a different morphology (from fusiform to circular). On the other hand,

when cells were treated with MCC950 or 1, 10, or 25 μM of RSV, the number of cells and morphology seemed to be at least partially recovered.

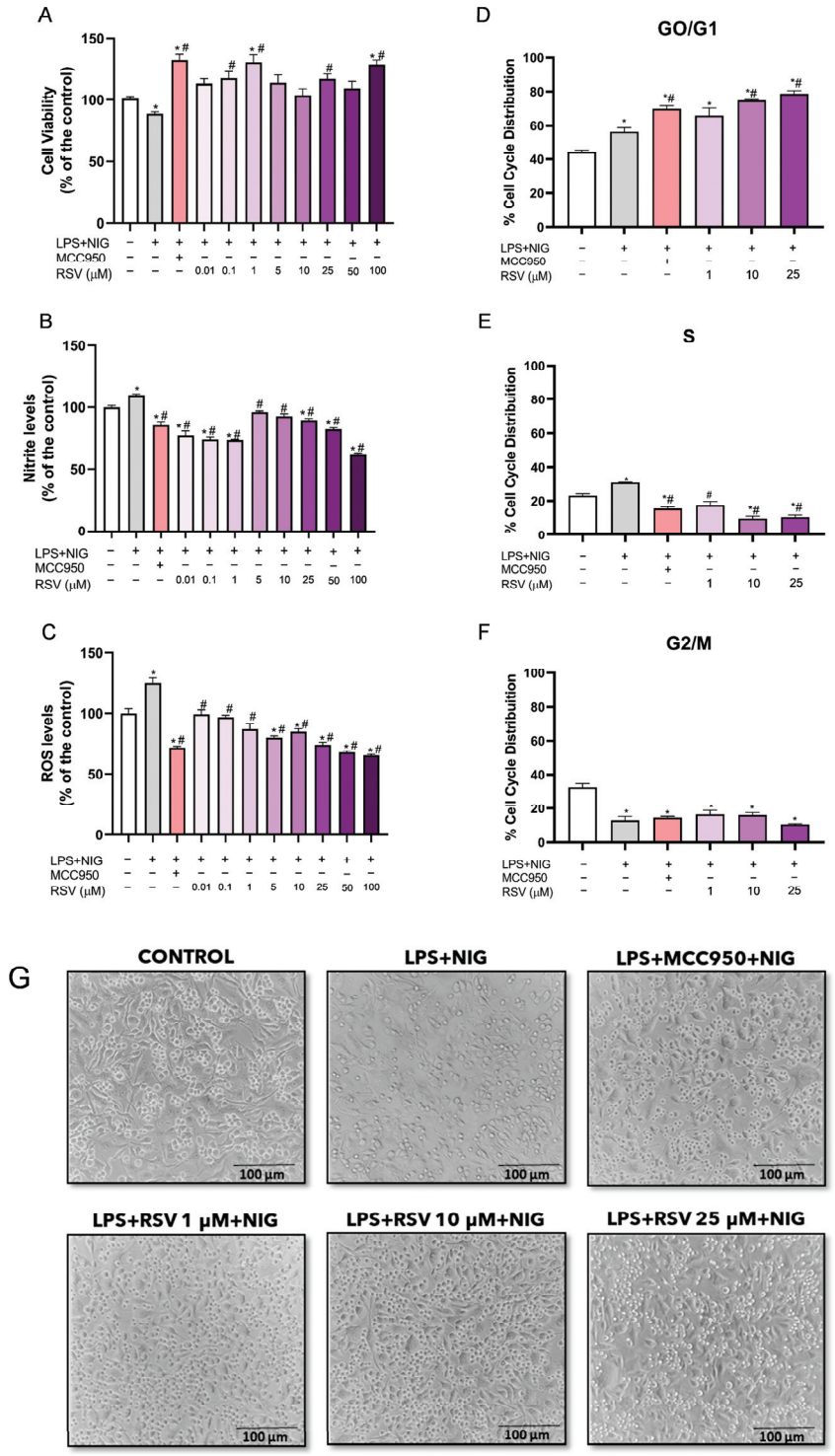


Figure 4. Effects of RSV on cell viability, oxidative profile, and cell cycle of BV-2 cells subjected to the NLRP3 activation protocol. (**A**) Cell viability; (**B**) nitrite levels; (**C**) ROS levels; (**D**–**F**) phases of cell cycle; (**G**) illustrative image of BV-2 cells and their respective treatments. Scale bar: 100 μm. The experiments were performed in triplicate. The statistical analysis was conducted using one-way ANOVA followed by the Tukey post hoc test. $p < 0.05$ was considered significant. Data are expressed as mean values ± SEM. * means comparison to the control group; # means comparisons to the activation group (LPS + NIG).

2.4. P2X7 and A1 Receptors Are Altered by RSV Treatment in Active Microglia

NLRP3 activation control increased P2X7 expression ($p = 0.0104$) and the protein density ($p = 0.003$) (compared to the control group), contributing to exacerbated inflammation and the constant formation of the NLRP3 inflammasome (Figure 5A,C). Meanwhile, the 10 and 25 μM concentrations of RSV ($p = 0.0164$; $p = 0.0048$, respectively) and MCC950 ($p = 0.0330$) restored the P2X7 expression, reducing the values when compared to the activation control (Figure 5A). Only 1 μM of RSV decreased the P2X7 levels compared to the control ($p = 0.0013$) and activation control ($p = 0.0002$) (Figure 5A). The relative protein density of P2X7 was reduced when using MCC ($p = 0.008$) and 10 μM RSV ($p = 0.0009$); the reduction is more evident in RSV compared to the activation control (Figure 5C).

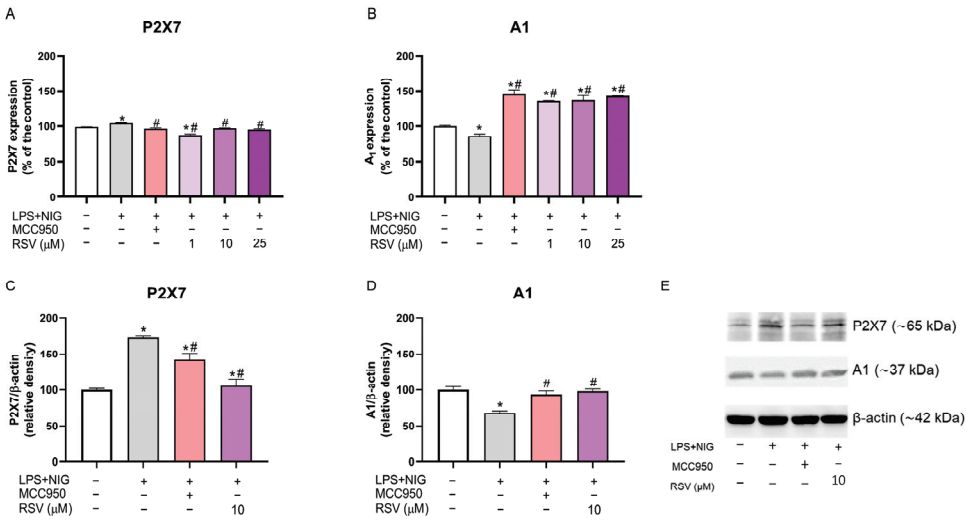


Figure 5. Effects of RSV on the purinergic receptor in BV-2 cells subjected to the NLRP3 activation protocol. (**A**) P2X7 expression by flow cytometer analysis; (**B**) A1 expression by flow cytometer analysis; (**C**) relative protein density of P2X7 receptor by Western blot analysis; (**D**) relative protein density of A1 receptor by Western blot analysis; (**E**) Western blot analysis showing the respective protein bands. The experiments were performed in triplicate. The statistical analysis was conducted using one-way ANOVA followed by the Tukey post hoc test. $p < 0.05$ was considered significant. Data are expressed as mean values ± SEM. * means comparison to the control group; # means comparisons to the activation group (LPS + NIG).

A1 receptor expression and protein levels were reduced in the NLRP3 activation control (LPS + NIG; $p = 0.0100$; $p = 0.0017$, compared to the control group) (Figure 5B,D). The MCC950 treatment showed overexpression of the A1 receptor ($p = 0.0013$, compared to the control and $p = 0.0072$ compared to the activation control—LPS + NIG) (Figure 5B). Additionally, all RSV concentrations increased this receptor's expression when compared to the control (1 μM—$p = 0.0042$; 10 μM—$p = 0.0035$; 25 μM—$p = 0.0017$) and activation control (1 μM—$p = 0.0008$; 10 μM—$p = 0.0007$; 25 μM—$p = 0.0004$) (Figure 5B). Moreover, 10 μM of RSV was sufficient to reestablish the A1 relative protein density, with no differences between the control and activation control ($p = 0.0027$) (Figure 5D). Figure 5E demonstrates the protein bands corresponding to P2X7 and A1, utilized to calculate the relative protein density (entire membranes are given in Supplementary Material S2).

2.5. NLRP3, CASP-1, and Cytokines Are Restored with RSV in Microglia Exposed to Inflammasome Activation Conditions

The mRNA levels and relative protein density of NLRP3, CASP-1, IL-1β, IL-6, TNF-α, and IL-10 in microglial cells exposed to NLRP3 inflammasome activation conditions and

pre-treated with RSV prior to NLRP3 inflammasome oligomerization were determined in this study (Figures 6 and 7).

All mRNA levels of the pro-inflammatory gene increased in cells exposed to LPS + NIG (activation control), compared to the negative control ($p < 0.0001$) (Figure 6A–E). On the other hand, for IL-10, no significant increase was observed (Figure 6F). The MCC950 inhibitor was potent in decreasing all pro-inflammatory mRNAs (Figure 6C–E), NLRP3, and CASP-1 (Figure 6A,B) when compared to the activation control (LPS + NIG) ($p < 0.0001$). An increase in IL-10 was detected compared to the control ($p = 0.0011$) and the activation control ($p = 0.0413$) (Figure 6F).

When added to the cells as a pre-treatment, RSV showed overexpressed pro-inflammatory mRNA levels compared to the control (Figure 6A–E). Nevertheless, RSV was satisfactory in reducing inflammasome components (NLRP3 and CASP-1) when compared to the activation group (LPS + NIG) using 1 μM ($p < 0.0001$), 10 μM ($p < 0.0001$), and 25 μM ($p < 0.0001$; $p = 0.0151$) (Figure 6A,B). In addition, a decrease in the IL-1β (1 μM—$p = 0.0004$; 10 μM—$p = 0.0062$; 25 μM—$p = 0.0009$), IL-6 (1 μM—$p < 0.0001$; 10 μM—$p = 0.0001$; 25 μM—$p = 0.0022$), and TNF-α (1 μM—$p = 0.033$; 10 μM—$p = 0.0048$; 25 μM—$p < 0.0001$) mRNA levels was evident when compared to the activation control (LPS + NIG) (Figure 6C–E).

Although the mRNA levels for IL-10 did not present statistically significant results in the activation control, the MCC inhibitor ($p = 0.0413$) and 1 and 10 μM of RSV ($p < 0.0001$) increased this measure, presenting more significance than MCC950; 1 μM and 10 μM of RSV elevated the IL-10 mRNA levels compared to the activation control (Figure 6F).

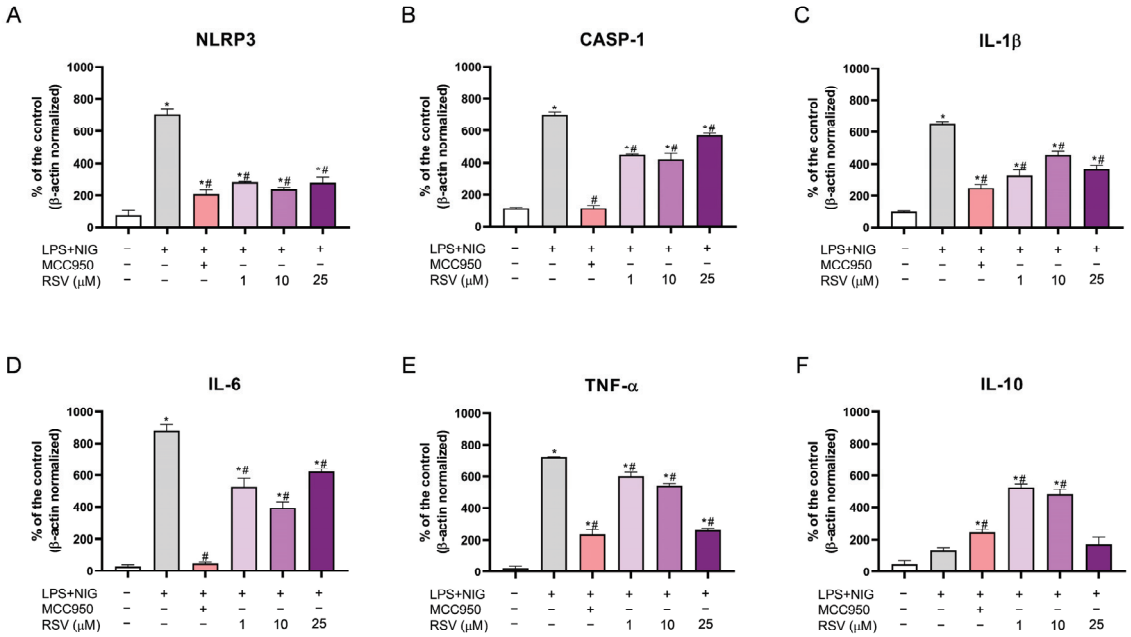


Figure 6. RSV modulation in the inflammatory mRNA levels in BV-2 microglial cells under exposure to NLRP3 inflammasome activation conditions. (**A**) NLRP3 mRNA levels; (**B**) caspase-1 mRNA levels; (**C**) IL-1β mRNA levels; (**D**) IL-6 mRNA levels; (**E**) TNF-α mRNA levels; (**F**) IL-10 mRNA levels. The experiments were performed in triplicate. The statistical analysis was conducted using one-way ANOVA followed by the Tukey post hoc test. $p < 0.05$ was considered significant. Data are expressed as mean values ± SEM. * means comparison to the control group; # means comparisons to the activation group (LPS + NIG).

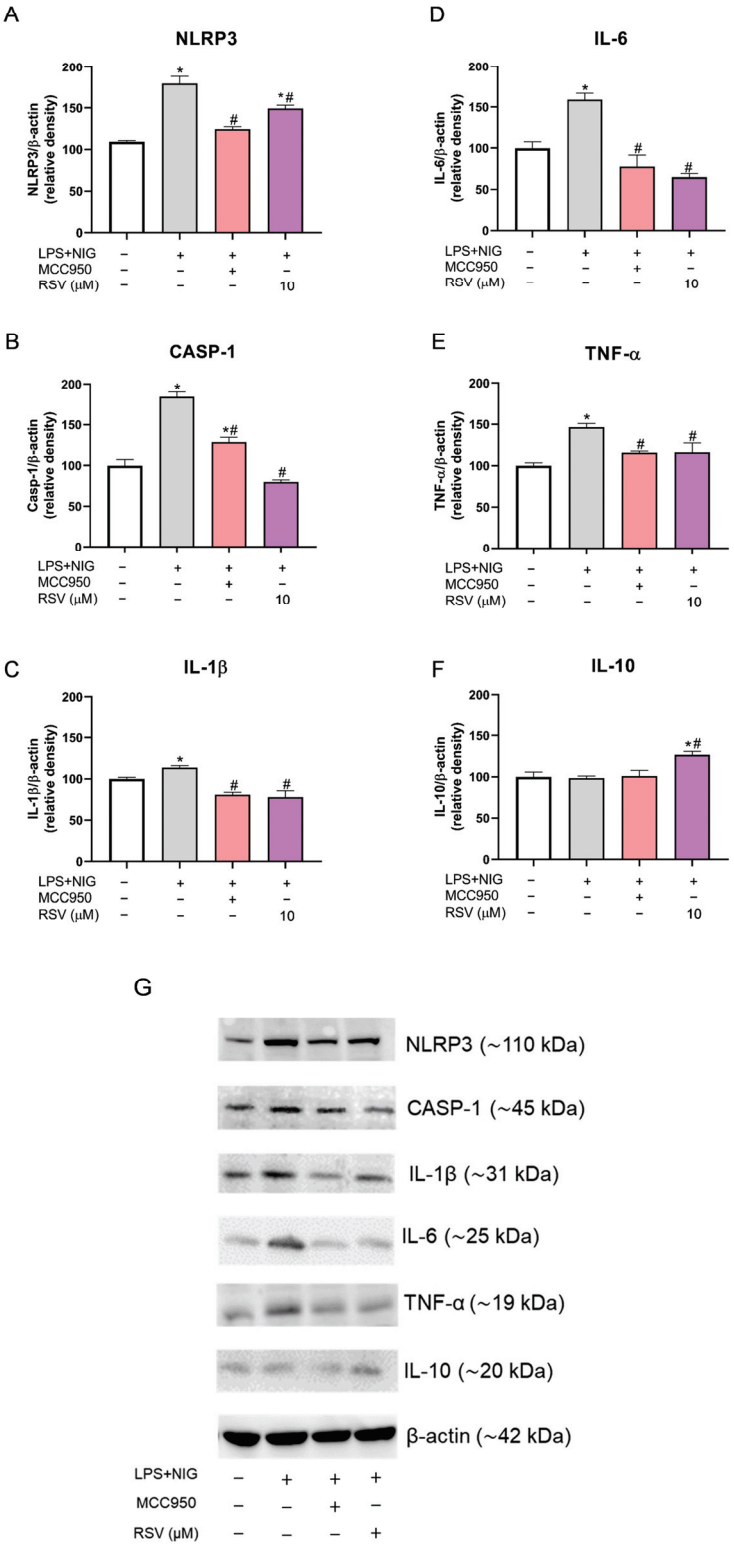


Figure 7. Western blot analysis shows RSV modulation in the relative protein density of inflammatory markers. (**A**) Relative protein density of NLRP3; (**B**) relative protein density of caspase-1; (**C**) relative protein density of IL-1β; (**D**) relative protein density of IL-6; (**E**) relative protein density of TNF-α; (**F**) relative protein density of IL-10; (**G**) Western blot analysis showing the respective protein bands. The experiments were performed in triplicate. The statistical analysis was conducted using one-way ANOVA followed by the Tukey post hoc test. $p < 0.05$ was considered significant. Data are expressed as mean values ± SEM. * means comparison to the control group; # means comparisons to the activation group (LPS + NIG).

The intermediate concentration of RSV (10 μM) was tested for Western blot analysis to perform protein density quantification (Figure 7), and the respective protein bands are shown in Figure 7G (entire membranes are given in Supplementary Material S2). The density of NLRP3 and CASP-1 was increased in microglia exposed to inflammasome activation conditions ($p = 0.0008$; $p < 0.0001$); however, MCC ($p = 0.0025$; $p = 0.0006$) and RSV ($p = 0.0231$; $p < 0.0001$) were effective in reducing these proteins (Figure 7A,B).

Pro-inflammatory cytokines (IL-1β, IL-6, and TNF-α) increased in cells with activated inflammasomes (Figure 7C–E), in agreement with the mRNA level results ($p = 0.048$; $p = 0.009$; $p = 0.0023$). MCC and 10 μM of RSV restored the protein levels of these markers, showing statistical significance only concerning the activation control (LPS + NIG) (Figure 7C–E). For IL-1β, the p values were 0.0201 and 0.0147, respectively. For IL-6, $p = 0.0012$ for MCC and $p = 0.0004$ for 10 μM RSV. The TNF-α p values were 0.0298 and 0.0316, respectively.

Again, although IL-10 did not show any difference between the control and activation control, 10 μM of RSV was effective in increasing the levels of this anti-inflammatory cytokine ($p = 0.0347$ compared to the control and $p = 0.0262$ compared to the activation control) (Figure 7F).

3. Discussion

RSV is a polyphenol found in more than 70 plant species, such as peanuts and grapes [30,32]. Many studies have already demonstrated that RSV has medicinal effects related to the ability to modulate several inflammatory pathways, including the inflammasome [34–37,40]. However, there is no proof obtained through a methodology that uses priming and assembly signals of NLRP3 in microglia. Therefore, this research adopted an innovative method to prove the inhibitory action of RSV on the NLRP3 inflammasome in association with the P2X7 and A1 purinergic receptors for the first time.

Initially, we evaluated, through in silico methods, the existence of ligation between RSV and the PYD region of the NLRP3 inflammasome, as this is the same domain where MCC950, a well-known NLRP3 inhibitory molecule, is capable of interacting to block its oligomerization [41,42]. Thus, it was also possible to compare the inhibitory potential of a synthetic molecule and one obtained from products of natural origin.

The affinity values between the target and ligand indicate biological compatibility between RSV and NLRP3 inflammasome PYD. However, the most promising result was presented by MCC950, indicating higher interaction potential than RSV. However, since MCC950 is a complete NLRP3 inhibitor, the results obtained for RSV could indicate a positive aspect as a partial inhibitor, perhaps bringing the NLRP3 inflammasome to basal levels under inflammatory situations. The RMSD measures the distance between the ligand atoms and the target, with values less than 2 Å indicating the quality of the methodological process [43,44]. From this metric, it can be seen that RSV and MCC950 are compatible with the PYD domain.

The results for Ki reflect the ligand concentration necessary to inhibit half of the target's maximum activity; the lower this value, the greater the inhibitory potential of the molecule. Despite the lower value for MCC950 (696.25 nM), it is highlighted that this result is an estimate calculated from the three-dimensional conformations of the target and ligand molecules and their binding energies. These values cannot correspond to those found in biological experimental models [43,44].

Although the molecular docking results are better for MCC950, it can be proven that RSV has a favorable inhibitory affinity, with a negative value of 6.7 and 12.27 μM necessary to inhibit half of the NLRP3 activity. MCC950, as a synthetic compound, can cause adverse effects when used as a treatment and act as a total inhibitor of NLRP3 [27]. Natural agents can be partial modulators, reducing intense inflammatory activations but without inhibiting what would be

basal and consequently presenting fewer adverse effects [27,45,46]. Thus, this is a promising natural compound that can inhibit the NLRP3 complex by binding in the PYD domain.

Despite the above, our in silico results also demonstrate that RSV has a simple molecular structure and exhibits exceptional affinity for the PYD domain of NLRP3, with binding free energy values ($\Delta G = -22.44 \pm 5.91$ kcal/mol) that are significantly more favorable than those observed for the reference inhibitor MCC950 (-13.63 ± 11.08 kcal/mol). These findings not only corroborate previous reports on the ability of RSV to modulate inflammatory pathways through multiple mechanisms [31,33] but also validate our experimental data, which demonstrated its effectiveness in reducing NLRP3, caspase-1, and IL-1β expression in activated microglia (discussed below).

Molecular dynamics analysis revealed that the RSV–PYD complex exhibited lower structural fluctuations (RMSD = 2.17 ± 0.33 Å) compared to both the APO system (2.26 ± 0.33 Å) and MCC950 (3.25 ± 0.60 Å), indicating the remarkable stabilization of the binding site. This pattern was confirmed by the reduced fluctuation in key residues (RMSF) and greater overall structural compactness (Rg = 21.55 ± 1.22 Å), suggesting that RSV effectively restricts PYD domain mobility. This effect may explain its unique ability to inhibit inflammasome oligomerization in vitro without causing the complete suppression characteristic of irreversible inhibitors such as MCC950.

The per-residue energy decomposition analysis highlighted fundamental differences in binding modes: while RSV showed stronger interaction with Leu39 ($\Delta G = -3.57$ kcal/mol), MCC950 preferentially bound to Ile37 ($\Delta G = -2.40$ kcal/mol). This distinction reflects the fact that different pharmacological profiles of RSV promote more balanced modulation by stabilizing regions critical for PYD oligomerization, whereas MCC950 acts more drastically. These theoretical findings correlate well with our experimental results, in which RSV preserved microglial homeostasis by simultaneously reducing P2X7 expression and the ROS levels while increasing IL-10 and A1 receptor expression, for example.

Our observations corroborate findings that demonstrate the ability of flavonoids derived from *Euterpe oleracea* (apigenin, catechin, and epicatechin) to modulate NLRP3 activation [47,48]. Although RSV was not evaluated in these studies, our data indicate that it shares similar structural and dynamic properties with these phenolic compounds, interacting with the same PYD domain and promoting conformational stabilization compatible with functional inhibition. This convergence of mechanisms reinforces the therapeutic potential of natural NLRP3 modulators, which combine anti-inflammatory efficacy with cellular homeostasis preservation, as evidenced by the absence of cytotoxicity and positive modulation of anti-inflammatory pathways.

Remarkably, RSV surpassed the flavonoids in terms of binding affinity ($\Delta G = -22.44$ kcal/mol vs. -7.1 to -21.14 kcal/mol), suggesting potential advantages in potency and interaction stability. These findings position RSV as an outstanding therapeutic candidate for conditions driven by NLRP3 hyperactivation, offering a unique balance between efficacy and safety. Thus, RSV emerges as a natural NLRP3 modulator with dual potential to inhibit exacerbated inflammation through the conformational stabilization of the PYD domain while simultaneously preserving cellular homeostasis by modulating anti-inflammatory pathways—a distinctive pharmacological profile that clearly differentiates it from synthetic inhibitors such as MCC950.

Therefore, aiming to use RSV as a treatment for neuroinflammatory diseases through the modulation of the NLRP3 inflammasome, we prioritized analyzing this molecule isolated in microglial cells of the murine BV-2 cell line in the second stage. Through basic tests, we verified possible changes in cellular homeostasis that RSV could cause by itself. In this context, it is important to note that we used a specific positive damage control for each test (MTT

and PicoGreen—dimethylsulfoxide (DMSO) reagent; nitrite levels—sodium nitroprusside (SNP) reagent; ROS levels—H_2O_2 reagent), which guaranteed the functionality of the test and that the results presented were reliable. Unusual behavior of the DMSO and H_2O_2 controls for dsDNA, as well as the ROS levels, was observed during the time periods analyzed. It is important to mention that each incubation time had a positive control due to the capacity of the cells (in general) to try to recover their homeostasis, even under stress conditions. This capacity for recovery could explain why we observed a difference between the incubation periods for dsDNA and ROS release in the control groups, which is not conducive to comparing the results of these controls at different times (24 versus 48 versus 72 h). Furthermore, the DMSO and H_2O_2 results (raw data) were normalized to those of the negative control group for each corresponding incubation time. Despite this, using these positive controls, we found a statistically significant increase in dsDNA and ROS release for all incubations, proving that the assay performed reflected the real results. Therefore, we considered the RSV concentration curve (0.01–100 μM) at three exposure times (24, 48, and 72 h).

Few metabolic cell changes are observed at intermediate RSV concentrations (0.1–25 μM) in BV-2 cells. Only the two highest (50 and 100 μM) deserve attention and indicate that caution is needed for their use, since more significant changes that are harmful to cells could be seen, especially when $p < 0.001$, as demonstrated. Furthermore, for these concentrations, a cell viability reduction is evident for all analyzed times. Interestingly, the lowest concentration of RSV (0.01 μM) decreased the cell viability at all times analyzed. Concentration-dependent effects are not necessarily expected when using natural product molecules (as observed here). This effect could be explained by the hormetic pharmacology phenomenon, which is found in many products when applying a curve composed of many concentrations, including RSV [49–51]. In addition, when RSV was in contact with cells for 72 h, cells that were still alive showed an increased proliferation rate compared to previous incubation times (24 and 48 h). This effect could occur due to RSV's beneficial properties, perhaps providing a suitable environment for cellular growth based on its bioactive properties, as well as its antioxidant capacity, for example [32,33,52].

According to the results, there was considerable membrane damage with the exposure of free dsDNA in the supernatant, indicating cell disruption/death and harmful changes to cellular metabolism and homeostasis. The dsDNA release levels are not dependent on the time of exposure to RSV. We observed that the statistical changes were more significant in the first time period—24 h. The nuclease enzymes present in the medium or secreted by the cells degrade DNA, which would explain the significant decrease in dsDNA levels at 48 and 72 h. Another explanation for our results is the antioxidant and pro-survival effects. For example, after an initial cell death response and DNA release (especially at higher concentrations), surviving cells may activate autophagy or repair mechanisms (through RSV's action), reducing the continued release of extracellular DNA.

According to these explanations, it is suggested that the safest concentrations of RSV to be used in vitro and to treat inflammatory diseases are between 0.1 and 25 μM. The results presented here are the concentrations used in scientific references that have applied RSV in models of diseases that affect the CNS and present inflammatory characteristics, ranging from 5 to 25 μM [53–55].

Recent research found that RSV is safe in concentrations lower than 25 μM in RAW 264.7 cells, and, from 50 μM, the cell viability decreases by 64.42% [56]. Additionally, a review study indicates that RSV's cytotoxic effects are directly affected by cellular mitochondria, per our findings, as measured via the succinate dehydrogenase viability in the MTT test [57]. Concentrations above 50 μM are cytotoxic as they impair cellular energy homeostasis, affecting respiration, the mitochondrial structure, and the membrane potential. They activate the cell

death pathway through apoptosis, with changes in calcium homeostasis, the activation of caspases, and the release of cytochrome c [57–59]. However, at low concentrations (<50 μM), RSV can activate cytoprotective mechanisms by increasing antioxidant protection and reducing ROS and NO, including at the mitochondrial level [57,60].

After confirming the in silico interaction of RSV with the NLRP3 PYD domain and studying the behavior of microglia exposed to this molecule, we continued to evaluate it in an innovative model of NLRP3 inflammasome activation in the BV-2 cell line of microglia. Our results, for the first time, reveal the preventive mechanism of action of RSV in this pathway through the P2X7 receptor, avoiding NLRP3 oligomerization.

Cells exposed to the inflammasome activation conditions (LPS + NIG) demonstrated impaired cell viability, along with an increase in nitrite and ROS levels. NLRP3 inflammasome activation decreases the mitochondrial activity of the enzyme succinate dehydrogenase, leading to mitochondrial dysfunction and the release of ROS and reactive nitrogen species, such as NO. This mechanism could be explained by the passage of electrons, which should transit between the mitochondrial complexes for ATP formation but now pass through the pore openings of the mitochondrial membrane, leading to the greater production of ROS and NO [61–63]. Inverse results were observed for MCC950.

Regarding RSV, mainly at the 25 μM concentration, it recovered cell viability; this effect was statistically similar to that in the control group but different from that in the activation control. Surprisingly, all concentrations of RSV reduced the levels of nitrite and ROS compared to the activation control. Still, the 5 and 10 μM concentrations for nitrite and 1 μM for ROS stood out, as some concentrations did not show a statistical difference from the negative control but only from the activation control. RSV reduces ROS production and nitrite levels through multiple interconnected mechanisms [30,52]. It acts as a direct antioxidant by scavenging ROS such as superoxide anion and peroxynitrite and inhibits the activity of NADPH oxidase, a primary source of ROS in inflammatory cells [30,60]. It activates the Nrf2 pathway, promoting the expression of antioxidant enzymes (superoxide dismutase, catalase, and glutathione peroxidase, for example), and modulates nitric oxide synthase activity, inhibiting iNOS under inflammatory conditions, thereby reducing excessive NO production and subsequent nitrite levels [34]. Additionally, it improves mitochondrial function via the activation of SIRT1 and exerts anti-inflammatory effects by inhibiting the NF-κB and MAPK pathways, contributing to a cellular environment with lower oxidative and nitrosative stress [57,64]. Bartra et al. [55] confirmed the protective effects of RSV via the modulation of the SIRT1, Nrf2, and NF-κB pathways in order to reduce inflammation and induce antioxidant enzymes. Furthermore, they propose that SIRT1 mediates the antioxidant and anti-inflammatory mechanisms in LPS-activated microglial cells [55].

Therefore, 1, 10, and 25 μM were chosen for the cytometer and RT-qPCR analyses and 10 μM for protein density measurement (an intermediate concentration of RSV used in the curve, which had an effect in the initial experiments).

The cell cycle results may confirm the microglial activation, with a significant increase in cells in the S phase under LPS + NIG exposure, characterizing the cell cycle arrest in this phase, induced by pro-inflammatory factors [65]. The increase in cells at the S phase suggests that these cells do not complete the cell cycle and achieve the G2/M phases. These results could explain why we observed a decrease in cellular viability/number in the MTT assay for the LPS + NIG group. Microglia, when active, recruit immune system components, including NLRP3 inflammasome activation [65–67]. Meanwhile, MCC950 again showed the opposite results, remaining in the G0/G1 phase of cell rest, blocking replication, and reducing intense cell proliferation and possible uncontrolled inflammation. All RSV concentrations (1, 10, and 25 μM) increased the G0/G1 phase, compared to the

cell and activation controls, indicating the existence of microglia in a resting period or in preparation for division. However, in the S phase, only the activation control increased this parameter, while RSV had the opposite effect compared to inflamed cells. It can be assumed that RSV can slow down or even block the activated cell's replication, remaining at rest for longer. Only a decrease in the G2/M phase was observed, with no difference between the activation or inhibition controls and RSV pre-treatments. It is believed that, due to the short protocol time (6 h), it was impossible to visualize the complete replication cycle of microglia (12–24 h) without changes in this G2/M phase.

Depending on the cell type and pathophysiological context, RSV can affect the cell cycle through different mechanisms. For example, in hepatocellular carcinoma cells, it can induce G2/M phase arrest [68]. In mononucleated cells from rat brains induced in a model of multiple sclerosis, RSV exhibited similar actions to those found in our research, with a decrease in the S phase and an increase in the cell rest phase [69]. Furthermore, in another study, 25 μM of RSV decreased cell proliferation, verified by analyzing the cycle of astrocytes and microglia in an in vitro Alzheimer's disease model, suggesting that inhibition of the NF-κB pathway is responsible for preventing cell cycle progression in inflammatory conditions [70].

The purinergic receptor P2X7 is activated during inflammation, while the A1 receptor activates neuroprotective/anti-inflammatory mechanisms [11–13,71]. The analysis of these receptors in cells with an activated NLRP3 inflammasome is essential to verify the modulation of inflammatory processes and evaluate the inflammatory balance and bioactive properties of RSV in this context, which has not yet been reported.

The P2X7 receptor (activated by extracellular ATP) is essential in releasing inflammatory cytokines, such as IL-1β and IL-18, and activating the NLRP3 inflammasome [11]. The A1 receptor acts as an anti-inflammatory mediator, activated by adenosine (an ATP metabolite) to negatively regulate inflammation [13]. For example, it inhibits the production of pro-inflammatory cytokines, modulates the activity of immune cells, and promotes tissue protection in situations of cellular injury or stress [71]. Therefore, while P2X7 promotes inflammation, A1 acts as an anti-inflammatory counterbalance. The interaction between these receptors reflects the dynamic regulation of the inflammatory process. Dysregulation of ATP and adenosine signaling is implicated in several chronic inflammatory diseases [72]. Examining the interaction between these receptors can reveal potential points of intervention.

Priming and assembly signals increase the P2X7 expression and protein levels, developing pro-inflammatory factors. P2X7 receptor activation can induce and activate the NLRP3 inflammasome [16,73–75]. This process can be interconnected and crucial in regulating the inflammatory and immune response. It is essential for host defense, but, in excess, it is also harmful [11,12,76].

The results reveal that the three RSV concentrations and MCC950 restored the expression values of this receptor compared to the control and activation control. Therefore, it is suggested that RSV modulates the P2X7 receptor, controls neuroinflammation, and reduces tissue damage generated by intense receptor and inflammasome activation.

Consequently, it can prevent the release of pro-inflammatory factors and regulate the activation of the NLRP3 inflammasome, which, in an exacerbated manner, would cause tissue damage [76,77]. At present, our results are the first to demonstrate the joint action of RSV on the P2X7 receptor and the NLRP3 inflammasome.

For the A1 purinergic receptor, a loss of neuroprotective effects was observed through a decrease in its expression and protein levels during NLRP3 activation, but, again, RSV managed to reverse this effect, increasing A1 receptor expression and the protein density. Among the mechanisms through which it exerts neuroprotective effects, the inhibitory

capacity for the release of excitatory neurotransmitters, such as glutamate, stands out, reducing neuronal excitotoxicity and attenuating the CNS inflammatory response by reducing the production of pro-inflammatory cytokines—mainly TNF-α and IL-1β—and inhibiting the activation of immune cells, such as microglia [13,19].

In a study with an animal model for intracerebral hemorrhage, it was seen that the neuroprotective action of RSV was linked to the adenosine A1 receptor. This activation resulted in an improvement in behavioral parameters in rats and also in the morphological tissue characteristics of the brain and a decrease in the expression of TNF-α [78].

Some studies suggest that RSV exerts effects in terms of controlling ATP and ADP degradation through the modulation of enzymes and purinergic receptors [38,79–81]. For this, we hypothesize that the stimulation of microglia with LPS and NIG promotes cell activation and ATP release as a danger signal that binds to the P2X7 receptor. RSV controls ATP degradation and release, explaining how this treatment regulates the receptor (less expression) and reduces the NLRP3 mRNA levels and protein density. However, the A1 receptor's overexpression by RSV could be explained by adenosine-enhancing agents inhibiting degradation and elevating the endogenous adenosine levels to control the protective responses in the brain by potentiating A1 receptor activation [13,80].

The P2X7 receptor and the NLRP3 inflammasome in microglia proliferate (cell cycle), inducing the production and release of pro-inflammatory cytokines, such as IL-6, IL-1β, and TNF-α, to eliminate the stressor [65]. Upon analyzing the inflammatory mRNA and protein levels, we can also confirm the direct action of RSV in modulating these cytokines.

Exposure to LPS + NIG increased the cytokine levels, which was reversed with the use of RSV, attenuating the exacerbated inflammation. On the other hand, IL-10 is an anti-inflammatory cytokine responsible for restoring cellular homeostasis in pro-inflammatory conditions [82]. It was evident that 1 and 10 μM of RSV, when compared to the action of MCC950, acted as a neuroprotector, with better results in increasing IL-10, inhibiting microglial activation and inflammation.

Furthermore, the NLRP3 and CASP-1 mRNA and protein levels (inflammasome components) were decreased with RSV pre-treatment. It is possible to note that ROS levels are also reduced with RSV, which contributed to helping to restore the expression of these inflammasome components, as, when increased, they function as inducers for the formation and activation of NLRP3 [3,62]. Recent studies highlight the P2X7/NLRP3/IL-1β pathway in neurological diseases [18,83–87]. In a mouse epilepsy protocol, the gene expression of P2X7, NLRP3, and IL-1β was increased [83]. Moreover, in an animal model with a depression-like phenotype, the P2X7/NLRP3/IL-1β pathway was noted as a potential mechanism for the study of treatments that target this pathological condition [87].

RSV exhibits pleiotropic properties, and certain effects may only occur at specific concentrations due to the modulation of multiple molecular pathways in a concentration-dependent manner [30,50]. Evidence from the literature indicates that lower concentrations of RSV (<50 μM) tend to activate cytoprotective and antioxidant pathways, such as SIRT1. In contrast, higher concentrations (>50 μM) induce alterations in calcium homeostasis, mitochondrial membrane potential disruption, and caspase activation, ultimately leading to apoptosis—particularly in cancer cells [57,58]. In this context, Calabrese et al. [50] also emphasize that high RSV concentrations are required to trigger pro-apoptotic mechanisms in tumor cells, while concentrations below 50 μM are more associated with antioxidant and anti-inflammatory effects [50]. Therefore, the cellular responses to RSV appear to be concentration-dependent, which may explain the variability observed across different studies and underscores the importance of considering the concentration as a key factor when interpreting experimental outcomes.

In summary, this is the first study that shows the anti-inflammatory potential of RSV through the modulation of the NLRP3 inflammasome in association with purinergic receptors, such as P2X7 and A1. This characterizes this scientific investigation as an innovative approach, bringing new perspectives that will allow for future investigations.

4. Materials and Methods

4.1. Computer Simulation

4.1.1. Molecular Docking

Molecular docking simulations were performed to investigate the interaction between RSV and the NLRP3 protein PYD region (pyrin domain). The AutoDock Vina software version 1.1.2 (AD Vina) integrated with the AMDOCK tool was used [43,88]. The target was the PYD domain of NLRP3, the same area where MCC950 inhibits the inflammasome [27]. The chemical structure of the PYD region was obtained from the Protein Data Bank (PDB ID: 2NAQ), and the ligand molecules, RSV (ID: 445144) and MCC950 (ID: 9910393), were obtained from the PubChem database. The complexes formed were evaluated and analyzed regarding their binding energy/affinity, inhibition constant (estimated Ki), and RMSD (distance between atoms). The Functional Transformation SITE (FTSITE) tool was used to identify and visualize binding sites between ligand targets [88]. Similar computational methodologies have been employed in previous studies by our research group [47,89,90]. The molecular docking process was carried out using the AMDOCK software 1.6.2 version, and Pymol 3.1 version (Schrödinger, LLC, New York, NY, USA) was used for visualization and image generation.

4.1.2. Molecular Dynamics Simulation

The three-dimensional structures of the ligands were optimized at the HF/6-31G* level using the Gaussian09 software [91], and partial atomic charges were calculated using the RESP method [92]. System preparation was performed using the tLEaP module from the AMBER package [93], applying the GAFF force field for the ligands and ff14SB for the protein [94]. The PYD domain (PDB ID: 2NAQ) was protonated at physiological pH (7.4) using the H++ server [95] and solvated in a TIP3P water box [96,97] with periodic boundaries. Ions were added to neutralize the system and simulate an ionic strength of 0.15 M, reproducing physiological conditions.

The systems were subjected to a multi-step minimization protocol involving solvent relaxation, the relaxation of protein hydrogens, the joint relaxation of protein and solvent hydrogens, and global system minimization. The hybrid steepest descent/conjugate gradient algorithm was used at all stages to ensure efficient energy convergence. This protocol was developed based on methodologies previously established in our studies [48, 89,98]. The simulations were conducted for 100 ns in the isothermal–isobaric ensemble (NPT), maintaining a constant temperature and pressure using a Langevin thermostat [99] and Berendsen barostat [100]. The trajectories generated were used for structural analyses and binding free energy calculations.

4.1.3. Generalized Born and Surface Area Continuum Solvation Calculations (MM/GBSA)

The affinity between the ligands and NLRP3 was quantified using the MM/GBSA method [101], available in the AmberTools23 package [93]. The calculations were based on 500 snapshots extracted from the last 10 ns of the molecular dynamics simulations. The energy components considered included van der Waals interactions, electrostatic contributions, polar solvation (generalized Born model), and non-polar solvation (solvent-accessible surface area). Energy decomposition per residue was carried out to identify the most relevant residues contributing to the stabilization of the complexes. The mathematical

foundation underlying the binding free energy calculations has already been previously published [102].

4.2. Microglia Cell Culture

A microglia cell line (BV-2) (Rio de Janeiro Cell Bank Code—0356, Rio de Janeiro, RJ, Brazil) was cultured with RPMI-1640 medium (Vitrocell®, Campinas, SP, Brazil) modified to contain 10 mM HEPES (Ludwig®, Alvorada, RS, Brazil) with 10% fetal bovine serum (FBS) (Vitrocell®, Campinas, SP, Brazil) and supplemented with 1% of antibiotics (100 U/mL penicillin; 100 mg/mL streptomycin) (Vitrocell®, Campinas, SP, Brazil). Cells were cultured at 37 °C with 5% CO_2 and atmospheric oxygen.

4.3. Resveratrol In Vitro Safety Profile

RSV was tested in BV-2 cells for its safety profile. RSV (C14H12O3; molecular weight 228, 25 g/mol; Sigma-Aldrich®, St. Louis, MO, USA) was initially diluted in DMSO (Sigma-Aldrich®, St. Louis, MO, USA) and RPMI-1640 medium (1% DMSO in initial solution), followed by dilutions into the culture medium, and added to the cells at concentrations of 0.01, 0.1, 1, 5, 10, 25, 50, and 100 µM [103]. In RSV 100 µM, the existing percentage of DMSO was 0.1% (non-toxic to BV-2 cells). The incubation times were 24, 48, and 72 h [47,104,105]. After incubation, the cell viability, extracellular double-strand DNA quantity (dsDNA), nitrite (indirect nitric oxide—NO measurement), and ROS levels were evaluated (Figure 8A). In this step, controls to induce cell damage were used: 15% DMSO for cell viability and dsDNA, 10 µM SNP (Sigma-Aldrich®, St. Louis, MO, USA) for nitrite levels, and 25 µM hydrogen peroxide (H_2O_2, Sigma-Aldrich®, St. Louis, MO, USA) for ROS levels [47,104–107]. Moreover, the negative damage control (control) group was composed of cells without treatment. It is important to mention that this experiment was conducted in triplicate and with the corresponding assays.

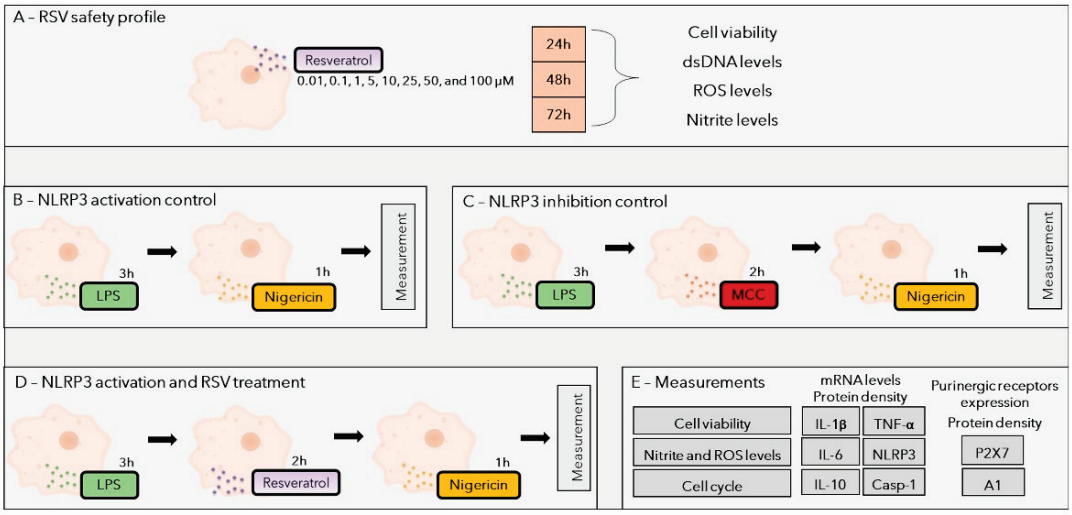


Figure 8. An illustrated scheme of the in vitro experiments performed in this study. (**A**) Investigation of RSV cytotoxicity in BV-2 cells. Microglial cells were exposed to a concentration curve of RSV for 24, 48, and 72 h, followed by cell viability, dsDNA, ROS, and nitrite level measurements; (**B**) scheme of exposure to treatments that induced NLRP3 activation; (**C**) scheme of exposure scheme to treatments that inhibited NLRP3 activation for the inhibition control; (**D**) scheme of exposure to RSV treatment to test its action in the NLRP3 inflammasome; (**E**) list of measurements performed in (**B**–**D**) treatments.

4.4. Resveratrol Pre-Treatment and NLRP3 Inflammasome Activation

Initially, 1 μg/mL of LPS (Sigma-Aldrich®, St. Louis, MO, USA) was added to the microglia cells for 3 h [20,108]. After priming–inducing factor addition, the same cells were treated with RSV (0.01–100 μM) for 2 h [20,103,109], followed by 10 μM NIG (Sigma-Aldrich®, St. Louis, MO, USA) exposure for 1 h [47,109]. RSV was used to test its anti-inflammatory capacity through the prevention of NLRP3 inflammasome oligomerization, potentially acting as an inhibitor between the use of LPS (increases the inflammasome transcription factors) and NIG (induces the assembly (oligomerization) of the inflammasome proteins). As an inhibition control, 100 nM of MCC950 (Sigma-Aldrich®, St. Louis, MO, USA) was added, between the use of LPS and NIG, for 2 h [20,109]. MCC950 is already known to inhibit the oligomerization of the NLRP3 inflammasome's ASC domain. We also used an activation control; for this, LPS and NIG were added to the cells at the exact times described here. This innovative in vitro protocol for the activation of the NLRP3 inflammasome followed guidelines from Zhou et al. [20], Davidson et al. [47], and El Sabbagh et al. [109] with modifications. The experimental model is schematized in Figure 8B–E. After all incubations, the cell viability, parameters related to oxidative stress, cell cycle arrest, and the expression of purinergic receptors, cytokines, CASP-1, and NLRP3 were evaluated. This experiment was conducted in triplicate, as well as the corresponding assays.

4.5. Experimental Assays

4.5.1. Determination of Cell Viability and Extracellular Double-Strand DNA

Cell viability was evaluated using the MTT assay. After the removal of the supernatant, the cells were incubated with 20 μL of MTT reagent (3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide) (Sigma-Aldrich®, St. Louis, MO, USA) at 5 mg/mL and 200 μL of PBS 1X for 1 h at 37 °C. After incubation, the reagent was removed and 200 μL of DMSO was added to solubilize the formazan crystals generated by the mitochondrial succinate dehydrogenase enzyme. Absorbance was determined at 560 nm in the plate reader SpectraMax I3 (Molecular Devices, San Jose, CA, USA) [110].

dsDNA was quantified in the cellular supernatant using 96-well black plates. First, 10 μL of PicoGreen® dsDNA reagent (Thermo Fisher Scientific®, Waltham, MA, USA), diluted 1:200 in Tris–EDTA 1X buffer, was added to 10 μL of the samples and 80 μL of Tris-EDTA 1X buffer for 5 min at room temperature, avoiding light exposure. The fluorescence intensity was determined at 480 nm of excitation and 520 nm of emission (SpectraMax I3, Molecular Devices, San Jose, CA, USA) [111].

4.5.2. Nitrite and ROS Level Measurement Analysis

Using an indirect protocol to detect NO, organic nitrite was measured. First, 50 μL of Griess reagent (sulfanilamide 1% and N-1-naphthylenediamine-bihydrochloride 0.1%, proportion—1:1) was added to 50 μL of the supernatant sample for 15 min at room temperature and under light protection. The absorbance was determined at 540 nm (SpectraMax I3, Molecular Devices, San Jose, CA, USA) [112].

To measure ROS production, 0.1 mM of 2′-7′ dichlorodihydrofluorescein diacetate (DCFH-DA) reagent (10 μL) (Sigma-Aldrich®, St. Louis, MO, USA) was added to 50 μL of the sample and 100 mM of Tris–HCl buffer (65 μL). This assay was carried out using black 96-well plates. A fluorescence reading was performed at excitation of 485 nm and emission of 520 nm after 1 h of incubation at room temperature with light protection (SpectraMax I3, Molecular Devices, San Jose, CA, USA) [113].

4.5.3. Cell Cycle Evaluation

With the three most effective concentrations of RSV determined through the previous assays, the treatments were performed again to carry out the cell cycle evaluation. After the treatment protocols, propidium iodide (PI) (Sigma-Aldrich®, St. Louis, MO, USA) was used in this analysis. Cells were collected and washed with PBS 1X to remove the medium and other debris. After washing, microglia were centrifuged at 1500 rpm for 5 min (low-speed differential centrifugation). The pellet cells were fixed with 70% ice-cold ethanol and centrifuged again for 5 min at 1500 rpm. The cell pellets were resuspended in a solution containing 0.05% Triton, 100 μg/mL RNase, and 50 μg/mL PI, following 30 min of incubation at 37 °C. The BV-2 microglial cells were analyzed using a flow cytometer (BD FACSVerse; BD Biosciences, San Jose, CA, USA), and the data were examined using the FlowJo V10 software (FlowJo, Ashland, OR, USA) [114].

4.5.4. Flow Cytometer Analysis

The three most effective concentrations of RSV were determined previously via the initial assays to perform the analysis of A1 and P2X7 receptor expression using a flow cytometer (BD FACSVerse; BD Biosciences, San Jose, CA, USA), and the quantification was performed using the FlowJo V10 software (FlowJo, Ashland, OR, USA). BV-2 microglial cells were fixed in 4% paraformaldehyde (PFA) and washed with PBS containing 2% FBS. Primary antibodies for A1 and P2X7 purinergic receptors (1:500, Life Technologies, Carlsbad, CA, USA) were used in the samples for 35 min. After washing with PBS, Alexa Fluor 555 or 647 secondary antibodies (1:500, Life Technologies, Carlsbad, CA, USA) were added for the same incubation time. In the flow cytometer, forty thousand events were acquired for the samples. Forward and side light-scatter signals were used to exclude dead cells and debris [115].

4.5.5. RT-qPCR Analysis

RNA isolation followed the manufacturer's standards, using Trizol reagent (TRI Reagent—Sigma-Aldrich®, St. Louis, MO, USA). Samples were quantified (RNA ng/μL) using a NanoDrop Lite (Thermo Fisher Scientific®, Waltham, MA, USA) and cDNA conversion using a commercial iScritTM cDNA synthesis kit (Bio-Rad, CA, USA). IL-1β, IL-6, IL-10, TNF-α, NLRP3, and CASP-1 mRNA quantification was conducted using specific primers (Supplementary Material S3) (Ludwig®, Alvorada, RS, Brazil) and the components provided in the commercial GoTaq® qPCR master kit (Promega, Fitchburg, WI, USA) in a QuantStudio 5 real-time PCR device (ThermoFischer Scientific, Waltham, MA, USA). The cycle conditions were as follows: 50 °C—2 min, 95 °C—2 min and 40 cycles of 95 °C—15 s and 60 °C—30 s. The mRNA level data were analyzed using delta–delta Ct calculation and relative expression conversion. It is important to mention that these analyses were carried out using the three main concentrations of RSV that presented effective potential in the initial experiments.

4.5.6. Western Blot Analysis

Finally, one concentration of RSV was selected to assess the protein expression of different important proteins involved in the inflammatory cascade and purinergic system. This selection was based on the intermediate concentration of RSV used in the curve. Microglia were homogenized in ice-cold radioimmunoprecipitation assay buffer with 1 mM protease and phosphatase inhibitors. The protein concentrations were determined using the Lowry methodology (Sigma-Aldrich). Samples were separated using sodium dodecyl sulfate–polyacrylamide gel electrophoresis and transferred into nitrocellulose membranes

(Amersham Biosciences, Slough, UK). After blocking, the membrane samples were incubated overnight at 4 °C with primary antibodies: NLRP3 (dilution 1:500, Santa Cruz Biotechnology, Santa Cruz, CA, USA), CASP-1 (dilution 1:500, Santa Cruz Biotechnology, CA, USA), IL-1β (dilution 1:500, Santa Cruz Biotechnology, CA, USA), P2X7 receptor (dilution 1:500, Santa Cruz Biotechnology, CA, USA), A1 receptor (dilution 1:500, Santa Cruz Biotechnology, CA, USA), TNF-α (dilution 1:500, Santa Cruz Biotechnology, CA, USA), IL-6 (dilution 1:500, Santa Cruz Biotechnology, CA, USA), and IL-10 (dilution 1:500, Santa Cruz Biotechnology, CA, USA). Afterward, anti-rabbit or anti-mouse secondary antibodies were incubated in these membranes (dilution 1:10.000, Santa Cruz Biotechnology) for 90 min at room temperature. The membranes were incubated with an enhanced chemifluorescent substrate (Amersham Biosciences) and analyzed with an Amersham Imager 600. The membranes were reprobed and tested for β-actin immunoreactivity as a control for the protein concentration. The band intensity was quantified via ImageJ 1.6.0 version to determine the relative protein density [116].

4.6. Statistical Analysis

The results were plotted in Microsoft Excel (Microsoft Office Professional Plus 2019 Version) and then converted to percentages relative to the negative control. One-way ANOVA was used to compare groups with a significance level of $p < 0.05$. Tukey's post hoc test was applied to correct multiple comparisons with $\alpha = 0.05$. Data were statistically analyzed in the GraphPad Prism 8.0 software (GraphPad Prism®, San Diego, CA, USA) and presented as the mean ± standard error of the mean (SEM). Comparations with $p < 0.05$ were considered statistically significant.

5. Conclusions

The present study confirmed the anti-neuroinflammatory action of RSV in a BV-2 cell line of microglia exposed to NLRP3 inflammasome activation conditions, highlighting the role of P2X7 and A1 in this context for the first time. We demonstrated the inhibitory ligation of RSV with the NLRP3 PYD domain by in silico methods. RSV preventive treatment can reduce pro-inflammatory NLRP3 and CASP-1 levels by decreasing P2X7 cellular receptors while increasing A1. Finally, RSV is a possible candidate for further investigations as a potential treatment against neuroinflammatory diseases, especially for those characterized by the exacerbated activation of the P2X7 purinergic receptor and NLRP3 inflammasome in their pathophysiology.

6. Limitations

This present study used in silico and in vitro methodologies and, for the first time, presents the integration of purinergic receptors (P2X7 and A1) and the NLRP3 inflammasome, using RSV as a possible treatment to control neuroinflammation. However, this study also has limitations. Here, we used a cell line obtained commercially. In this regard, using primary cultured microglia, cerebral organoids (3D culture), and in vivo evaluations could be promising alternatives to perform additional experiments to obtain similar and complementary results. Indeed, by considering more complex models mimicking a living organism, we could confirm and complement our findings.

Furthermore, this study did not evaluate the in vitro binding capacity of RSV with the PYD domain of NLRP3. However, the results suggest that RSV can act as an anti-inflammatory agent by regulating the NLRP3 mRNA levels and protein density. We also emphasize that our experimental analyses focused on the cytokines NLRP3 and CASP-1, which confirmed the action of RSV through priming signaling and the regulation of NLRP3 at the gene expression and protein density levels.

It is important to mention that this study did not directly analyze its biological potential in the NLRP3 assembly mechanism. On the other hand, our findings provide a potential reference for other research groups interested in analyzing this specific mechanism. We suggest that P2X7 acts as an induction mechanism for NLRP3 assembly, considering the elevated levels of the CASP-1 protein detected by the Western blot analysis (pro-caspase-1 cleaved to CASP-1 via inflammasome activation). In summary, RSV decreases the P2X7 purinergic receptor's expression, and we believe that it reduces inflammasome assembly and releases pro-inflammatory markers, including CASP-1 and IL-1β. We strongly believe that our results are relevant and contribute to the scientific literature, bringing new perspectives regarding the mechanistic understanding of RSV's anti-inflammatory effects through an innovative method of NLRP3 inflammasome activation, highlighting the P2X7 receptor's role in this context.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ph18070950/s1, Supplementary Material S1: RSV cytotoxicity measurements in microglial cells (BV-2 cell line)—in vitro safety profile; Supplementary Material S2: Complete membranes corresponding to the Western blot presented in Figures 7 and 8 of the article; Supplementary Material S3: Primers forward (F) and reverse (R) for gene expression.

Author Contributions: Conceptualization, B.F.B., A.K.M. and A.S.d.S.; methodology, B.F.B., A.F.d.S., and J.A.P.d.R.; formal analysis and investigation, B.F.B., C.E.A., M.V.d.S. and P.M.C.; resources and funding acquisition, S.B.F., M.R.C.S., V.M.M.M., N.B.B., A.K.M. and A.S.d.S.; supervision and project administration, N.B.B., A.K.M. and A.S.d.S.; data curation, B.F.B., writing—original draft preparation, B.F.B.; writing—review and editing, C.E.A., M.V.d.S., P.M.C., A.F.d.S., S.B.F., J.A.P.d.R., M.R.C.S., N.B.B., A.K.M. and A.S.d.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—CAPES/PROEX process n° 88887.677231/2022-00 and the FAPERGS-Gaúcho Researcher Program process n° 82630.812.29195.25032024.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in this article. Further inquiries can be directed to the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

ASC	Apoptosis-associated adapter protein containing the CARD domain
CASP-1	Caspase-1
CNS	Central nervous system
DAMPs	Damage-associated patterns
DCFH-DA	2′-7′-Dichlorodihydrofluorescein diacetate
DMSO	Dimethylsulfoxide
dsDNA	Extracellular double-stranded DNA
FSB	Fetal bovine serum
FTSITE	Functional transformation SITE
IL-10	Interleukin-10
IL-18	Interleukin-18
IL-1β	Interleukin-1 beta

IL-6	Interleukin-6
LPS	Lipopolysaccharide
MM/GBSA	Generalized Born and surface area continuum solvation calculations
NIG	Nigericin
NLRP3	NOD-like receptor family pyrin domain containing 3
NO	Nitric oxide
NPT	Isothermal–isobaric ensemble
PAMPs	Pathogen-associated molecular patterns
PFA	Paraformaldehyde
PI	Propidium iodide
Rg	Radius of gyration
RMSD	Root mean square deviation
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
RMSF	Root mean square fluctuation
RSV	Resveratrol
SASA	Solvent-accessible surface area
SNP	Sodium nitroprusside
TNF-α	Tumor necrosis factor-alpha

References

1. Rabolli, V.; Lison, D.; Huaux, F. The Complex Cascade of Cellular Events Governing Inflammasome Activation and IL-1β Processing in Response to Inhaled Particles. *Part. Fibre Toxicol.* **2016**, *13*, 1–17. [CrossRef] [PubMed]
2. Groslambert, M.; Py, B.F. Spotlight on the NLRP3 Inflammasome Pathway. *J. Inflamm. Res.* **2018**, *11*, 359–374. [CrossRef] [PubMed]
3. Song, N.; Li, T. Regulation of NLRP3 Inflammasome by Phosphorylation. *Front. Immunol.* **2018**, *9*, 2305. [CrossRef]
4. Kelley, N.; Jeltema, D.; Duan, Y.; He, Y. The NLRP3 inflammasome: An overview of mechanisms of activation and regulation. *Int. J. Mol. Sci.* **2019**, *20*, 3328. [CrossRef]
5. Swanson, K.V.; Deng, M.; Ting, J.P.-Y. The NLRP3 Inflammasome: Molecular Activation and Regulation to Therapeutics. *Nat. Rev. Immunol.* **2019**, *19*, 477–489. [CrossRef]
6. Schroder, K.; Sagulenko, V.; Zamoshnikova, A.; Richards, A.A.; Cridland, J.A.; Irvine, K.M.; Stacey, K.J.; Sweet, M.J. Acute Lipopolysaccharide Priming Boosts Inflammasome Activation Independently of Inflammasome Sensor Induction. *Immunobiology* **2012**, *217*, 1325–1329. [CrossRef]
7. Elliott, E.I.; Sutterwala, F.S. Initiation and Perpetuation of NLRP3 Inflammasome Activation and Assembly. *Immunol. Rev.* **2015**, *265*, 35–52. [CrossRef] [PubMed]
8. He, Y.; Hara, H.; Núñez, G. Mechanism and Regulation of NLRP3 Inflammasome Activation. *Trends Biochem. Sci.* **2016**, *41*, 1012–1021. [CrossRef]
9. Abbracchio, M.P.; Burnstock, G.; Verkhratsky, A.; Zimmermann, H. Purinergic Signalling in the Nervous System: An Overview. *Trends Neurosci.* **2009**, *32*, 19–29. [CrossRef]
10. Di Virgilio, F. Purinergic Mechanism in the Immune System: A Signal of Danger for Dendritic Cells. *Purinergic Signal.* **2005**, *1*, 205–209. [CrossRef]
11. Lister, M.F.; Sharkey, J.; Sawatzky, D.A.; Hodgkiss, J.P.; Davidson, D.J.; Rossi, A.G.; Finlayson, K. The Role of the Purinergic P2X7 Receptor in Inflammation. *J. Inflamm.* **2007**, *4*, 5. [CrossRef] [PubMed]
12. Oliveira-Giacomelli, Á.; Petiz, L.L.; Andrejew, R.; Turrini, N.; Silva, J.B.; Sack, U.; Ulrich, H. Role of P2X7 Receptors in Immune Responses During Neurodegeneration. *Front. Cell. Neurosci.* **2021**, *15*, 662935. [CrossRef]
13. Stone, T.W.; Ceruti, S.; Abbracchio, M.P. Adenosine Receptors and Neurological Disease: Neuroprotection and Neurodegeneration. *Handb. Exp. Pharmacol.* **2009**, *193*, 535–587. [CrossRef]
14. Pang, F.; Yang, Y.; Huang, S.; Yang, Z.; Zhu, Z.; Liao, D.; Guo, X.; Zhou, M.; Li, Y.; Tang, C. Electroacupuncture Alleviates Depressive-like Behavior by Modulating the Expression of P2X7/NLRP3/IL-1β of Prefrontal Cortex and Liver in Rats Exposed to Chronic Unpredictable Mild Stress. *Brain Sci.* **2023**, *13*, 436. [CrossRef]
15. Wang, C.; Cui, C.; Xie, X.; Chen, B.; Feng, L.; Jiang, P. Calcitriol Attenuates Lipopolysaccharide-Induced Neuroinflammation and Depressive-like Behaviors by Suppressing the P2X7R/NLRP3/Caspase-1 Pathway. *Psychopharmacology* **2024**, *241*, 1329–1343. [CrossRef]

16. Thawkar, B.S.; Kaur, G. Inhibitors of NF-KB and P2X7/NLRP3/Caspase 1 Pathway in Microglia: Novel Therapeutic Opportunities in Neuroinflammation Induced Early-Stage Alzheimer's Disease. *J. Neuroimmunol.* **2019**, *326*, 62–74. [CrossRef]
17. Moreira-Souza, A.C.A.; Almeida-da-Silva, C.L.C.; Rangel, T.P.; Rocha, G.d.C.; Bellio, M.; Zamboni, D.S.; Vommaro, R.C.; Coutinho-Silva, R. The P2X7 Receptor Mediates Toxoplasma Gondii Control in Macrophages through Canonical NLRP3 Inflammasome Activation and Reactive Oxygen Species Production. *Front. Immunol.* **2017**, *8*, 1257. [CrossRef] [PubMed]
18. Bhattacharya, A.; Jones, D.N.C. Emerging Role of the P2X7-NLRP3-IL1β Pathway in Mood Disorders. *Psychoneuroendocrinology* **2018**, *98*, 95–100. [CrossRef]
19. Stockwell, J.; Jakova, E.; Cayabyab, F.S. Adenosine A1 and A2A Receptors in the Brain: Current Research and Their Role in Neurodegeneration. *Molecules* **2017**, *22*, 676. [CrossRef]
20. Zhou, X.; Fernando, S.M.; Pan, A.Y.; Laposa, R.; Cullen, K.R.; Klimes-Dougan, B.; Andreazza, A.C. Characterizing the NLRP3 Inflammasome in Mood Disorders: Overview, Technical Development, and Measures of Peripheral Activation in Adolescent Patients. *Int. J. Mol. Sci.* **2021**, *22*, 12513. [CrossRef]
21. Chu, J.Q.; Gao, F.F.; Wu, W.; Li, C.; Pan, Z.; Sun, J.; Wang, H.; Huang, C.; Lee, S.H.; Quan, J.H.; et al. Expression Profiles of NOD-like Receptors and Regulation of NLRP3 Inflammasome Activation in Toxoplasma Gondii-Infected Human Small Intestinal Epithelial Cells. *Parasites Vectors* **2021**, *14*, 1–13. [CrossRef] [PubMed]
22. Gov, L.; Schneider, C.A.; Lima, T.S.; Pandori, W.; Lodoen, M.B. NLRP3 and Potassium Efflux Drive Rapid IL-1β Release from Primary Human Monocytes during Toxoplasma Gondii Infection. *J. Immunol.* **2017**, *199*, 2855–2864. [CrossRef] [PubMed]
23. Wang, Z.; Zhang, S.; Xiao, Y.; Zhang, W.; Wu, S.; Qin, T.; Yue, Y.; Qian, W.; Li, L. NLRP3 Inflammasome and Inflammatory Diseases. *Oxid. Med. Cell. Longev.* **2020**, *2020*, 4063562. [CrossRef]
24. Mangan, M.S.J.; Olhava, E.J.; Roush, W.R.; Seidel, H.M.; Glick, G.D.; Latz, E. Targeting the NLRP3 Inflammasome in Inflammatory Diseases. *Nat. Rev. Drug Discov.* **2018**, *17*, 588–606. [CrossRef] [PubMed]
25. Millan, M.J.; Goodwin, G.M.; Meyer-Lindenberg, A.; Ove Ogren, S. Learning from the Past and Looking to the Future: Emerging Perspectives for Improving the Treatment of Psychiatric Disorders. *Eur. Neuropsychopharmacol.* **2015**, *25*, 599–656. [CrossRef]
26. Adams, S.D.; Kouzani, A.Z.; Tye, S.J.; Bennet, K.E.; Berk, M. An Investigation into Closed-Loop Treatment of Neurological Disorders Based on Sensing Mitochondrial Dysfunction. *J. Neuroeng. Rehabil.* **2018**, *15*, 8. [CrossRef]
27. Ma, Q. Pharmacological Inhibition of the NLRP3 Inflammasome: Structure, Molecular Activation, and Inhibitor-NLRP3 Interaction. *Pharmacol. Rev.* **2023**, *75*, 487–520. [CrossRef]
28. Boeing, H.; Bechthold, A.; Bub, A.; Ellinger, S.; Haller, D.; Kroke, A.; Leschik-Bonnet, E.; Müller, M.J.; Oberritter, H.; Schulze, M.; et al. Critical Review: Vegetables and Fruit in the Prevention of Chronic Diseases. *Eur. J. Nutr.* **2012**, *51*, 637–663. [CrossRef]
29. González-Ponce, H.A.; Rincón-Sánchez, A.R.; Jaramillo-Juárez, F.; Moshage, H. Natural Dietary Pigments: Potential Mediators against Hepatic Damage Induced by over-the-Counter Non-Steroidal Anti-Inflammatory and Analgesic Drugs. *Nutrients* **2018**, *10*, 117. [CrossRef]
30. Gambini, J.; Inglés, M.; Olaso, G.; Lopez-Grueso, R.; Bonet-Costa, V.; Gimeno-Mallench, L.; Mas-Bargues, C.; Abdelaziz, K.M.; Gomez-Cabrera, M.C.; Vina, J.; et al. Properties of Resveratrol: In Vitro and In Vivo Studies about Metabolism, Bioavailability, and Biological Effects in Animal Models and Humans. *Oxid. Med. Cell. Longev.* **2015**, *2015*, 837042. [CrossRef]
31. Zhang, L.X.; Li, C.X.; Kakar, M.U.; Khan, M.S.; Wu, P.F.; Amir, R.M.; Dai, D.F.; Naveed, M.; Li, Q.Y.; Saeed, M.; et al. Resveratrol (RV): A Pharmacological Review and Call for Further Research. *Biomed. Pharmacother.* **2021**, *143*, 112164. [CrossRef]
32. Tian, B.; Liu, J. Resveratrol: A Review of Plant Sources, Synthesis, Stability, Modification and Food Application. *J. Sci. Food Agric.* **2020**, *100*, 1392–1404. [CrossRef]
33. Meng, T.; Xiao, D.; Muhammed, A.; Deng, J.; Chen, L.; He, J. Anti-Inflammatory Action and Mechanisms of Resveratrol. *Molecules* **2021**, *26*, 229. [CrossRef] [PubMed]
34. Chiang, M.C.; Nicol, C.J.B.; Lo, S.S.; Hung, S.W.; Wang, C.J.; Lin, C.H. Resveratrol Mitigates Oxygen and Glucose Deprivation-Induced Inflammation, NLRP3 Inflammasome, and Oxidative Stress in 3D Neuronal Culture. *Int. J. Mol. Sci.* **2022**, *23*, 11678. [CrossRef] [PubMed]
35. Zemheri-Navruz, F.; Ince, S.; Arslan-Acaroz, D.; Acaroz, U.; Demirel, H.H.; Demirkapi, E.N. Resveratrol Alleviates Pyraclostrobin-Induced Lipid Peroxidation, Oxidative Stress, and DNA Damage in Rats. *Environ. Sci. Pollut. Res.* **2023**, *30*, 6414–6423. [CrossRef]
36. Fan, W.; Chen, S.; Wu, X.; Zhu, J.; Li, J. Resveratrol Relieves Gouty Arthritis by Promoting Mitophagy to Inhibit Activation of Nlrp3 Inflammasomes. *J. Inflamm. Res.* **2021**, *14*, 3523–3536. [CrossRef] [PubMed]
37. Tufekci, K.U.; Eltutan, B.I.; Isci, K.B.; Genc, S. Resveratrol Inhibits NLRP3 Inflammasome-Induced Pyroptosis and MiR-155 Expression in Microglia Through Sirt1/AMPK Pathway. *Neurotox. Res.* **2021**, *39*, 1812–1829. [CrossRef]

38. Xie, Z.; Ying, Q.; Luo, H.; Qin, M.; Pang, Y.; Hu, H.; Zhong, J.; Song, Y.; Zhang, Z.; Zhang, X. Resveratrol Alleviates Retinal Ischemia-Reperfusion Injury by Inhibiting the NLRP3/Gasdermin D/Caspase-1/Interleukin-1β Pyroptosis Pathway. *Investig. Opthalmol. Vis. Sci.* **2023**, *64*, 28. [CrossRef]
39. Bottari, N.B.; Reichert, K.P.; Fracasso, M.; Dutra, A.; Assmann, C.E.; Ulrich, H.; Schetinger, M.R.C.; Morsch, V.M.; Da Silva, A.S. Neuroprotective Role of Resveratrol Mediated by Purinergic Signalling in Cerebral Cortex of Mice Infected by Toxoplasma Gondii. *Parasitol. Res.* **2020**, *119*, 2897–2905. [CrossRef]
40. Zou, P.; Liu, X.; Li, G.; Wang, Y. Resveratrol Pretreatment Attenuates Traumatic Brain Injury in Rats by Suppressing NLRP3 Inflammasome Activation via SIRT1. *Mol. Med. Rep.* **2018**, *17*, 3212–3217. [CrossRef]
41. Ohto, U.; Kamitsukasa, Y.; Ishida, H.; Zhang, Z.; Murakami, K.; Hirama, C.; Maekawa, S.; Shimizu, T. Structural Basis for the Oligomerization-Mediated Regulation of NLRP3 Inflammasome Activation. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2121353119. [CrossRef] [PubMed]
42. Sharif, H.; Wang, L.; Wang, W.L.; Magupalli, V.G.; Andreeva, L.; Qiao, Q.; Hauenstein, A.V.; Wu, Z.; Núñez, G.; Mao, Y.; et al. Structural Mechanism for NEK7-Licensed Activation of NLRP3 Inflammasome. *Nature* **2019**, *570*, 338–343. [CrossRef] [PubMed]
43. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898. [CrossRef] [PubMed]
44. 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*, 455–461. [CrossRef]
45. Vong, C.T.; Tseng, H.H.L.; Yao, P.; Yu, H.; Wang, S.; Zhong, Z.; Wang, Y. Specific NLRP3 Inflammasome Inhibitors: Promising Therapeutic Agents for Inflammatory Diseases. *Drug Discov. Today* **2021**, *26*, 1394–1408. [CrossRef]
46. Blevins, H.M.; Xu, Y.; Biby, S.; Zhang, S. The NLRP3 Inflammasome Pathway: A Review of Mechanisms and Inhibitors for the Treatment of Inflammatory Diseases. *Front. Aging Neurosci.* **2022**, *14*, 879021. [CrossRef]
47. Davidson, C.B.; El Sabbagh, D.E.S.; Machado, A.K.; Pappis, L.; Sagrillo, M.R.; Somacal, S.; Emanuelli, T.; Schultz, J.V.; Augusto Pereira da Rocha, J.; dos Santos, A.F.; et al. Euterpe Oleracea Mart. Bioactive Molecules: Promising Agents to Modulate the NLRP3 Inflammasome. *Biology* **2024**, *13*, 729. [CrossRef]
48. da Rocha, E.C.M.; da Rocha, J.A.P.; da Costa, R.A.; Costa, A.d.S.S.d.; Barbosa, E.d.S.; Josino, L.P.C.; Brasil, L.d.S.N.d.S.; Vendrame, L.F.O.; Machado, A.K.; Fagan, S.B.; et al. High-Throughput Molecular Modeling and Evaluation of the Anti-Inflammatory Potential of Açaí Constituents against NLRP3 Inflammasome. *Int. J. Mol. Sci.* **2024**, *25*, 8112. [CrossRef]
49. Juhasz, B.; Mukherjee, S.; Das, D.K. Hormetic Response of Resveratrol against Cardioprotection. *Exp. Clin. Cardiol.* **2010**, *15*, 2–6.
50. Calabrese, E.J.; Mattson, M.P.; Calabrese, V. Resveratrol Commonly Displays Hormesis: Occurrence and Biomedical Significance. *Hum. Exp. Toxicol.* **2010**, *29*, 980–1015. [CrossRef]
51. Borriello, A.; Bencivenga, D.; Caldarelli, I.; Tramontano, A.; Borgia, A.; Pirozzi, A.V.A.; Oliva, A.; Ragione, F. Della Resveratrol and Cancer Treatment: Is Hormesis a Yet Unsolved Matter? *Curr. Pharm. Des.* **2013**, *19*, 5384–5393. [CrossRef] [PubMed]
52. Galiniak, S.; Aebisher, D.; Bartusik-Aebisher, D. Health Benefits of Resveratrol Administration. *Acta Biochim. Pol.* **2019**, *66*, 13–21. [CrossRef]
53. Lang, J.; Gao, L.; Wu, J.; Meng, J.; Gao, X.; Ma, H.; Yan, D. Resveratrol Attenuated Manganese-Induced Learning and Memory Impairments in Mice Through PGC-1Alpha-Mediated Autophagy and Microglial M1/M2 Polarization. *Neurochem. Res.* **2022**, *47*, 3414–3427. [CrossRef] [PubMed]
54. Janpaijit, S.; Lertpatipanpong, P.; Sillapachaiyaporn, C.; Baek, S.J.; Charoenkiatkul, S.; Tencomnao, T.; Sukprasansap, M. Anti-Neuroinflammatory Effects of Cleistocalyx Nervosum Var. Paniala Berry-Seed Extract in BV-2 Microglial Cells via Inhibition of MAPKs/NF-KB Signaling Pathway. *Heliyon* **2022**, *8*, e11869. [CrossRef]
55. Bartra, C.; Yuan, Y.; Vuraić, K.; Valdés-Quiroz, H.; Garcia-Baucells, P.; Slevin, M.; Pastorello, Y.; Suñol, C.; Sanfeliu, C. Resveratrol Activates Antioxidant Protective Mechanisms in Cellular Models of Alzheimer's Disease Inflammation. *Antioxidants* **2024**, *13*, 177. [CrossRef]
56. Lu, Y.N.; Shen, X.Y.; Lu, J.M.; Jin, G.N.; Lan, H.W.; Xu, X.; Piao, L.X. Resveratrol Inhibits Toxoplasma Gondii-Induced Lung Injury, Inflammatory Cascade and Evidences of Its Mechanism of Action. *Phytomedicine* **2023**, *108*, 154522. [CrossRef] [PubMed]
57. Madreiter-Sokolowski, C.T.; Sokolowski, A.A.; Graier, W.F. Dosis Facit Sanitatem—Concentration-Dependent Effects of Resveratrol on Mitochondria. *Nutrients* **2017**, *9*, 1117. [CrossRef]
58. Sareen, D.; Darjatmoko, S.R.; Albert, D.M.; Polans, A.S. Mitochondria, Calcium, and Calpain Are Key Mediators of Resveratrol-Induced Apoptosis in Breast Cancer. *Mol. Pharmacol.* **2007**, *72*, 1466–1475. [CrossRef]
59. Ma, X.; Tian, X.; Huang, X.; Yan, F.; Qiao, D. Resveratrol-Induced Mitochondrial Dysfunction and Apoptosis Are Associated with Ca2+ and MCICR-Mediated MPT Activation in HepG2 Cells. *Mol. Cell. Biochem.* **2007**, *302*, 99–109. [CrossRef]

60. Plauth, A.; Geikowski, A.; Cichon, S.; Wowro, S.J.; Liedgens, L.; Rousseau, M.; Weidner, C.; Fuhr, L.; Kliem, M.; Jenkins, G.; et al. Hormetic Shifting of Redox Environment by Pro-Oxidative Resveratrol Protects Cells against Stress. *Free Radic. Biol. Med.* **2016**, *99*, 608–622. [CrossRef]
61. Latz, E.; Xiao, T.S.; Stutz, A. Activation and Regulation of the Inflammasomes. *Nat. Rev. Immunol.* **2013**, *13*, 397–411. [CrossRef] [PubMed]
62. Tschopp, J.; Schroder, K. NLRP3 Inflammasome Activation: The Convergence of Multiple Signalling Pathways on ROS Production? *Nat. Rev. Immunol.* **2010**, *10*, 210–215. [CrossRef] [PubMed]
63. Schroder, K.; Tschopp, J. The Inflammasomes. *Cell* **2010**, *140*, 821–832. [CrossRef] [PubMed]
64. Fukui, M.; Choi, H.J.; Zhu, B.T. Mechanism for the Protective Effect of Resveratrol against Oxidative Stress-Induced Neuronal Death. *Free Radic. Biol. Med.* **2010**, *49*, 800–813. [CrossRef]
65. Kiernan, E.A.; Smith, S.M.C.; Mitchell, G.S.; Watters, J.J. Mechanisms of Microglial Activation in Models of Inflammation and Hypoxia: Implications for Chronic Intermittent Hypoxia. *J. Physiol.* **2016**, *594*, 1563–1577. [CrossRef]
66. Merighi, S.; Nigro, M.; Travagli, A.; Gessi, S. Microglia and Alzheimer's Disease. *Int. J. Mol. Sci.* **2022**, *23*, 12990. [CrossRef]
67. Paolicelli, R.C.; Sierra, A.; Stevens, B.; Tremblay, M.-E.; Aguzzi, A.; Ajami, B.; Amit, I.; Audinat, E.; Bechmann, I.; Bennett, M.; et al. Microglia States and Nomenclature: A Field at Its Crossroads. *Neuron* **2022**, *110*, 3458–3483. [CrossRef]
68. Guo, A.; Chang, Y.; Lin, J.; Guo, J.; He, Y.; Wang, C.; Wu, Z.; Xing, Y.; Jin, F.; Deng, Y. Resveratrol Enhances Anticancer Effects of Silybin on HepG2 Cells and H22 Tumor-Bearing Mice via Inducing G2/M Phase Arrest and Increasing Bax/Bcl-2 Ratio. *Comb. Chem. High Throughput Screen.* **2024**, *27*, 89–98. [CrossRef]
69. Gandy, K.A.O.; Zhang, J.; Nagarkatti, P.; Nagarkatti, M. Resveratrol (3,5,4′-Trihydroxy-Trans-Stilbene) Attenuates a Mouse Model of Multiple Sclerosis by Altering the MiR-124/Sphingosine Kinase 1 Axis in Encephalitogenic T Cells in the Brain. *J. Neuroimmune Pharmacol.* **2019**, *14*, 462–477. [CrossRef]
70. Zhao, H.; Wang, Q.; Cheng, X.; Li, X.; Li, N.; Liu, T.; Li, J.; Yang, Q.; Dong, R.; Zhang, Y.; et al. Inhibitive Effect of Resveratrol on the Inflammation in Cultured Astrocytes and Microglia Induced by Aβ 1–42. *Neuroscience* **2018**, *379*, 390–404. [CrossRef]
71. Pasquini, S.; Contri, C.; Borea, P.A.; Vincenzi, F.; Varani, K. Adenosine and Inflammation: Here, There and Everywhere. *Int. J. Mol. Sci.* **2021**, *22*, 7685. [CrossRef] [PubMed]
72. Alves, V.S.; Leite-Aguiar, R.; da Silva, J.P.; Coutinho-Silva, R.; Savio, L.E.B. Purinergic Signaling in Infectious Diseases of the Central Nervous System. *Brain. Behav. Immun.* **2020**, *89*, 480–490. [CrossRef] [PubMed]
73. Yue, N.; Huang, H.; Zhu, X.; Han, Q.; Wang, Y.; Li, B.; Liu, Q.; Wu, G.; Zhang, Y.; Yu, J. Activation of P2X7 Receptor and NLRP3 Inflammasome Assembly in Hippocampal Glial Cells Mediates Chronic Stress-Induced Depressive-like Behaviors. *J. Neuroinflammation* **2017**, *14*, 102. [CrossRef] [PubMed]
74. Murphy, N.; Cowley, T.R.; Richardson, J.C.; Virley, D.; Upton, N.; Walter, D.; Lynch, M.A. The Neuroprotective Effect of a Specific P2X7 Receptor Antagonist Derives from Its Ability to Inhibit Assembly of the NLRP3 Inflammasome in Glial Cells. *Brain Pathol.* **2012**, *22*, 295–306. [CrossRef]
75. Zhao, J.; Wang, H.; Dai, C.; Wang, H.; Zhang, H.; Huang, Y.; Wang, S.; Gaskin, F.; Yang, N.; Fu, S.M. P2X7 Blockade Attenuates Murine Lupus Nephritis by Inhibiting Activation of the NLRP3/ASC/Caspase 1 Pathway. *Arthritis Rheum.* **2013**, *65*, 3176–3185. [CrossRef]
76. Di Virgilio, F.; Dal Ben, D.; Sarti, A.C.; Giuliani, A.L.; Falzoni, S. The P2X7 Receptor in Infection and Inflammation. *Immunity* **2017**, *47*, 15–31. [CrossRef]
77. Pelegrin, P. P2X7 Receptor and the NLRP3 Inflammasome: Partners in Crime. *Biochem. Pharmacol.* **2021**, *187*, 114385. [CrossRef]
78. Aziz, N.A.W.; Iezhitsa, I.; Agarwal, R.; Bakar, N.S.; Abd. Latiff, A.; Ismail, N.M. Neuroprotection by Trans-Resveratrol in Rats With Intracerebral Hemorrhage: Insights into the Role of Adenosine A1 Receptors. *J. Neuropathol. Exp. Neurol.* **2022**, *81*, 596–613. [CrossRef]
79. Bottari, N.B.; Pillat, M.M.; Schetinger, M.R.C.; Reichert, K.P.; Machado, V.; Assmann, C.E.; Ulrich, H.; Dutra, A.; Morsch, V.M.; Vidal, T.; et al. Resveratrol-Mediated Reversal of Changes in Purinergic Signaling and Immune Response Induced by Toxoplasma Gondii Infection of Neural Progenitor Cells. *Purinergic Signal.* **2019**, *15*, 77–84. [CrossRef]
80. Fracasso, M.; Reichert, K.; Bottari, N.B.; da Silva, A.D.; Schetinger, M.R.C.; Monteiro, S.G.; da Silva, A.S. Involvement of Ectonucleotidases and Purinergic Receptor Expression during Acute Chagas Disease in the Cortex of Mice Treated with Resveratrol and Benznidazole. *Purinergic Signal.* **2021**, *17*, 493–502. [CrossRef]
81. Xie, J.; Liu, S.; Wu, B.; Li, G.; Rao, S.; Zou, L.; Yi, Z.; Zhang, C.; Jia, T.; Zhao, S.; et al. The Protective Effect of Resveratrol in the Transmission of Neuropathic Pain Mediated by the P2X7 Receptor in the Dorsal Root Ganglia. *Neurochem. Int.* **2017**, *103*, 24–35. [CrossRef] [PubMed]
82. Saraiva, M.; O'Garra, A. The Regulation of IL-10 Production by Immune Cells. *Nat. Rev. Immunol.* **2010**, *10*, 170–181. [CrossRef]

83. Yang, J.; Wang, C.; Xin, W.; Liu, J.; Ping, X.; Lu, Y.; Zhao, J.; Pei, L. Exploring the Therapeutic Potential of Baicalin: Mitigating Anxiety and Depression in Epileptic Rats. *Comb. Chem. High Throughput Screen.* **2025**, *28*, 1776–1794. [CrossRef] [PubMed]
84. Ozdamar Unal, G.; Kumbul, D.; Hekimler Ozturk, K.; Erkılınc, G.; Donmez, F.; Dogan Kıran, E.; Yuceer, R.O. The Effect of Vortioxetine on the NLRP3 Pathway and Microglial Activity in the Prefrontal Cortex in an Experimental Model of Depression. *Immunopharmacol. Immunotoxicol.* **2024**, *46*, 264–275. [CrossRef] [PubMed]
85. Rifat, A.; Ossola, B.; Bürli, R.W.; Dawson, L.A.; Brice, N.L.; Rowland, A.; Lizio, M.; Xu, X.; Page, K.; Fidzinski, P.; et al. Differential Contribution of THIK-1 K+ Channels and P2X7 Receptors to ATP-Mediated Neuroinflammation by Human Microglia. *J. Neuroinflammation* **2024**, *21*, 58. [CrossRef]
86. Chai, Y.; Cai, Y.; Fu, Y.; Wang, Y.; Zhang, Y.; Zhang, X.; Zhu, L.; Miao, M.; Yan, T. Salidroside Ameliorates Depression by Suppressing NLRP3-Mediated Pyroptosis via P2X7/NF-KB/NLRP3 Signaling Pathway. *Front. Pharmacol.* **2022**, *13*, 812362. [CrossRef]
87. Li, S.; Zhang, Y.; Wang, Y.; Zhang, Z.; Xin, C.; Wang, Y.; Rong, P. Transcutaneous Vagus Nerve Stimulation Modulates Depression-like Phenotype Induced by High-Fat Diet via P2X7R/NLRP3/IL-1β in the Prefrontal Cortex. *CNS Neurosci. Ther.* **2024**, *30*, e14755. [CrossRef]
88. Valdés-Tresanco, M.S.; Valdés-Tresanco, M.E.; Valiente, P.A.; Moreno, E. AMDock: A Versatile Graphical Tool for Assisting Molecular Docking with Autodock Vina and Autodock4. *Biol. Direct* **2020**, *15*, 12. [CrossRef]
89. da Rocha, J.A.P.; da Costa, R.A.; da Costa, A.d.S.S.; da Rocha, E.C.M.; Gomes, A.J.B.; Machado, A.K.; Fagan, S.B.; Brasil, D.d.S.B.; Lima e Lima, A.H. Harnessing Brazilian Biodiversity Database: Identification of Flavonoids as Potential Inhibitors of SARS-CoV-2 Main Protease Using Computational Approaches and All-Atom Molecular Dynamics Simulation. *Front. Chem.* **2024**, *12*, 1336001. [CrossRef]
90. dos Santos, A.F.; Martins, M.O.; Lameira, J.; de Oliveira Araújo, J.; Frizzo, M.S.; Davidson, C.B.; de Souza, D.V.; Machado, A.K.; Mortari, S.R.; Druzian, D.M.; et al. Evaluation Interaction of Graphene Oxide with Heparin for Antiviral Blockade: A Study of Ab Initio Simulations, Molecular Docking, and Experimental Analysis. *J. Mol. Model.* **2023**, *29*, 235. [CrossRef]
91. 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, Revision A.02*; Gaussian, Inc.: Wallingford, CT, USA, 2016.
92. Woods, R.J.; Chappelle, R. Restrained Electrostatic Potential Atomic Partial Charges for Condensed-Phase Simulations of Carbohydrates. *J. Mol. Struct. THEOCHEM* **2000**, *527*, 149–156. [CrossRef] [PubMed]
93. Case, D.A.; Aktulga, H.M.; Belfon, K.; Cerutti, D.S.; Cisneros, G.A.; Cruzeiro, V.W.D.; Forouzesh, N.; Giese, T.J.; Götz, A.W.; Gohlke, H.; et al. AmberTools. *J. Chem. Inf. Model.* **2023**, *63*, 6183–6191. [CrossRef]
94. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [CrossRef] [PubMed]
95. Gordon, J.C.; Myers, J.B.; Folta, T.; Shoja, V.; Heath, L.S.; Onufriev, A. H++: A Server for Estimating p Ka s and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33*, W368–W371. [CrossRef] [PubMed]
96. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]
97. Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105*, 1902–1921. [CrossRef]
98. da Costa, R.A.; da Rocha, J.A.P.; Pinheiro, A.S.; da Costa, A.D.S.S.; da Rocha, E.C.M.; Silva, R.C.; Gonçalves, A.d.S.; Santos, C.B.R.; Brasil, D.D.S.B. A Computational Approach Applied to the Study of Potential Allosteric Inhibitors Protease NS2B/NS3 from Dengue Virus. *Molecules* **2022**, *27*, 4118. [CrossRef]
99. Loncharich, R.J.; Brooks, B.R.; Pastor, R.W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-acetylalanyl-N′-methylamide. *Biopolymers* **1992**, *32*, 523–535. [CrossRef]
100. 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]
101. 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]
102. Da Costa, R.A.; Da Rocha, J.A.P.; Pinheiro, A.S.; Da Costa, A.S.S.; Da Rocha, E.C.M.; Josino, L.P.C.; Da, A.; Gonçalves, S.; Lima, A.H.L.; Brasil, D.S.B. In Silico Identification of Novel Allosteric Inhibitors of Dengue Virus NS2B/NS3 Serine Protease. *J. Serb. Chem. Soc* **2022**, *87*, 693–706. [CrossRef]
103. Bottari, N.B.; Schetinger, M.R.C.; Pillat, M.M.; Palma, T.V.; Ulrich, H.; Alves, M.S.; Morsch, V.M.; Melazzo, C.; de Barros, L.D.; Garcia, J.L.; et al. Resveratrol as a Therapy to Restore Neurogliogenesis of Neural Progenitor Cells Infected by Toxoplasma Gondii. *Mol. Neurobiol.* **2019**, *56*, 2328–2338. [CrossRef] [PubMed]

104. Copetti, P.M.; Bissacotti, B.F.; da Silva Gündel, S.; Bottari, N.B.; Sagrillo, M.R.; Machado, A.K.; Ourique, A.F.; Chitolina Schetinger, M.R.; Schafer da Silva, A. Pharmacokinetic Profiles, Cytotoxicity, and Redox Metabolism of Free and Nanoencapsulated Curcumin. *J. Drug Deliv. Sci. Technol.* **2022**, *72*, 103352. [CrossRef]
105. Bissacotti, B.F.; Copetti, P.M.; Bottari, N.B.; Gündel, S.d.S.; Machado, A.K.; Sagrillo, M.R.; Ourique, A.F.; Morsch, V.M.M.; da Silva, A.S. Impact of Free Curcumin and Curcumin Nanocapsules on Viability and Oxidative Status of Neural Cell Lines. *Drug Chem. Toxicol.* **2021**, *46*, 155–165. [CrossRef]
106. Pappis, L.; Ramos, A.P.; Fontana, T.; Sangoi, G.G.; Dornelles, R.C.; Dolwitsch, C.B.; Sagrillo, M.R.; Cadoná, F.C.; Machado, A.K.; Bauermann, L.d.F. Randia Ferox (Cham & Schltdl) DC.: Phytochemical Composition, in Vitro Cyto- and Genotoxicity Analyses. *Nat. Prod. Res.* **2022**, *36*, 4170–4176. [CrossRef]
107. Copetti, P.M.; Gündel, S.d.S.; de Oliveira, P.S.B.; Favarin, F.R.; Ramos, A.P.; Pintos, F.G.; Pappis, L.; Gündel, A.; Machado, A.K.; Ourique, A.F.; et al. Development, Characterisation, Stability Study and Antileukemic Evaluation of Nanoemulsions Containing Astrocaryum Aculeatum Extract. *Nat. Prod. Res.* **2020**, *36*, 1321–1326. [CrossRef]
108. de Souza, D.V.; Pappis, L.; Bandeira, T.T.; Sangoi, G.G.; Fontana, T.; Rissi, V.B.; Sagrillo, M.R.; Duarte, M.M.; Duarte, T.; Bodenstein, D.F.; et al. Açaí (Euterpe Oleracea Mart.) Presents Anti-Neuroinflammatory Capacity in LPS-Activated Microglia Cells. *Nutr. Neurosci.* **2022**, *25*, 1188–1199. [CrossRef] [PubMed]
109. El Sabbagh, D.; Attisano, L.; Andreazza, A.C.; Machado, A.K. A Dynamic Protocol to Explore NLRP3 Inflammasome Activation in Cerebral Organoids. *Int. J. Mol. Sci.* **2024**, *25*, 6335. [CrossRef]
110. Mosmann, T. Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays. *J. Immunol. Methods* **1983**, *65*, 55–63. [CrossRef]
111. Ahn, S.; Costa, J.; Emanuel, J. PicoGreen Quantitation of DNA: Effective Evaluation of Samples Pre- or Post-PCR. *Nucleic Acids Res.* **1996**, *24*, 2623–2625. [CrossRef]
112. Choi, W.; Shin, P.; Lee, J.; Kim, G. The Regulatory Effect of Veratric Acid on NO Production in LPS-Stimulated. *Cell. Immunol.* **2012**, *280*, 164–170. [CrossRef] [PubMed]
113. Costa, F.; Dornelles, E.; Mânica-Cattani, M.F.; Algarve, T.D.; de Souza Filho, O.C.; Sagrillo, M.R.; Garcia, L.F.M.; da Cruz, I.B.M. Influence of Val16Ala SOD2 Polymorphism on the In-Vitro Effect of Clomiphene Citrate in Oxidative Metabolism. *Reprod. Biomed. Online* **2012**, *24*, 474–481. [CrossRef] [PubMed]
114. William-Faltaos, S.; Rouillard, D.; Lechat, P.; Bastian, G. Cell Cycle Arrest and Apoptosis Induced by Oxaliplatin (L-OHP) on Four Human Cancer Cell Lines. *Anticancer Res.* **2006**, *26*, 2093–2099. [PubMed]
115. McLaren, F.H.; Svendsen, C.N.; Van der Meide, P.; Joly, E. Analysis of Neural Stem Cells by Flow Cytometry: Cellular Differentiation Modifies Patterns of MHC Expression. *J. Neuroimmunol.* **2001**, *112*, 35–46. [CrossRef]
116. Rebola, N.; Pinheiro, P.C.; Oliveira, C.R.; Malva, J.O.; Cunha, R.A. Subcellular Localization of Adenosine A1 Receptors in Nerve Terminals and Synapses of the Rat Hippocampus. *Brain Res.* **2003**, *987*, 49–58. [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.

pharmaceuticals

MDPI

Article

Systems Biology-Driven Discovery of Host-Targeted Therapeutics for Oropouche Virus: Integrating Network Pharmacology, Molecular Docking, and Drug Repurposing

Pranab Dev Sharma [1], Abdulrahman Mohammed Alhudhaibi [2,*], Abdullah Al Noman [3], Emad M. Abdallah [4], Tarek H. Taha [2] and Himanshu Sharma [5]

[1] Biotechnology Program, Department of Mathematics and Natural Science, BRAC University, Dhaka 1212, Bangladesh; pranab.dev.sharma@g.bracu.ac.bd
[2] Department of Biology, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11623, Saudi Arabia; thali@imamu.edu.sa
[3] School of Pharmacy, BRAC University, Dhaka 1212, Bangladesh; abdullah.al.noman@g.bracu.ac.bd
[4] Department of Biology, College of Science, Qassim University, Buraydah 51452, Saudi Arabia; 140208@qu.edu.sa
[5] Teerthanker Mahaveer College of Pharmacy, Teerthanker Mahaveer University, Moradabad 244001, Uttar Pradesh, India; amitsharmaaligarh786@gmail.com
* Correspondence: amalhudhaibi@imamu.edu.sa

Abstract: Background: Oropouche virus (OROV), part of the Peribunyaviridae family, is an emerging pathogen causing Oropouche fever, a febrile illness endemic in South and Central America. Transmitted primarily through midge bites (*Culicoides paraensis*), OROV has no specific antiviral treatment or vaccine. This study aims to identify host-targeted therapeutics against OROV using computational approaches, offering a potential strategy for sustainable antiviral drug discovery. **Methods:** Virus-associated host targets were identified using the OMIM and GeneCards databases. The Enrichr and DSigDB platforms were used for drug prediction, filtering compounds based on Lipinski's rule for drug likeness. A protein–protein interaction (PPI) network analysis was conducted using the STRING database and Cytoscape 3.10.3 software. Four key host targets—IL10, FASLG, PTPRC, and FCGR3A—were prioritized based on their roles in immune modulation and OROV pathogenesis. Molecular docking simulations were performed using the PyRx software to evaluate the binding affinities of selected small-molecule inhibitors—Acetohexamide, Deptropine, Methotrexate, Retinoic Acid, and 3-Azido-3-deoxythymidine—against the identified targets. **Results:** The PPI network analysis highlighted immune-mediated pathways such as Fc-gamma receptor signaling, cytokine control, and T-cell receptor signaling as critical intervention points. Molecular docking revealed strong binding affinities between the selected compounds and the prioritized targets, suggesting their potential efficacy as host-targeting antiviral candidates. Acetohexamide and Deptropine showed strong binding to multiple targets, indicating broad-spectrum antiviral potential. Further in vitro and in vivo validations are needed to confirm these findings and translate them into clinically relevant treatments. **Conclusions:** This study highlights the potential of using computational approaches to identify host-targeted therapeutics for Oropouche virus (OROV). By targeting key host proteins involved in immune modulation—IL10, FASLG, PTPRC, and FCGR3A—the selected compounds, Acetohexamide and Deptropine, demonstrate strong binding affinities, suggesting their potential as broad-spectrum antiviral candidates. Further experimental validation is needed to confirm their efficacy and potential for clinical application, offering a promising strategy for sustainable antiviral drug discovery.

Keywords: computational biology; drug development; virus–host protein interaction; neglected tropical diseases; molecular docking validation

1. Introduction

In an era when climate change and urbanization contribute to the spread of neglected diseases such as arboviruses (arthropod-borne viruses), the Oropouche virus (OROV) has emerged as a significant yet underrecognized threat, causing outbreaks in tropical regions such as Central and South America. Currently, no antivirals or vaccines are available to prevent its transmission. Although no fatalities have been reported, OROV can lead to systemic infections, affecting both the nervous and circulatory systems and potentially resulting in severe complications [1]. The Oropouche virus (OROV) belongs to the order Bunyavirales, family Peribunyaviridae, and genus Orthobunyavirus, classified under the *Orthobunyavirus oropoucheense* species. It is characterized by a negative-sense, single-stranded RNA genome enclosed within a spherical lipid envelope [2,3]. The transmission of the virus occurs principally via the bites of midges (*Culicoides paraensis*) and to a lesser extent by mosquitoes [4]. Epidemics of OROV have been reported in Brazil, Panama, and Peru, and increasing cases are pointing toward broader dissemination [5]. Its clinical features resemble other arboviral diseases including dengue fever, with fever, headache, myalgia, and sometimes neurological complications [6]. Surprisingly, limited information is available regarding the vectors responsible for the ongoing epidemic. Current evidence indicates that, unlike many other arthropod-borne orthobunyaviruses, Oropouche virus (OROV) is primarily transmitted by culicoid midges (Ceratopogonidae: Culicoides), rather than by mosquitoes or ticks [7].

Historically, this virus, OROV, was first identified in Trinidad and Tobago during an outbreak of febrile illness. Since its discovery, epidemics have been documented in various countries across South and Central America, including Brazil, Peru, and Panama [8]. In Brazil alone, the Oropouche virus has been responsible for over 500,000 reported cases, with recent outbreaks linked to fatalities, suspected vertical transmission, and possible associations with microcephaly in newborns. The true burden of Oropouche fever is likely underestimated due to underreporting and inadequate surveillance. Additionally, the limited availability of diagnostic methods, such as serology and RT-PCR, in resource-constrained settings further hampers efforts to monitor and control the virus's spread [9]. It, however, has no specific antiviral treatment or vaccine despite its epidemiological importance, given that further speculative research on possible suitable therapeutic interventions to target the cellular factors of the host exploited by viruses for replication and survival is quite a promising strategy in antiviral drug discovery. In contrast to direct-acting antivirals, which often face obstacles emerging from viral mutations and resistance, host-targeted therapeutics offer a much wider and perhaps more sustainable approach for the management of viral infections.

For neglected diseases such as Oropouche virus (OROV), which lack targeted therapeutic options, the integration of systems biology and computational drug discovery represents a transformative approach to addressing emerging viral threats [10]. Although the application of these strategies to arboviruses is still in its early stages, recent advances in network pharmacology and multi-omics analyses have uncovered host dependency factors essential for viral replication [11].

Despite the increasing public health threat posed by OROV, significant gaps remain in our understanding of its molecular etiology and in the development of targeted thera-

pies, highlighting a critical global research deficit. Current studies have largely focused on epidemiological surveillance, with limited mechanistic insight into host–virus interactions or immune evasion strategies [12]. Although host-targeted antivirals have shown promise against other arboviruses such as dengue and Zika [13], similar approaches remain largely unexplored for OROV, reflecting a broader trend of systematic neglect toward underprioritized pathogens within global drug development pipelines.

While computational strategies like network pharmacology have proven effective in identifying host-directed therapies for viruses such as SARS-CoV-2 [14], these methods have yet to be applied to OROV. As a result, its druggable host pathways remain uncharted. Furthermore, regional disparities in research funding have hindered the integration of drug repurposing strategies [15], an essential component of pandemic preparedness, into OROV-related investigations.

This study introduces a multi-layered computational framework that integrates network-based target prioritization, robust molecular docking, and drug repurposing—a cost-effective and efficient strategy particularly suitable for neglected diseases—to identify host-directed therapeutics against OROV. By mapping virus–host interactomes and conducting functional enrichment analyses, we elucidate immune-modulatory pathways (e.g., Fcγ receptor signaling, cytokine regulation) hijacked by OROV, thereby identifying mechanistically relevant host targets (IL10, FASLG, PTPRC, FCGR3A) implicated in both viral pathogenesis and immune homeostasis.

Our approach not only prioritizes repurposed drugs with established binding affinities (e.g., Methotrexate, Retinoic Acid), but also mitigates the risk of resistance commonly associated with direct-acting antivirals. Importantly, this work addresses a critical gap in OROV research, which has traditionally focused on epidemiological surveillance rather than mechanistic, host-targeted interventions.

The computational rigor of our method—including the application of Lipinski's rule-based filters, protein–protein interaction (PPI) network clustering, and Gene Ontology (GO) term enrichment—ensures translational relevance. Furthermore, the identification of druggable immune pathways offers broader implications for combating related arboviruses. By bridging systems biology with antiviral discovery, this study provides a scalable and adaptable framework for rapid therapeutic development, positioning host-targeted strategies as a cornerstone of pandemic preparedness in an era of climate-driven arboviral emergence.

While this study presents a computational approach for host-directed antiviral candidate finding, further in vitro and in vivo validations are needed for confirming efficacy. This study highlights the opportunities present in computational drug discovery to avert the cry of any mathematical interludes ushered by OROV. By coupling network pharmacology with molecular docking, this finds sundry drug suggestions on basis of advanced study in host-targeted therapeutics for the advancement of antiviral weaponry against Oropouche virus infections.

2. Results

2.1. Targets for Virus

A total of 214 target genes were obtained from the OMIM and Genecard databases after removing 110 duplicate genes. After that, those virus target genes were subjected to mapping into the UniPort database in the condition in Homo sapiens where 7 genes were not mapped and the other 207 targets were accepted for mapping. These 207 mapped genes were used in other steps.

2.2. Drug Prediction

We used the identified 207 genes for our drug prediction process to see which drugs came out associated with the targets by using the DSigDB database. From that finding, top 10 compounds came out for drug use (Figure 1).

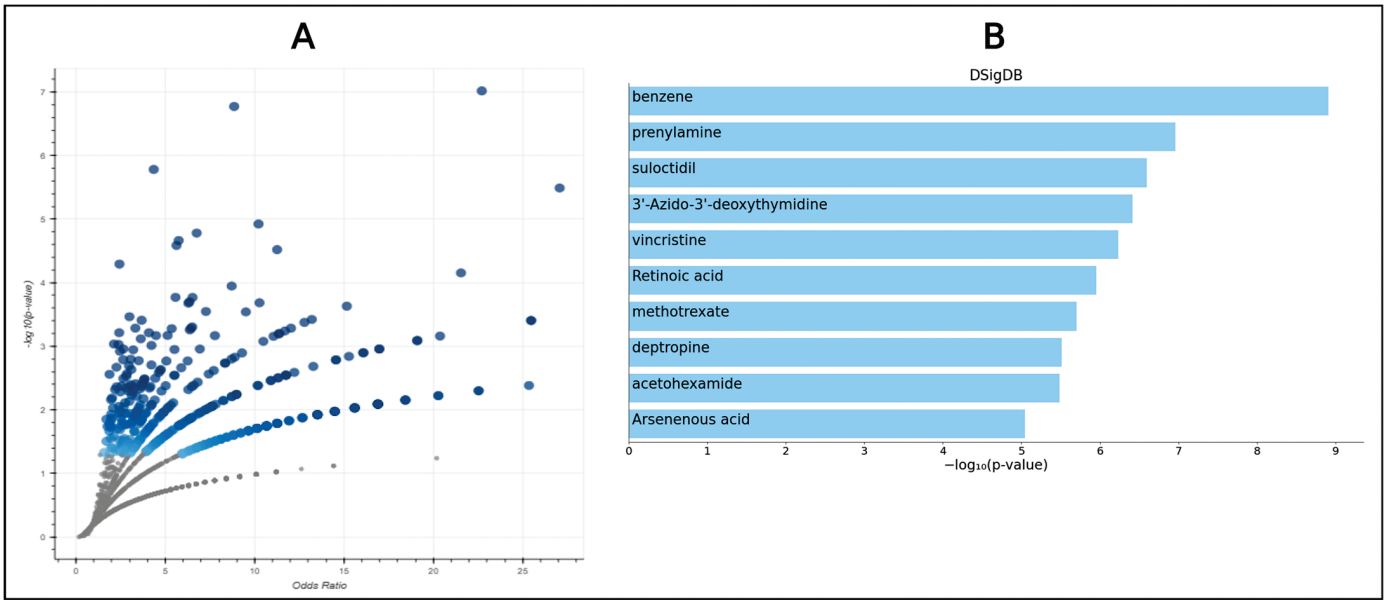


Figure 1. Volcano plot and bar chart for the selected top compounds. (**A**) In the volcano plot, each of the compound is represented by each dot where odds ratio lies along the *x*-axis and −log10(*p*-value) along the y-axis in relation to the enrichment results of the input query gene set. The term is further ordered based on the set enrichment *p*-value for the supplied input gene set, where larger and darker points signify greater terms. Moreover, the colors in this figure highlight significance levels, with darker blue representing highly significant values and lighter shades indicating lower significance. Gray points likely correspond to non-significant data, ensuring contrast in visualization. (**B**) Based on the −log(*p* value), the top 10 enriched terms are displayed for the input gene set. These terms have the highest overlap with the input query gene set. Moreover, the blue-colored bars correspond to terms with significant *p*-values (<0.05).

2.3. Compound Selection

For known toxicity, compounds should be discontinued from the market as they do not follow the criteria. Five compounds out of ten were removed from further study. The other five molecules were selected, which followed most of the pharmacokinetic criteria (Table 1).

Table 1. Different pharmacokinetic criteria.

Name	Weight	H-Bond Acceptor	H-Bond Donor	TPSA	MlogP
3-Azido-3-deoxythymidine	267.24 g/mol	7	2	134.07	−1.25
Retinoic Acid	300.44 g/mol	2	1	37.30	-
Methotrexate	454.44 g/mol	9	5	210.54	−1.13
Deptropine	333.47 g/mol	2	0	12.47	−4.12
Acetohexamide	324.40 g/mol	4	2	100.72	1.12

2.4. Target Collection for Compounds

According to the different databases, a total of 561 targets were identified for 3-Azido-3-deoxythymidine, 9540 targets for Retinoic Acid, 2691 targets for Methotrexate, 100 for

Deptropine, and a total of 115 targets were collected for Acetohexamide. Furthermore, those targets were used to map in the UniProt database by focusing on *Homo sapiens*, where 10,011 targets were mapped and 802 targets were not mapped and removed from the list.

2.5. PPI Network Analysis

From the previous analysis, 207 target genes were collected for the virus and 10,011 targets for potential drugs. By utilizing the Venny tool, we found out the common targets for both specific viruses and compounds. The graph shows a total of 146 common targets (Supplementary Table S1, Figure 2A).

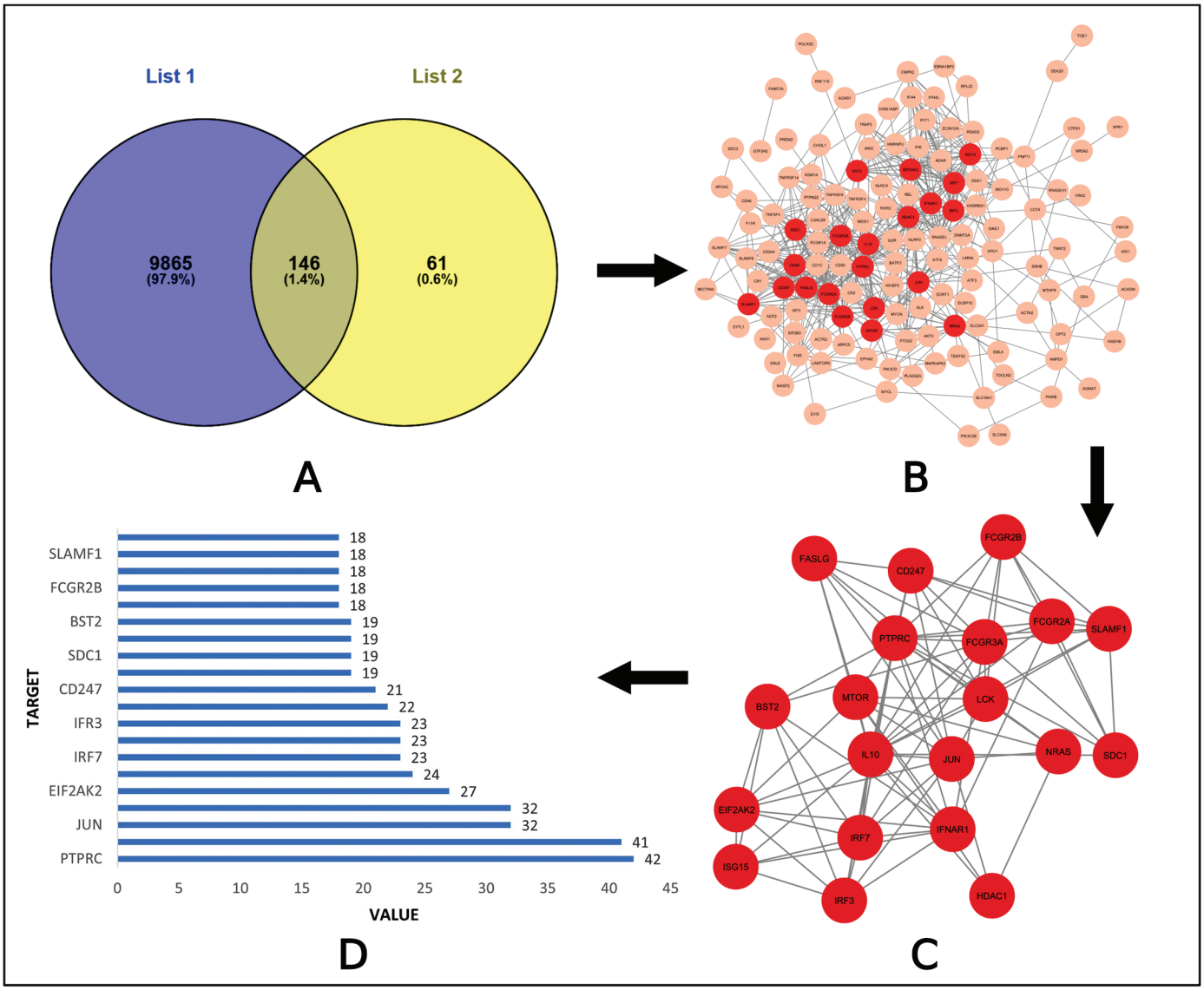


Figure 2. Screening the common targets and their networking interactions. (**A**) Venn diagram showing the intersect targets. (**B**) Protein–protein interaction for common targets. (**C**) Top 20 proteins among the targets. (**D**) List of top proteins with degree value.

The intersect genes were subjected to an insert into string for constructing a PPI network that had 133 nodes and 597 edges (Figure 2B). We focused on the degree value because a higher degree value means more interactions in the network and shows more importance. We calculated the top 20 hub genes based on the degree value (Figure 2C). The degree value for each top node is shown in Figure 2D. PTPRC, IL10, JUN, FCGR3A,

EIF2AK2, LCK, IRF7, IRF3, FCGR2A, IFNAR, CD247, MTOR, SDC1, HDAC1, BST2, NRAS, FCGR2B, FASLG, SLAMF1, and ISG15 were the identified key node proteins.

2.6. MCODE Analysis

To identify core therapeutic targets, the MCODE algorithm was applied, by which we constructed a modular network. Different key functions, such as positive regulation of cytokine production, response to virus, regulation of immune effector process, and positive regulation in immune response, were identified from the topological analysis in Metascape (Figure 3A). A subset network was generated to understand the relationship between pathways and targets (Figure 3B). This network helps us to find out different important functions of targets within different clusters. Moreover, Metascape performed MCODE cluster analysis of the 146 intersect genes. That analysis provided three target modules for treating Oropouche virus by using the selected compounds. These modules help to obtain the biological pathways that might be present in therapeutic impacts on the virus infection.

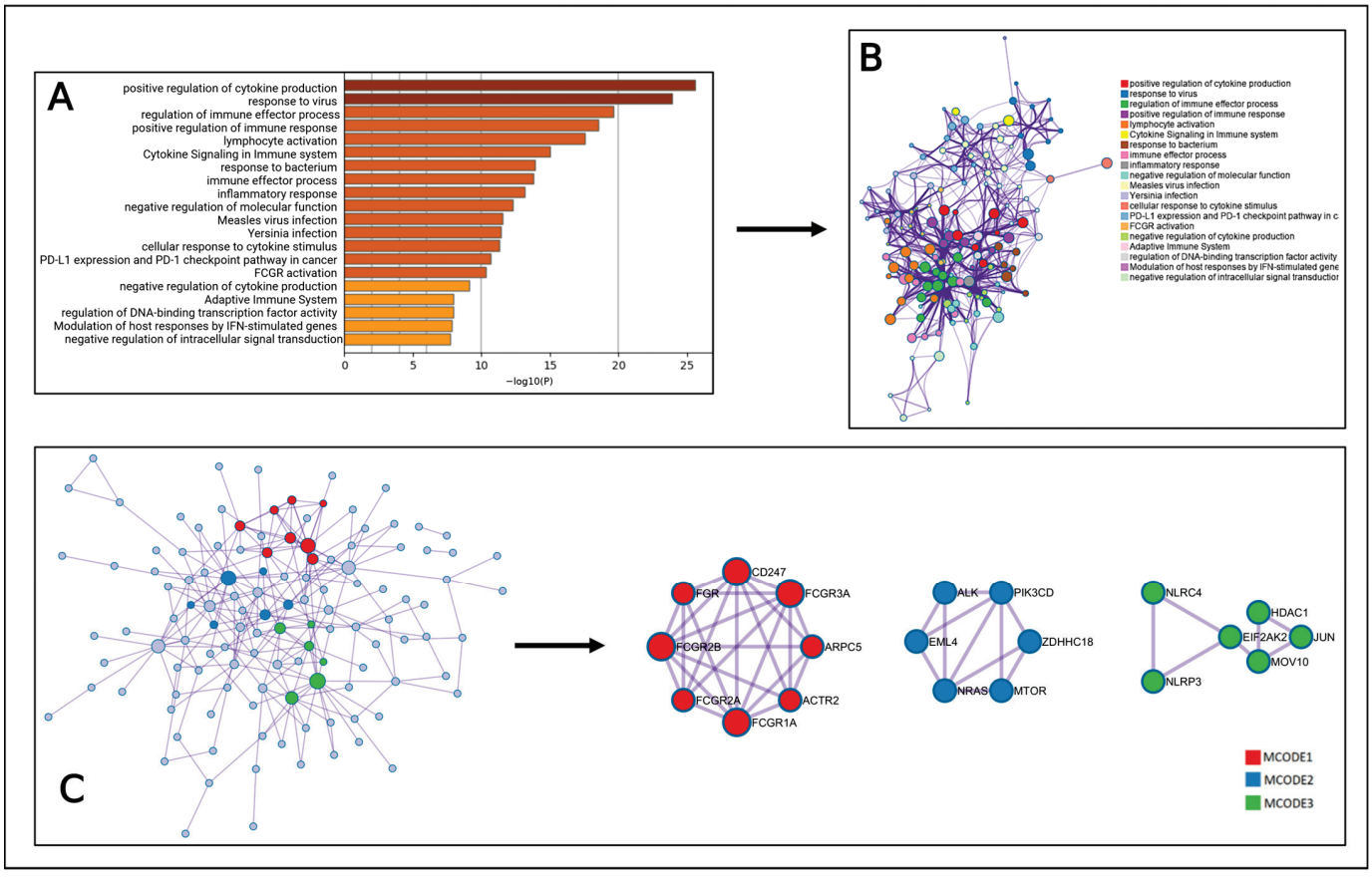


Figure 3. Cluster module analysis diagram. (**A**) Highly enriched terms. The colors in the bars represent the significance of different biological processes, pathways, or functions. Darker shades indicate higher significance levels based on the $-\log10(p)$ value, while lighter shades represent lower significance. This helps visually distinguish the most statistically relevant terms in the dataset. (**B**) Sub-network formation on specific interaction. (**C**) Different cluster analysis.

MCODE 1 is enriched with three different pathway and eight genes, including FGR, CD247, FCGR3A, ARPC5, ACTR2, FCGR1A, FCGR2A, and FCGR2B (Figure 3C, Table 2). Additionally, MCODE 2 and MCODE 3 both showed three pathways, where EML4, NRAS, MTOR, ZDHHC18, PIK3CD, and ALK are the genes present in MCODE 2 and NLRP3, NLRC4, EIF2AK2, HDAC1, JUN, and MOV10 are for MCODE 3 (Figure 3C, Table 2).

Table 2. Different MCODE cluster analysis.

Color	MCODE	GO	Description	$\log_{10}(p)$
Red	MCODE_1	R-HSA-2029480	Fc-gamma receptor (FCGR)-dependent phagocytosis	−17.9
Red	MCODE_1	R-HSA-2029481	FCGR activation	−17.3
Red	MCODE_1	GO:0038094	Fc-gamma receptor signaling pathway	−16.3
Blue	MCODE_2	hsa05235	PD-L1 expression and PD-1 checkpoint pathway in cancer	−11.9
Blue	MCODE_2	WP4255	Non-small cell lung cancer	−9.4
Blue	MCODE_2	hsa05223	Non-small cell lung cancer	−9.3
Green	MCODE_3	R-HSA-3700989	Transcriptional regulation by TP53	−6.9
Green	MCODE_3	R-HSA-9700026	Signaling by ALK in cancer	−6.7
Green	MCODE_3	R-HSA-9725370	Signaling by ALK fusions and activated point mutants	−6.7

2.7. Gene Ontology and KEGG Pathway Analysis

To perform GO functional and KEGG pathway enrichment analysis, the 146 intersecting genes were analyzed in the DAVID online platform. The analysis displayed a total of 312 entries, where 162 were involved in biological process (BP) (Supplementary Table S2), 31 entries for cellular component (CC) (Supplementary Table S3), 49 for molecular function (MF) (Supplementary Table S4), and 70 results for KEGG pathway (Supplementary Table S5). Innate immune response, positive regulation of transcription by RNA polymerase II, defense response to virus, immune response, and chromatin remodeling were present in biological process. The CC terms displayed cytosol, cytoplasm, membrane, plasma membrane, and nucleus, while the MF terms included organic cyclic compound binding, nucleic acid binding, identical protein binding, enzyme binding, and DNA binding. Top 10 entries based on the gene count value are shown in the enrichment bar chart (Figure 4A).

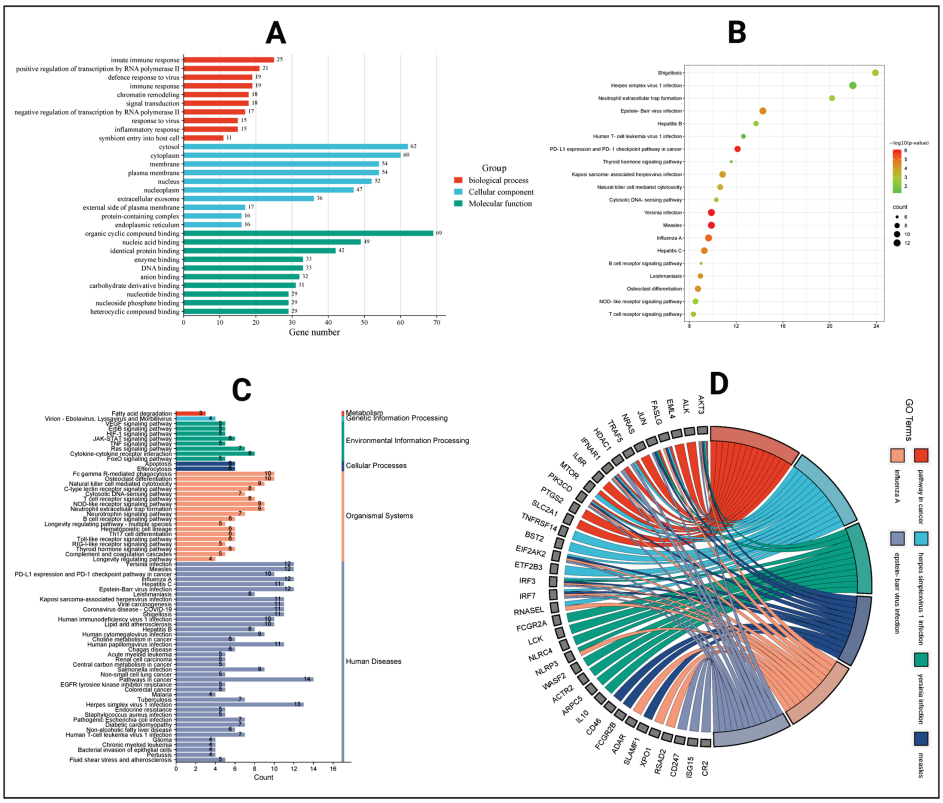


Figure 4. Gene ontology and KEGG pathway analysis. (**A**) Different gene ontology functions in bar chart. (**B**) KEGG enrichment in bubble plot. (**C**) Different systems and actions of KEGG pathways. (**D**) Pathway network for KEGG pathways.

A total of 70 pathways were identified by the KEGG pathway analysis, where the pathway in cancer and the herpes simplex virus 1 infection pathway had the highest gene value (Figure 5). Another 20 pathways are presented in the bubble plots selected by *p*-value (Figure 4B). Furthermore, all the 70 pathways of KEGG data were mapped based on actions and functions, which selected 1 pathway in metabolism, 1 in genetic information process, 8 pathways in environmental information, 2 are present in cellular process, 18 in organism systems, and human disease displayed 40 pathways (Figure 4C). Moreover, with the highest amount of genes the top six pathways are presented in the chord diagram (Figure 4D).

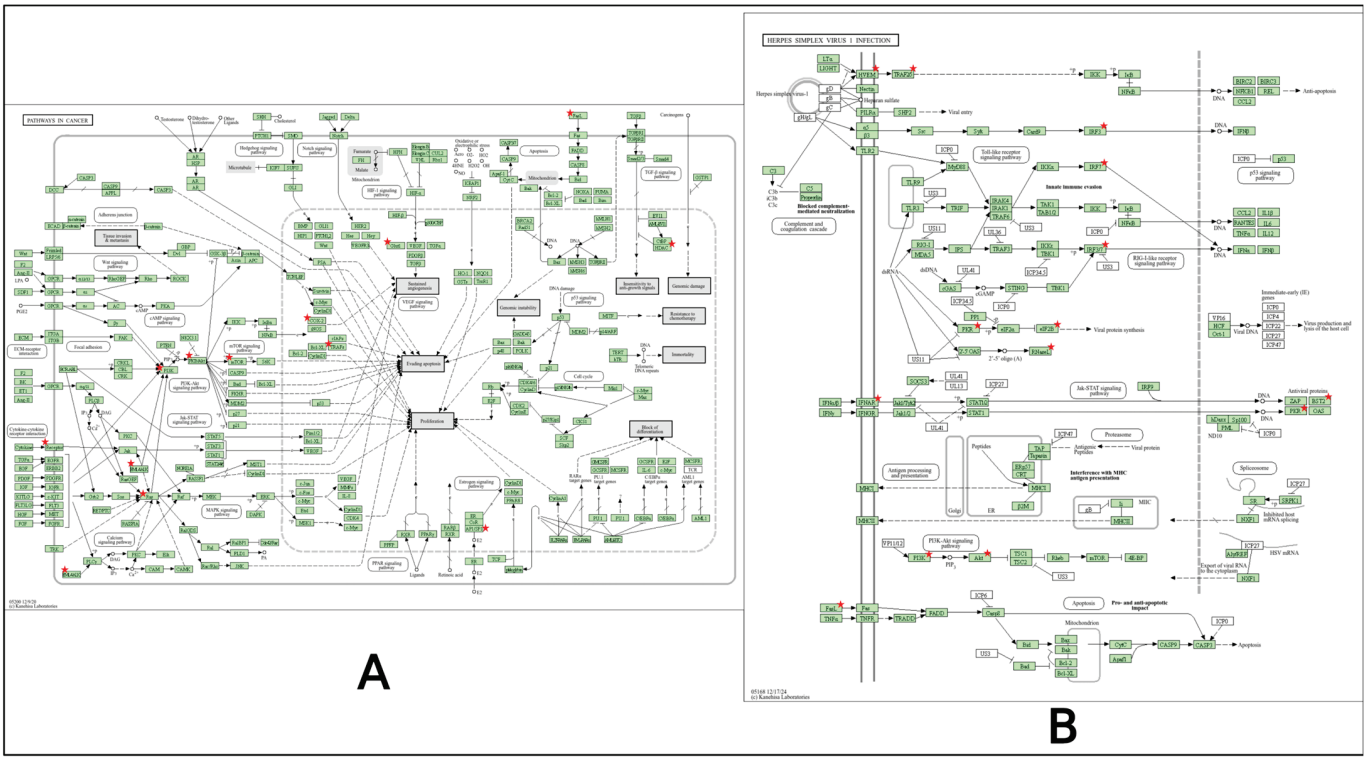


Figure 5. Top 2 pathways based on the gene count value. Red stars indicate key regulatory nodes within the pathway, highlighting critical molecular interactions associated with cancer progression and Herpes simplex virus 1 infection. These elements represent significant genes or proteins involved in cellular signaling and disease mechanisms. (**A**) Pathway in cancer. (**B**) Herpes simplex virus 1 infection.

2.8. Target to Pathway Network

Using the Cytoscape 3.10.3 software, we constructed a target to pathway network involving 162 nodes and 1018 edges (Figure 6A). This network shows interaction among virus, targets, drug, and pathways. The red diamond shape represents the virus, green triangle shapes are the pathways that were identified in MCODE clustering, V-shapes indicates drug components, and the orange-colored circles are the targets.

After that, a network was constructed with the top 20 targets based on their degree value by using the cytohubba plugin in Cytoscape (Figure 6B), and their degree values are presented in the bar chart (Figure 6C). The top 20 targets from protein–protein interaction and top 20 targets from target–pathway network interaction was used to intersect to obtain the common targets and were visualized with a Venn diagram (Figure 6D). Finally, PTPRC, IL10, JUN, FCGR3A, EIF2AK2, LCK, IRF7, CD247, MTOR, SDC1, HDAC1, BST2, NRAS, FCGR2B, FASLG, and ISG15 were identified as core targets.

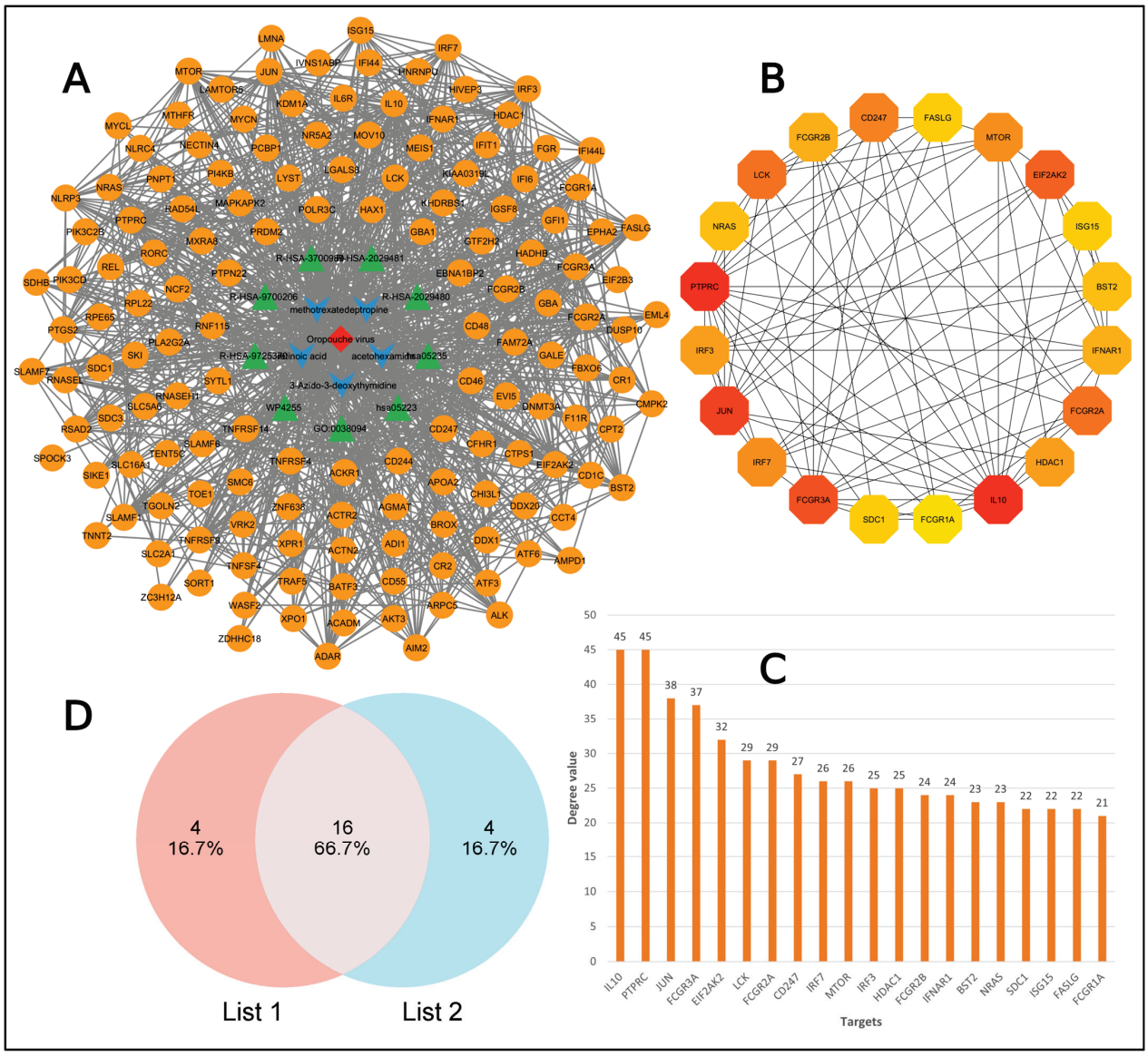


Figure 6. Visualization of target–pathway network interaction. (**A**) Virus–drug–pathway–target network diagram. (**B**) Network of top 20 targets. (**C**) Bar chart for top 20 genes with the degree value. (**D**) Core target identification between PPI network (list 1) and target–pathway network (list 2).

2.9. Key Target Identification

The PPI network from the String database was use in the Cytoscape software to obtain top ten genes for the five different topological properties (Maximal Clique Centrality (MCC), Closeness, BottleNeck, Edge Percolated Component (EPC), Degree value, and Radiality). We created a simple network of those 16 targets to visualize their interactions (Figure 7A). The upset plot shows four targets that are common for all six topological properties in the first bar (Supplementary Table S6, Figure 7B). The first bar represents FASLG, FCGR3A, IL10, and PTPRC as common targets for all the selected properties. Also, the heat map in Figure 7C shows the common genes between two properties. Moreover, we checked different Gene Ontology properties for the top 16 targets. Top five terms for biological processes (BP), cellular component (CC), molecular function (MF) (only has three terms), and KEGG pathways of those genes are visualized in Figure 7D. Defense response, low-affinity lgG receptor activity, Fc-gamma receptor III complex, and T-cell receptor signaling pathway some are the necessary pathways for the 16 hub genes.

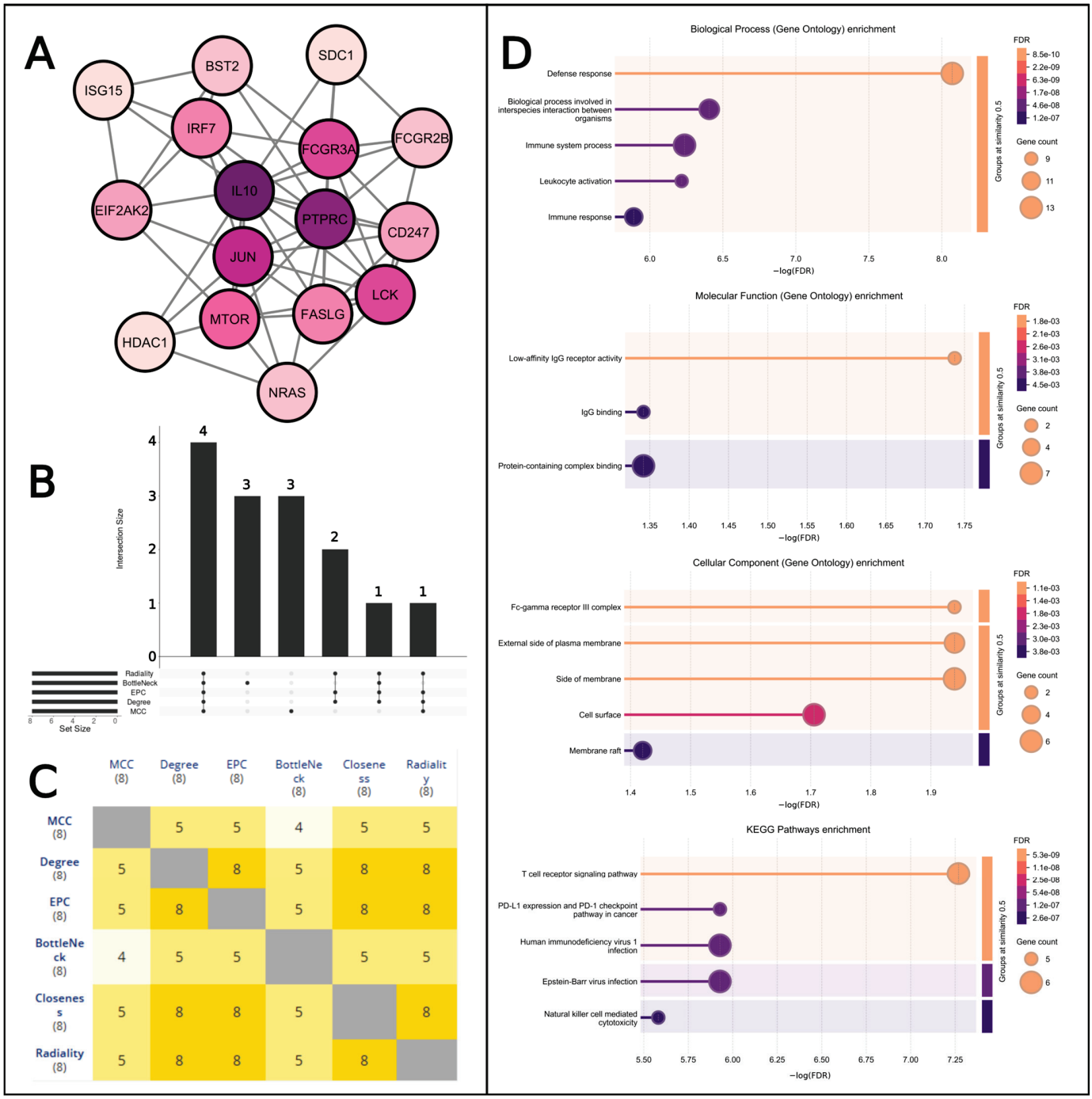


Figure 7. Identification of common targets for different properties and their pathways. (**A**) Network construction between the hub genes. Here, the color gradient likely reflects variations in degree centrality, with darker shades indicating nodes with higher connectivity and lighter shades representing lower-degree values. (**B**) Determining common targets based on topological properties. (**C**) Heat map for showing interaction among two properties. (**D**) Enrichment analysis for the regulatory network. The color gradient in the bars represents False Discovery Rate (FDR) values, providing a visual cue for statistical significance—warmer tones (orange to pink) indicate higher confidence, while cooler shades (purple) reflect decreasing significance. The orange color marks a strong statistical threshold, ensuring data reliability with minimal false positives while acting as a midpoint between highly significant and slightly higher FDR values. This strategic color choice enhances clarity, making robust findings stand out while maintaining a smooth visual transition across the significance scale.

2.10. Docking Result

By yielding the binding energies among the compounds (small molecules) and proteins (large molecules), the docking interactions among the identified compounds and proteins were validated. The lower the binding affinity, the stronger the interaction between a compound and protein. The binding results of the targets and compounds are summarized in Table 3 and all the affinities were less than −5 kcal/mol.

Table 3. Comparison of SWISSDOCK and original docking results for molecular binding analysis.

Compound Name and CID	Protein Name and Binding Affinity							
	PyRx	SwissDock	PyRx	SwissDock	PyRx	SwissDock	PyRx	SwissDock
	IL10		FASLG		PTPRC		FCGR3A	
Acetohexamide, CID: 1989	−7.3 kcal/mol	−7.3 kcal/mol	−6.5 kcal/mol	−6.8 kcal/mol	−7.3 kcal/mol	−7.1 kcal/mol	−6.7 kcal/mol	−7.0 kcal/mol
Deptropine, CID: 203911	−8.3 kcal/mol	−7.0 kcal/mol	−7.4 kcal/mol	−6.9 kcal/mol	−7.0 kcal/mol	−6.4 kcal/mol	−7.1 kcal/mol	−6.8 kcal/mol
Methotrexate, CID: 126941	−6.8 kcal/mol	−7.8 kcal/mol	−7.4 kcal/mol	−7.4 kcal/mol	−7.7 kcal/mol	−7.4 kcal/mol	−6.7 kcal/mol	−8.1 kcal/mol
Retinoic Acid, CID: 449171	−8.1 kcal/mol	−7.4 kcal/mol	−6.7 kcal/mol	−7.1 kcal/mol	−6.3 kcal/mol	−7.1 kcal/mol	−5.8 kcal/mol	−6.9 kcal/mol
3-Azido-3-deoxythymidine, CID: 5726	−5.3 kcal/mol	−6.8 kcal/mol	−5.7 kcal/mol	−6.8 kcal/mol	−5.6 kcal/mol	−6.8 kcal/mol	−5.1 kcal/mol	−6.7 kcal/mol

The five identified compounds were selected to determine the combinations with four core targets. The visualized results are presented in Figures 8–11. Usually, a binding value lower than −7 kcal/mol means a strong predicted binding, and a value lower than −5 kcal/mol suggests moderate binding [16].

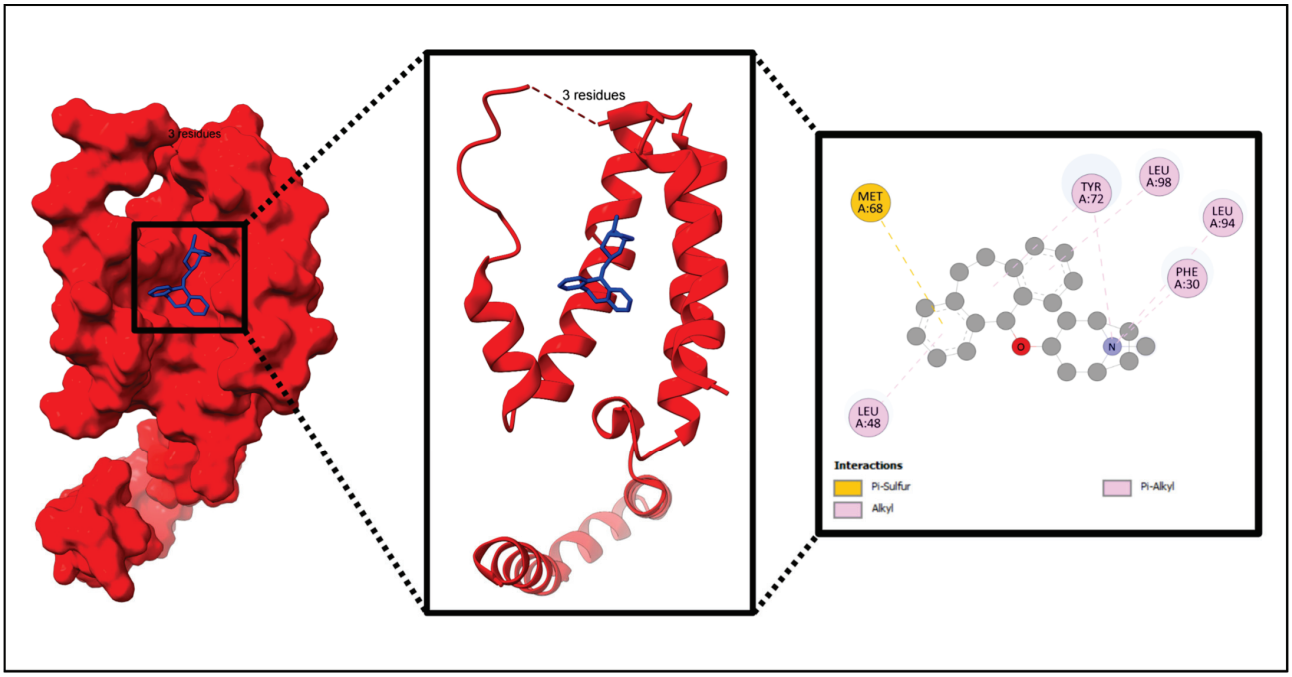


Figure 8. Molecular interaction of IL10 with the selected compound (IL10 + Deptropine). In this figure, red color indicates the protein and blue color indicates the compound. Additional data related to Figure 8 can be found in Supplementary Figures S1–S4.

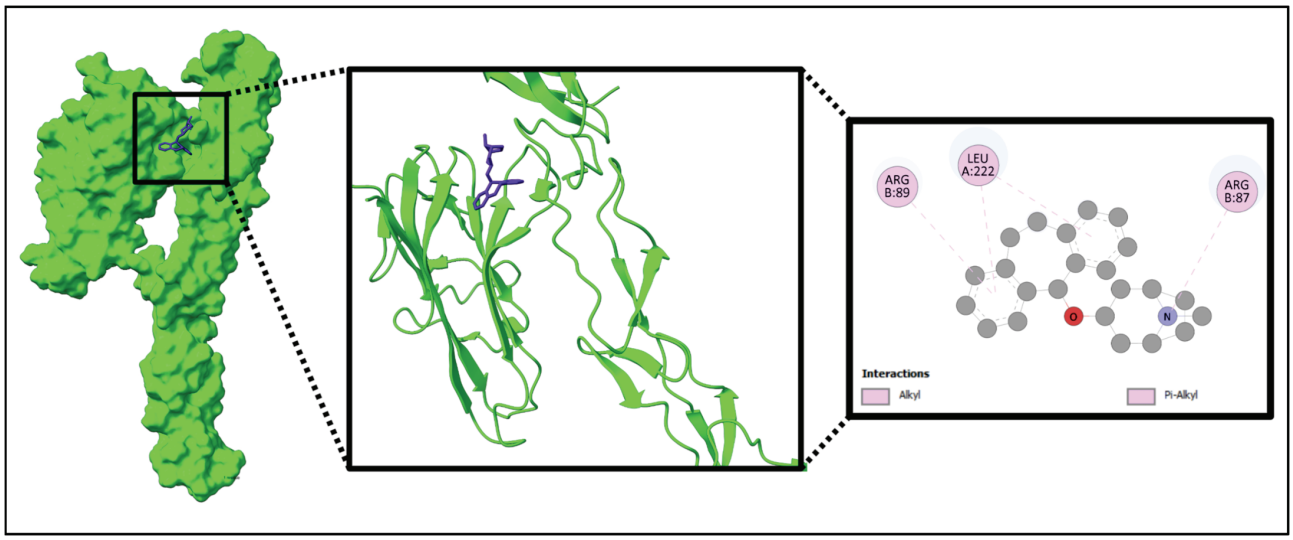


Figure 9. Molecular interaction of FASLG with the selected compound (FASLG + Deptropine). In this figure, green color indicated the protein and blue color indicated the compound. Additional data related to Figure 9 can be found in Supplementary Figures S5–S8.

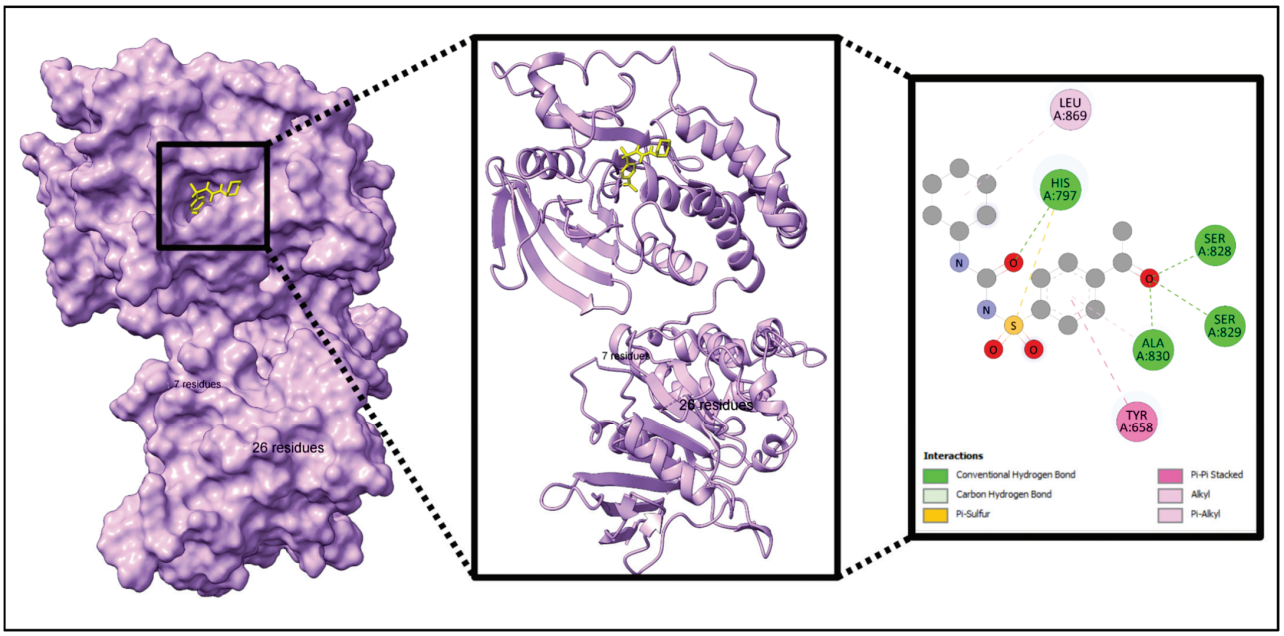


Figure 10. Molecular interaction of PTPRC with the selected compound (PTPRC + Acetohexamide). In this figure, violate color indicates the protein and yellow color indicates the compound. Additional data related to Figure 10 can be found in Supplementary Figures S9–S12.

The molecular binding interactions from PyRx reveals that Deptropine demonstrates a superior binding affinity across all proteins, particularly to IL10 (−8.3 kcal/mol), significantly outperforming 3-azido-3-deoxythymidine, which consistently exhibited the weakest binding (−5.3 kcal/mol) (Table 3). Acetohexamide and Methotrexate display comparable binding profiles with a strong affinity to IL10 and PTPRC, though Methotrexate binds slightly more strongly to PTPRC (−7.3 vs. −7.7 kcal/mol) (Table 3). Retinoic Acid shows a strong binding to IL10 (−8.1 kcal/mol), with a more gradual decrease across other proteins compared to the steeper drops observed with other compounds. Among proteins, IL10 generally exhibits the strongest compound interactions (except with Acetohexamide where

PTPRC binding is stronger), followed by PTPRC, then FASLG with intermediate affinity, while FCGR3A demonstrates a relatively weaker binding (Table 3). The overall binding patterns suggest that aromatic compounds (Deptropine, Methotrexate) and larger molecules bind more strongly, while IL10's consistent strong binding across compounds indicates it may possess a more accessible binding pocket or favorable electrostatic environment (Figure 12, Table 3).

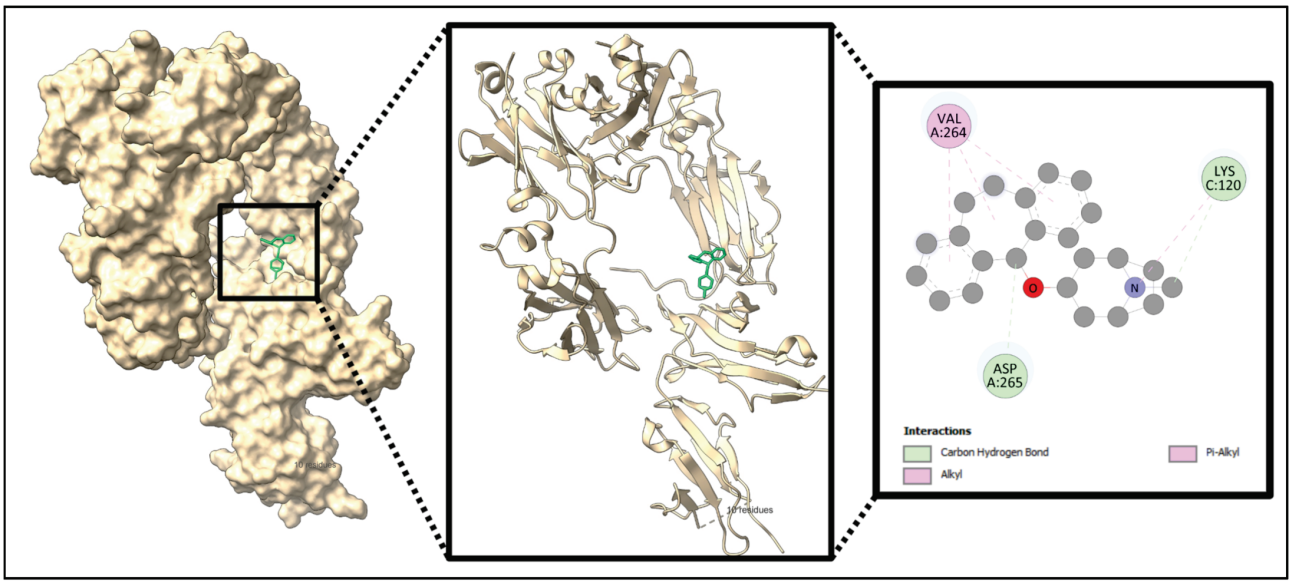


Figure 11. Molecular interaction of FCGR3A with the selected compound (FCGR3A + Deptropine). In this figure, beige color indicates the protein and green color indicates the compound. Additional data related to Figure 10 can be found in Supplementary Figures S13–S16.

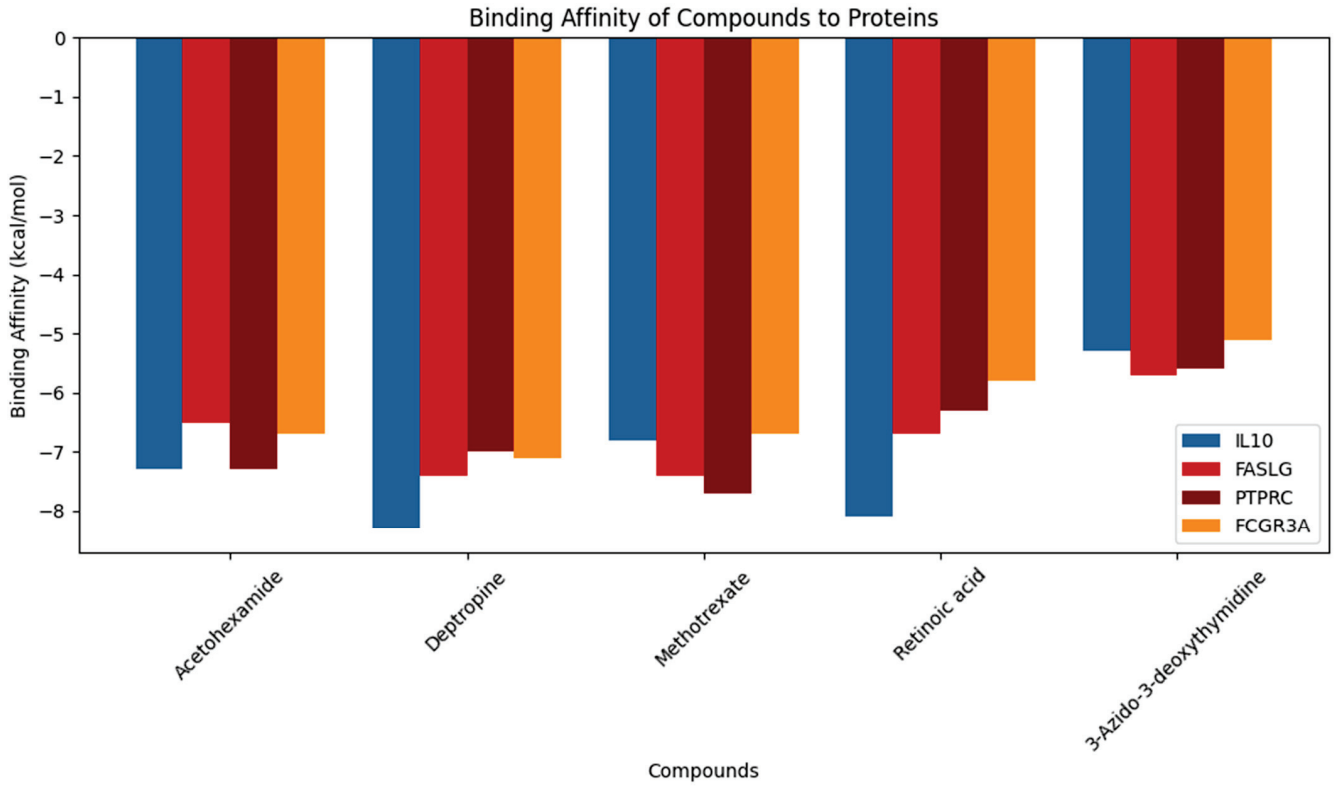


Figure 12. Binding affinity comparison.

In this paper, IL10 with Acetohexamide, Deptropine, and Retinoic Acid; FASLG with Deptropine and Methotrexate; PTPRC with Acetohexamide and Methotrexate; and FCGR3A with Deptropine exhibited the strongest binding affinities according to the results from PyRx (Table 3).

The molecular docking analysis comparing the selected compounds with known inhibitors revealed promising findings based on PyRx (Tables 3 and 4, Figures 8–11 and 13). Deptropine exhibited a stronger binding affinity (Table 3, Table 4) than the reference inhibitor JTE-607 (Table 3) for IL10 (−8.3 kcal/mol vs. −6.7 kcal/mol), suggesting a potentially enhanced interaction. Similarly, Methotrexate demonstrated better binding with PTPRC compared to its known inhibitor 2-(4-Acetylanilino)-3-chloronaphthoquinone (−7.7 kcal/mol vs. −7.3 kcal/mol). Additionally, Deptropine showed a higher binding affinity to FCGR3A than BI-1206 (−7.1 kcal/mol vs. −6.5 kcal/mol). However, for FASLG, the known inhibitor ONL1204 exhibited a stronger affinity (−7.7 kcal/mol) than the best-performing selected compounds Methotrexate and Deptropine (−7.4 kcal/mol each) (Tables 3 and 4). Overall, these results suggest that Deptropine and Methotrexate are promising candidates for IL10, PTPRC, and FCGR3A.

Table 4. Binding affinity of the target proteins with their known inhibitors.

Compound Name and CID	Protein Name	Interaction Between Protein and Compound	Binding Affinity
2-(4-Acetylanilino)-3-chloronaphthoquinone; CID: 781109	PTPRC	PTPRC + 2-(4-Acetylanilino)-3-chloronaphthoquinone	−7.3 kcal/mol
JTE-607; CID: 9938544	IL10	IL10 + JTE-607	−6.7 kcal/mol
BI-1206; CID: 170872117	FCGR3A	FCGR3A + BI-1206	−6.5 kcal/mol
ONL1204; CID: 122677428	FASLG	FASLG + ONL1204	−7.7 kcal/mol

The binding affinity results from the two docking tools (PyRx and SwissDock) reveal both alignments and notable variations across different compounds and proteins (Table 3). While the general trends remain consistent, specific differences highlight methodological nuances between the tools.

For Acetohexamide (CID: 1989), both tools predicted similar binding affinities for IL10 (−7.3 kcal/mol), indicating consistency in the docking results. However, SwissDock predicted a stronger binding for FASLG (−6.8 kcal/mol vs. −6.5 kcal/mol) and FCGR3A (−7.0 kcal/mol vs. −6.7 kcal/mol), with the difference ranging from 0.3 to 0.5 kcal/mol. Conversely, the first tool (PyRx) predicted a stronger binding for PTPRC (−7.3 kcal/mol vs. −7.1 kcal/mol), though the variation remains minor at 0.2 kcal/mol (Table 3).

Deptropine (CID: 203911) shows notable differences, where the first tool consistently predicted stronger binding affinities across all proteins. The largest discrepancy occurred for IL10 (−8.3 kcal/mol vs. −7.0 kcal/mol), showing a difference of 1.3 kcal/mol, while FASLG (−7.4 kcal/mol vs. −6.9 kcal/mol), PTPRC (−7.0 kcal/mol vs. −6.4 kcal/mol), and FCGR3A (−7.1 kcal/mol vs. −6.8 kcal/mol) exhibited differences ranging from 0.3 to 0.6 kcal/mol. This variation suggests that SwissDock may apply different scoring functions or ligand flexibility considerations (Table 3).

Methotrexate (CID: 126941) presented mixed results, with SwissDock predicting a stronger binding for IL10 (−7.8 kcal/mol vs. −6.8 kcal/mol) and FCGR3A (−8.1 kcal/mol vs. −6.7 kcal/mol), showing substantial differences of 1.0 and 1.4 kcal/mol, respectively. Meanwhile, the binding affinities for FASLG remained identical at −7.4 kcal/mol across both tools. The first tool predicted a stronger binding for PTPRC (−7.7 kcal/mol vs. −7.4 kcal/mol), but the difference is smaller at 0.3 kcal/mol (Table 3).

Retinoic Acid (CID: 449171) generally shows a stronger binding in SwissDock, with notable differences for FASLG (−7.1 kcal/mol vs. −6.7 kcal/mol), PTPRC (−7.1 kcal/mol vs. −6.3 kcal/mol), and FCGR3A (−6.9 kcal/mol vs. −5.8 kcal/mol). The largest deviation occurred in FCGR3A at 1.1 kcal/mol, suggesting SwissDock models binding interactions differently, possibly considering additional hydrophobic or electrostatic contributions (Table 3).

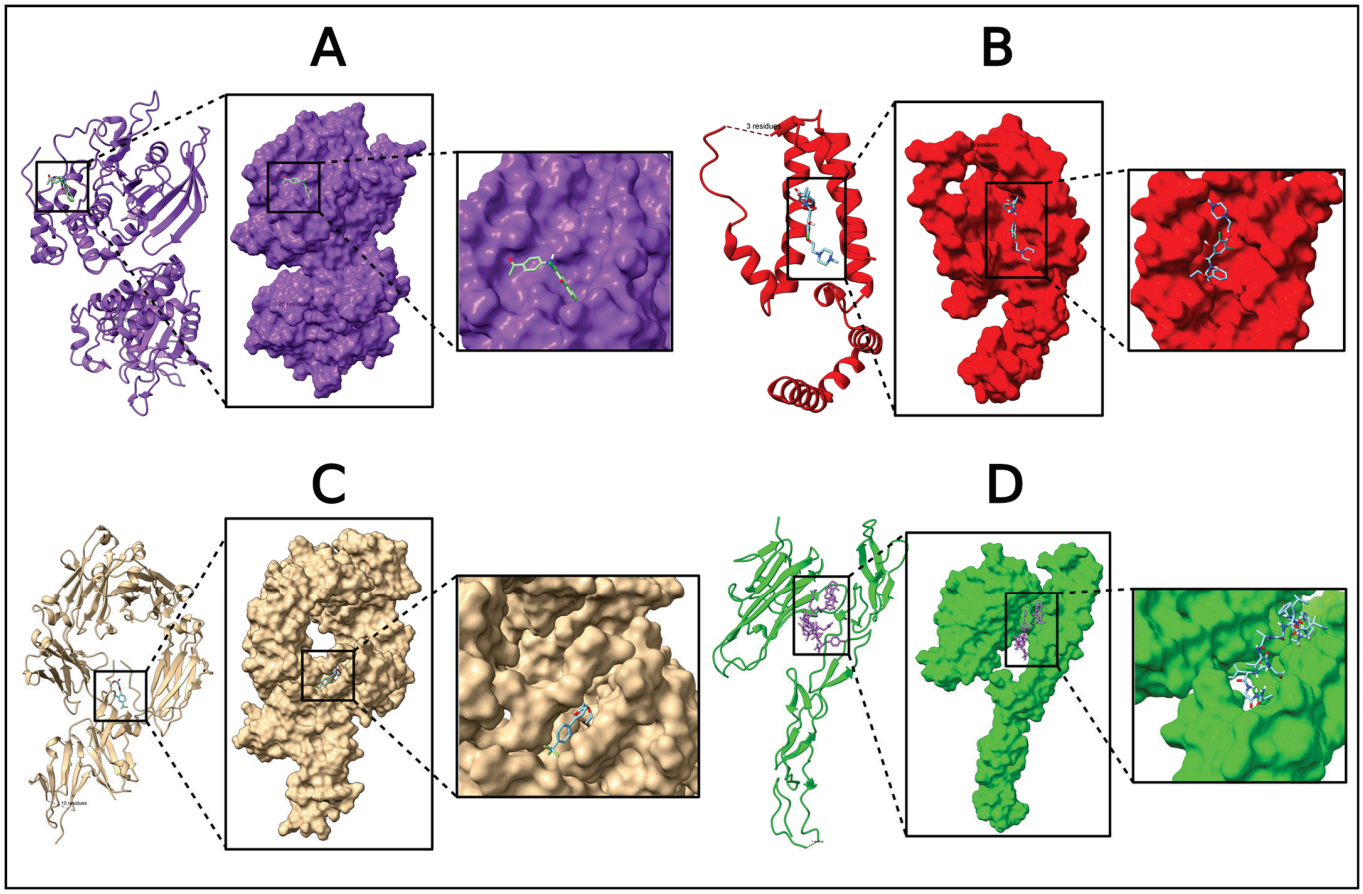


Figure 13. Molecular interactions of target proteins with their know inhibitors. (**A**) PTPRC (violet color) + 2-(4-Acetylanilino)-3-chloronaphthoquinone (sky blue with red), (**B**) IL10 (red color) + JTE-607 (sky blue with red), (**C**) FCGR3A (beige color) + BI-1206 (sky blue with red), and (**D**) FASLG (Green color) + ONL1204 (sky blue with red).

Finally, 3-Azido-3-deoxythymidine (CID: 5726) demonstrated the most substantial deviation across all proteins, where SwissDock consistently predicted a stronger binding affinity. The differences ranged from 1.1 to 1.7 kcal/mol, with SwissDock showing a consistently stronger binding.

3. Discussion

This study employed a systematic computational approach to identify host-targeted therapeutics for Oropouche virus (OROV) by integrating network pharmacology, molecular docking, and drug repurposing. The findings suggest that targeting host factors could be an effective antiviral strategy, reducing the risk of resistance that often arises with direct-acting antivirals. Our study identified IL10, FASLG, PTPRC, and FCGR3A as key immune-related host proteins involved in OROV infection. Small-molecule inhibitors Acetohexamide, Deptropine, Methotrexate, Retinoic Acid,

and 3-Azido-3-deoxythymidine exhibited strong binding affinities with these targets, highlighting their potential for repurposing as antiviral agents.

Our findings align with recent studies on host-directed antiviral strategies for arboviruses, including dengue virus (DENV), Zika virus (ZIKV), and Chikungunya virus (CHIKV) [17,18]. Prior research has demonstrated that targeting immune regulators such as IL10 can modulate viral infection outcomes. A study on DENV suggested that IL10 dysregulation aids viral immune evasion, and blocking IL10 pathways enhances viral clearance [19,20]. Similarly, PTPRC (CD45) has been implicated in viral pathogenesis, particularly in HIV and Epstein–Barr virus (EBV), where T-cell activation plays a crucial role in disease progression [21]. The identification of these immune checkpoints in OROV infection suggests that the virus may employ immune modulation mechanisms similar to other viral pathogens, reinforcing the viability of these proteins as therapeutic targets [22,23].

Our protein–protein interaction (PPI) analysis revealed the clustering of immune-related pathways, including Fc-gamma receptor signaling, cytokine regulation, and T-cell receptor signaling. These pathways have also been reported as crucial intervention points in other arboviral infections, where Fc-gamma receptor signaling modulates viral antibody-dependent enhancement (ADE) [24]. Studies have shown that flaviviruses like ZIKV and DENV exploit Fc-mediated pathways to enhance viral infection, emphasizing the potential role of FCGR3A in viral pathogenesis [25,26]. Our analysis extends these findings to OROV, identifying FCGR3A (CD16a) as a potential antiviral target, aligning with prior work on immune modulation in viral infections [27].

For a proper molecular docking, the active site of each protein was identified using Discovery Studio [28], a widely recognized computational tool known for its robust structural analysis capabilities [29–31]. Unlike online tools that merely confirm the presence of an active site, Discovery Studio allows a more refined approach, enabling precise characterization based on ligand interactions, binding pocket geometry, and energetic considerations [32,33]. This level of structural validation strengthens the reliability of the identified active sites, ensuring consistency with established biochemical principles.

The molecular docking analysis demonstrated that Deptropine and Acetohexamide exhibited the strongest binding affinities across multiple targets, suggesting their broad-spectrum antiviral potential. Similar computational approaches have identified repurposable drugs against other viruses, where docking studies highlighted the importance of drug binding to host immune regulators. Notably, our binding affinities for Deptropine (-8.3 kcal/mol with IL10) and Acetohexamide (-7.3 kcal/mol with PTPRC) are comparable to findings from previous antiviral docking studies, where binding affinities below -7 kcal/mol were associated with strong target engagement [16]. These results suggest that the identified compounds could serve as promising candidates for further investigation in antiviral drug discovery, notably, Retinoic Acid, known for its role in immune modulation [34]. Among them, Deptropine and Acetohexamide exhibited strong binding affinities to multiple targets, suggesting broad-spectrum antiviral potential. They have demonstrated antiviral properties against other RNA viruses [35]. Methotrexate, primarily used as an immunosuppressant [36], may influence viral replication through the indirect regulation of host immune responses [37].

The molecular docking results provide insights into the binding potential of the selected compounds compared to known inhibitors. Deptropine exhibited a stronger binding affinity than JTE-607 for IL10 (-8.3 kcal/mol vs. -6.7 kcal/mol), indicating a potentially enhanced interaction. Similarly, Methotrexate demonstrated improved binding to PTPRC compared to its known inhibitor 2-(4-Acetylanilino)-3-chloronaphthoquinone

(−7.7 kcal/mol vs. −7.3 kcal/mol). Additionally, Deptropine showed a better binding affinity for FCGR3A than BI-1206 (−7.1 kcal/mol vs. −6.5 kcal/mol), suggesting that it may be a promising candidate for targeting this protein.

The binding affinities predicted by PyRx and SwissDock exhibited both concordance and divergence, underscoring the importance of multi-tool validation in computational drug discovery. While there was consistency with general trends (e.g., Methotrexate's strong binding to PTPRC), notable discrepancies emerged, such as Deptropine's IL10 affinity (−8.3 kcal/mol in PyRx vs. −7.0 kcal/mol in SwissDock). These variations likely stem from methodological differences: PyRx's AutoDock Vina employs a gradient-optimized scoring function, whereas SwissDock's EADock DSS incorporates solvation effects and broader conformational sampling. The stronger FCGR3A binding predicted by SwissDock for Methotrexate (−8.1 kcal/mol vs. −6.7 kcal/mol in PyRx) suggests its algorithm may better capture hydrophobic interactions, a critical factor for this receptor. Based on integrated docking results from PyRx and SwissDock, Deptropine emerges as the most promising candidate against OROV. It showed the strongest binding to IL10 (−8.3 kcal/mol in PyRx, −7.0 kcal/mol in SwissDock), outperforming the known inhibitor JTE-607 (−6.7 kcal/mol). Deptropine also exhibited robust affinity for FCGR3A (−7.1 kcal/mol in PyRx, −6.8 kcal/mol in SwissDock), surpassing BI-1206 (−6.5 kcal/mol). While Methotrexate bound well to PTPRC (−7.7 kcal/mol), its weaker interaction with IL10 (−6.8 kcal/mol) limits its broad-spectrum potential. Acetohexamide showed moderate binding (−6.5 to −7.3 kcal/mol), while Retinoic Acid and 3-Azido-3-deoxythymidine had consistently weaker affinities (> −7 kcal/mol). Deptropine's multi-target inhibition of IL10, FCGR3A, and FASLG suggests it could disrupt immune evasion, viral entry, and apoptosis pathways simultaneously.

IL10, a key target in our study, plays a dual role in viral infections. While it modulates inflammatory responses to prevent excessive tissue damage, it can also suppress antiviral immunity, enabling viral persistence [20,27]. The strong docking interactions of Acetohexamide, Deptropine, and Retinoic Acid with IL10 suggest that these drugs could modulate IL10 signaling to enhance antiviral immunity, similar to IL10-targeting strategies explored in HCV and DENV research [19]. Additionally, FASLG (Fas Ligand) is crucial for apoptosis regulation, and many viruses, including DENV and HIV, manipulate apoptotic pathways to avoid immune clearance [38]. Our docking results show that Methotrexate and Deptropine bind strongly to FASLG, suggesting a potential mechanism to restore apoptosis and eliminate infected cells. Furthermore, PTPRC (CD45) is essential for T-cell receptor signaling and immune activation, and its involvement in OROV infection suggests a novel immune evasion strategy by the virus [21]. Our findings align with previous studies on HIV and EBV, where CD45-targeting drugs altered disease progression [21]. Finally, FCGR3A plays an essential role in antibody-dependent cellular cytotoxicity, a key mechanism in viral clearance [27,39]. Though its role in OROV infection is not well-documented, its involvement in other viral infections suggests that targeting this pathway may enhance antiviral immunity [27].

While literature-reported binding sites offer valuable reference points, protein–ligand interactions are inherently dynamic and context-dependent. The observed binding of the proposed compounds to alternative residues suggests potential mechanistic diversity rather than deviation from established findings [40–42]. This variability does not diminish the relevance of the compounds; rather, it broadens the scope for further experimental validation [43–45]. Such insights can pave the way for discovering novel therapeutic interactions, emphasizing the importance of considering both canonical and alternative binding modes in drug discovery [46–48].

Despite the promising findings, our study has several limitations. First, while computational approaches provide valuable insights, they lack experimental validation. Our aim of this study is to systematically integrate host–virus interaction networks with drug prediction with docking validation. Similar approaches have been employed for other arboviruses, such as dengue and chikungunya, where host-targeting strategies have yielded promising antiviral candidates [17,49]. Future studies should include in vitro antiviral assays to confirm the efficacy of identified drug candidates. Additionally, molecular dynamics (MD) simulations should be performed to assess the stability of drug–protein interactions beyond docking. Second, the study primarily focuses on network pharmacology and docking but lacks a full multi-omics approach. Integrating transcriptomics and proteomics data could enhance target validation. Finally, future work should involve in vitro studies in human cell lines and animal models to evaluate the antiviral efficacy of Acetohexamide, Deptropine, and Methotrexate.

In conclusion, this study presents a computational framework for identifying host-targeted antiviral candidates against Oropouche virus. By targeting immune-modulating proteins such as IL10, FASLG, PTPRC, and FCGR3A, we propose a novel antiviral strategy that aligns with recent host-directed antiviral research. However, experimental validation remains critical, and future studies should explore in vitro/in vivo assays to confirm the therapeutic potential of the identified compounds. Integrating multi-omics data and molecular dynamics simulations will further strengthen the translational value of these findings in antiviral drug discovery.

4. Materials and Methods

4.1. Target Identification for Oropouche Virus

The targets associated with the Oropouche virus were collected from the OMIM (Online Mendelian Inheritance in Man) database (https://www.omim.org/, accessed on 6 March 2025) and GeneCards database (https://www.genecards.org/, accessed on 6 March 2025) [50,51]. The keywords "Oropouche virus" and "Oropouche fever" were set as search options in those databases. Related target symbols corresponding to the Oropouche virus were accumulated.

4.2. Identified Drugs

Within the Enrichr platform (https://maayanlab.cloud/Enrichr/, accessed on 6 March 2025), the DSigDB database (https://dsigdb.tanlab.org/DSigDBv1.0/, accessed on 6 March 2025) was applied to detect the drug molecules that are connected to the target genes for the virus [52]. The filtering criterion was adjusted in $p < 0.05$. From that analysis, the top 10 compounds were identified as the potential applicants for inhibiting the Oropouche virus [53].

4.3. Selected Compounds

We followed Lipinski's rule parameters for drug criteria such as molecular weight (<500 g/mol), hydrogen bond acceptors (<10), hydrogen bond donors ($\leq$5), topological polar surface area (<150), and Moriguchi octanol–water partition coefficient ($\leq$4.15). To calculate those criteria, the SwissADME tool was used [54].

4.4. Target Gene Identification for Selected Compounds

Different databases were utilized to collect the target genes for each selected compound. The name of the compounds was used as a keyword for searching. For obtaining the target genes, we used CTD (Comparative Toxicogenomics Database) (https://ctdbase.org/, accessed on 6 March 2025), Swiss Target Prediction tool (http://www.swisstargetprediction.ch/, accessed

on 6 March 2025), Drug Bank (https://go.drugbank.com/, accessed on 25 February 2025), and SymMap (http://www.symmap.org/, accessed on 6 March 2025) [55,56]. Then, the UniPort database was utilized to cross-reference the detected targets for *Homo sapiens* [57].

4.5. Intersection of Targets

To determine the common targets for the virus and compounds, the Venny 2.0 tool (https://bioinfogp.cnb.csic.es/tools/venny/index2.0.2.html, accessed on 6 March 2025) was used. For further analysis, those common targets were applied, and these are the connective targets between considered compounds as drugs and the virus [58].

4.6. Protein–Protein Interaction Network

The String database (https://string-db.org/, accessed on 6 March 2025) was applied to create the PPI network among those intersected targets and specifying "Homo sapiens" as the species [59]. The interaction file then used to visualize them in the Cytoscape 3.10.3 software [59]. After that, the top twenty common targets were identified by the CytoHubba plugin of Cytoscape [60].

4.7. Analysis by MCODE

For creating functional modules and clustering similar proteins, we used MetaScape (https://metascape.org/, accessed on 6 March 2025) with molecular complex detection (MCODE) [61]. With the aid of a custom MCODE tool, a clustering analysis was performed for common targets where a minimum overlap of 3, a *p*-value cutoff of 0.05, and a minimum enrichment of 1.5 were used. Using MCODE analysis, protein relationship complexes were identified and organized into gene clusters, where each target can be associated with particular gene functions. Moreover, other regulatory transcription factors that modulated the specified targets were also predicted to inform the user of possible target interactions.

4.8. Gene Ontology Analysis

By using the DAVID online platform (https://davidbioinformatics.nih.gov/, accessed on 6 March 2025), Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomics (KEGG) pathways were analyzed for the intersection targets of the virus and compounds where biological processes (BPs), cellular components (CCs), molecular functions (MFs), and KEGG pathways were included [62]. In this analysis, the top ten GO entries and seventy KEGG pathways were identified based on gene count value. The bioinformatics online platforms (https://www.bioinformatics.com.cn/en, accessed on 6 March 2025) were used for visualization. Furthermore, the overlapping genes underwent KEGG mapping analysis by focusing on Homo sapiens. Metabolism, environmental information processing, cellular processes, systems of the organism, and human diseases are the pathways that were included in that analysis.

4.9. Compound–Pathway–Target Interaction Network

From the Metascape analysis, the complex biological pathways were identified by using MCODE cluster calculations. Those complex biological pathways were used to construct a compound–pathway–target interaction network by using the intersecting genes and compounds.

4.10. Identified Key Targets

The identified final targets were used to compare them to find out key targets. For further analysis, those final targets were used in the STRING (https://string-db.org/, accessed on 6 March 2025) online platform to create a network among those genes. The required score

was set to 0.700 to obtain a high confidence in network construction, and the species was set as Homo sapiens. After that, the Cytoscape 3.10.3 software along with the Cytohubba plugin were utilized to visualize and analyze the network. To find out the key targets, five different topological properties (Maximal Clique Centrality (MCC), Closeness, BottleNeck, Edge Percolated Component (EPC), Degree value, and Radiality) were set. The UpSetR Shinny App (https://gehlenborglab.shinyapps.io/upsetr/, accessed on 6 March 2025) was utilized to see the number of common genes for those different topological factors [63]. We also applied molbiotools (https://molbiotools.com/listcompare.php, accessed on 6 March 2025) for comparing all six properties.

4.11. Use of Positive Control Inhibitors

4.11.1. Using Positive Controls

To ensure the reliability and accuracy of our molecular docking simulations, we included positive control inhibitors for each target protein. Positive controls are known compounds that are expected to bind to the target proteins and provide a benchmark for validating the docking methodology [64]. By docking these inhibitors alongside the target proteins, it is helpful to assess the docking protocol's ability to correctly predict protein–ligand interactions, even in the absence of literature-reported binding affinities [65,66].

4.11.2. Selection of Positive Control Inhibitors

For each of the target proteins (FASLG, FCGR3A, IL10, and PTPRC), the following positive control inhibitors were selected based on their known interactions or theoretical binding to these proteins:

- FASLG (Fas Ligand): One promising inhibitor for FASLG is ONL1204, a small peptide antagonist of the Fas receptor. It has demonstrated neuroprotective effects in models of glaucoma, reducing neuroinflammation and preventing axon degeneration [67].
- FCGR3A (Fc gamma receptor IIIa): The inhibitor BI-1206 was selected for FCGR3A based on its a monoclonal antibody that blocks Fc gamma receptor IIIa to enhance immune responses in cancer therapy [68].
- IL10 (Interleukin 10): A well-characterized positive control inhibitor for IL10 (Interleukin-10) is JTE-607, which selectively inhibits inflammatory cytokine synthesis, including IL10 [69].
- PTPRC (Protein Tyrosine Phosphatase Receptor Type C): A well-characterized positive control inhibitor for PTPRC (CD45) is CD45 Inhibitor VI, also known as 2-(4-Acetylanilino)-3-chloronaphthoquinone. It is a chloronaphthoquinone compound that potently inhibits PTPRC tyrosine phosphatase activity in a non-substrate-competitive and irreversible manner [70].

4.12. Molecular Docking

4.12.1. Protein and Compound Preparation

First of all, each protein was added into the Discovery Studio software for removing heteroatoms such as water and other ligands. After that, those proteins were inserted into the Chimera X 1.19 software for further preparation [71]. In this process, hydrogens were added, charges and missing atoms were added, and the prepared proteins were saved as a PDB file for docking [72]. For docking, we used the PyRx software, which has the accessibility of Autodock Vina. Each protein was added there and converted into a macromolecule, and the next grid box for each protein was adjusted for docking.

The saved SDF files of compounds were directly added into PyRx for docking since PyRx has access to Open Babel [73,74]. The charge of each compound was minimized, and

those minimized charged compounds were converted into a PDBQT file to continue the docking process [75].

4.12.2. Identification of Binding Site and Grid Box for Protein

The binding sites/active sites for each target were identified by using Discovery Studio's predictions and the grid boxes optimized around the binding areas suggested by the software [76–78]. Active sites help to adjust the grid box according to protein. Discovery Studio is a useful software and helped to find out the active sites quickly [79]. While the binding affinity is an important parameter, ligand interactions also depend on chemical properties, steric interactions, and protein flexibility, which influence binding to specific residues [80–82]. For IL10 (PDB ID: 2H24), LEU26, PHE30, MET68, ILE69, TYR72, LEU73, VAL76, MET77, DER93, LEU94, GLY95, and LEU98 were the amino acids for the active site. LEU85, ARG89, CYS132, PRO133, ALA136, SER159, SER161, ASN215, TYR218, GLN220, and LEU222 were the identified amino acids of the active site for FASLG (PDB ID: 4MSV). Furthermore, PRO244, PRO245, LYS246, GLU258, THR260, ASP265, THR299, ARG301, GLN295, HIS44, ASP64, SER66, and ARG155 were the identified amino acids of the active site for FCGR3A (PDB ID: 3AY4). Moreover, ARG637, TYR658, VAL659, ARG736, LYS736, HIS797, SER828, SER829, ALA830, GLY831, VAL 832, GLY833, and ARG834 were identified for PTPRC (PDB ID: 1YGR).

The grid box was set into the center: X = 15.7665, Y = 26.4005, Z = 13.5283 and dimensions: X = 32.4150, Y = 34.1901, Z = 27.6981 for IL10 (PDB ID: 2H24). The box was fixed into the center: X = −18.0398, Y = −7.5283, Z = −52.8173 and dimensions: X = 37.2960, Y = 43.1758, Z = 30.4286 for FASLG (PDB ID: 4MSV). The box was selected into the center: X = 18.0506, Y = −8.2066, Z = 55.3804 and dimensions: X = 39.4536, Y = 42.7338, Z = 36.2982 for PTPRC (PDB ID: 1YGR). Finally, for FCGR3A (PDB ID: 3AY4), the grid box was fixed into the center: X = 9.2275, Y = 22.3744, Z = 110.7291 and dimensions: X = 26.8820, Y = 25.0730, Z = 38.1790.

4.13. Alternative Docking

To further validate the docking results, molecular docking was performed using SwissDock, a web-based docking service employing the EADock DSS algorithm [83]. This alternative approach was used to cross-check the binding affinities of selected compounds with the target proteins, ensuring consistency with previous docking findings.

For this online tool, the targets were used directly in the PDB format, as it has an option for preparing proteins [84]. The smiles of the compound were inserted and prepared by the tool [84]. The grid box size and position for each protein were set according to the active site of each protein, which was identified by Discovery Studio in the previous method.

5. Conclusions

This study employed a systems biology-driven approach integrating network pharmacology, molecular docking, and drug repurposing to identify host-targeted therapeutics against the Oropouche virus (OROV). By prioritizing key host proteins—IL10, FASLG, PTPRC, and FCGR3A—we uncovered their critical roles in immune modulation and viral pathogenesis. Deptropine and Acetohexamide emerged as promising candidates, demonstrating strong binding affinities across multiple targets, suggesting their potential as broad-spectrum antiviral agents.

Our findings highlight the viability of host-directed antiviral strategies to circumvent viral resistance, a major limitation of direct-acting antivirals. The computational validation of these compounds, supported by multi-docking analyses, provides a robust foundation

for further experimental studies. However, in vitro and in vivo assays are essential to confirm their efficacy and safety.

This research bridges a critical gap in OROV therapeutics, offering a cost-effective, rapid-response framework for neglected tropical diseases. Future work should incorporate multi-omics data, molecular dynamics simulations, and clinical trials to advance these findings toward real-world applications. By leveraging computational drug discovery, this study contributes to pandemic preparedness and underscores the importance of host-targeted interventions in combating emerging arboviral threats.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ph18050613/s1. Table S1: Target genes common in both virus and drugs. Table S2: List of biological processes. Table S3: List of cellular components. Table S4: List of molecular functions. Table S5: List of keg pathway. Table S6: Targets with different topological properties. Figure S1: Molecular interaction of IL10 with the selected compound (IL10 + Acetohexamide). Figure S2: Molecular interaction of IL10 with the selected compound (IL10 + Methotrexate). Figure S3: Molecular interaction of IL10 with the selected compound (IL10 + Retinoic acid). Figure S4: Molecular interaction of IL10 with the selected compound (IL10 + 3-Azido-3-deoxythymidine). Figure S5: Molecular interaction of FASLG with the selected compounds (FASLG + Acetohexamide). Figure S6: Molecular interaction of FASLG with the selected compounds (FASLG + Methotrexate). Figure S7: Molecular interaction of FASLG with the selected compounds (FASLG + Retinoic acid). Figure S8: Molecular interaction of FASLG with the selected compounds (FASLG + 3-Azido-3-deoxythymidine). Figure S9: Molecular interaction of PTPRC with the selected compound (PTPRC + Deptropine). Figure S10: Molecular interaction of PTPRC with the selected compound (PTPRC + Methotrexate). Figure S11: Molecular interaction of PTPRC with the selected compound (PTPRC + Retinoic acid). Figure S12: Molecular interaction of PTPRC with the selected compound (PTPRC + 3-Azido-3-deoxythymidine). Figure S13: Molecular interaction of FCGR3A with the selected compound (FCGR3A + Acetohexamide). Figure S14: Molecular interaction of FCGR3A with the selected compound (FCGR3A + Methotrexate). Figure S15: Molecular interaction of FCGR3A with the selected compound (FCGR3A + Retinoic acid). Figure S16: Molecular interaction of FCGR3A with the selected compound (FCGR3A + 3-Azido-3-deoxythymidine).

Author Contributions: Conceptualization, P.D.S.; methodology, P.D.S.; software, P.D.S., A.A.N. and H.S.; validation, P.D.S., A.A.N. and A.M.A.; formal analysis, P.D.S. and A.A.N.; investigation, P.D.S.; resources, A.A.N. and P.D.S.; data curation, P.D.S.; writing—original draft, A.A.N., P.D.S. and H.S.; writing—review and editing, P.D.S., E.M.A., A.M.A. and T.H.T.; visualization, A.M.A. and E.M.A.; supervision, E.M.A.; project administration, P.D.S. and A.A.N.; funding acquisition, A.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2501).

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 no conflicts of interest.

References

1. Zhang, Y.; Liu, X.; Wu, Z.; Feng, S.; Lu, K.; Zhu, W.; Sun, H.; Niu, G. Oropouche virus: A neglected global arboviral threat. *Virus Res.* **2024**, *341*, 199318. [CrossRef] [PubMed]
2. Travassos Da Rosa, J.F.; De Souza, W.M.; De Paula Pinheiro, F.; Figueiredo, M.L.; Cardoso, J.F.; Acrani, G.O.; Teixeira Nunes, M.R. Oropouche Virus: Clinical, Epidemiological, and Molecular Aspects of a Neglected Orthobunyavirus. *Am. J. Trop. Med. Hyg.* **2017**, *96*, 1019–1030. [CrossRef] [PubMed]

3. de Armas Fernández, J.R.; Peña García, C.E.; Acosta Herrera, B.; Betancourt Plaza, I.; Gutiérrez de la Cruz, Y.; Resik Aguirre, S.; Kourí Cardellá, V.; Guzmán Tirado, M.G. Report of an unusual association of Oropouche Fever with Guillain-Barré syndrome in Cuba, 2024. *Eur. J. Clin. Microbiol. Infect. Dis.* **2025**, *43*, 2233–2237. [CrossRef]
4. Carpenter, S.; Groschup, M.H.; Garros, C.; Felippe-Bauer, M.L.; Purse, B.V. Culicoides biting midges, arboviruses and public health in Europe. *Antivir. Res.* **2013**, *100*, 102–113. [CrossRef]
5. Gräf, T.; Delatorre, E.; do Nascimento Ferreira, C.; Rossi, A.; Santos, H.G.G.; Pizzato, B.R.; Nascimento, V.; Souza, V.; de Lima, G.B.; Dezordi, F.Z.; et al. Expansion of Oropouche virus in non-endemic Brazilian regions: Analysis of genomic characterisation and ecological drivers. *Lancet Infect. Dis.* **2024**, *25*, 379–389. [CrossRef] [PubMed]
6. Tortosa, F.; Castillo, G.G.; Izcovich, A.; Luz, K.; dos Santos, T.; Gonzalez-Escobar, G.; Ragusa, M.A.; Gresh, L.; Mendez-Rico, J.A.; Reveiz, L. Revisión sistemática viva de las manifestaciones clínicas de la fiebre de Oropouche: Claves para diferenciarla del dengue y otras arbovirosis. *Rev. Panam. Salud Publica* **2024**, *48*, e136. [CrossRef]
7. Elliott, R.M. Orthobunyaviruses: Recent genetic and structural insights. *Nat. Rev. Microbiol.* **2014**, *12*, 673–685. [CrossRef]
8. Bonifay, T.; Le Turnier, P.; Epelboin, Y.; Carvalho, L.; De Thoisy, B.; Djossou, F.; Duchemin, J.B.; Dussart, P.; Enfissi, A.; Lavergne, A.; et al. Review on Main Arboviruses Circulating on French Guiana, An Ultra-Peripheric European Region in South America. *Viruses* **2023**, *15*, 1268. [CrossRef]
9. Akingbola, A.; Adegbesan, A.; Ojo, O.; Ezendu, A.; Shekoni, M. The rising concern of Oropouche fever: A call for enhanced surveillance and research in emerging arboviral threats. *Infect. Dis.* **2024**, *56*, 1015–1019. [CrossRef]
10. Jiao, Y.; Shi, C.; Sun, Y. The Use of Xuanbai Chengqi Decoction on Monkeypox Disease through the Estrone-Target AR Interaction. *Front Microbiol* **2023**, *14*, 1234817. [CrossRef]
11. Peinado, R.d.S.; Eberle, R.J.; Arni, R.K.; Coronado, M.A. A Review of Omics Studies on Arboviruses: Alphavirus, Orthobunyavirus and Phlebovirus. *Viruses* **2022**, *14*, 2194. [CrossRef] [PubMed]
12. Moreira, F.R.R.; Dutra, J.V.R.; de Carvalho, A.H.B.; Reis, C.R.; Rios, J.S.H.; de Oliveira Ribeiro, M.; Arruda, M.B.; Alvarez, P.; Souza, R.P.; Voloch, C.; et al. Oropouche virus genomic surveillance in Brazil. *Lancet Infect. Dis.* **2024**, *24*, e664–e666. [CrossRef] [PubMed]
13. Rocha, R.F.; Del Sarto, J.L.; Marques, R.E.; Costa, V.V.; Teixeira, M.M. Host target-based approaches against arboviral diseases. *Biol. Chem.* **2018**, *399*, 203–217. [CrossRef]
14. Fang, J.; Wu, Q.; Ye, F.; Cai, C.; Xu, L.; Gu, Y.; Wang, Q.; Liu, A.L.; Tan, W.; Du, G.H. Network-Based Identification and Experimental Validation of Drug Candidates Toward SARS-CoV-2 via Targeting Virus–Host Interactome. *Front. Genet.* **2021**, *12*, 728960. [CrossRef]
15. Krishnamurthy, N.; Grimshaw, A.A.; Axson, S.A.; Choe, S.H.; Miller, J.E. Drug repurposing: A systematic review on root causes, barriers and facilitators. *BMC Health Serv. Res.* **2022**, *22*, 970. [CrossRef]
16. Pantsar, T.; Poso, A. Binding Affinity via Docking: Fact and Fiction. *Molecules* **2018**, *23*, 1899. [CrossRef]
17. Bhutkar, M.; Singh, V.; Dhaka, P.; Tomar, S. Virus-host protein-protein interactions as molecular drug targets for arboviral infections. *Front. Virol.* **2022**, *2*, 959586. [CrossRef]
18. Huang, Y.J.S.; Higgs, S.; Vanlandingham, D.L. Arbovirus-mosquito vector-host interactions and the impact on transmission and disease pathogenesis of arboviruses. *Front. Microbiol.* **2019**, *10*, 434047. [CrossRef]
19. Wilson, E.B.; Brooks, D.G. The Role of IL-10 in Regulating Immunity to Persistent Viral Infections. In *Negative Co-Receptors and Ligands*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 39–65. [CrossRef]
20. Tsai, T.T.; Chuang, Y.J.; Lin, Y.S.; Wan, S.W.; Chen, C.L.; Lin, C.F. An emerging role for the anti-inflammatory cytokine interleukin-10 in dengue virus infection. *J. Biomed. Sci.* **2013**, *20*, 40. [CrossRef]
21. Al Barashdi, M.A.; Ali, A.; McMullin, M.F.; Mills, K. Protein tyrosine phosphatase receptor type C (PTPRC or CD45). *J. Clin. Pathol.* **2021**, *74*, 548–552. [CrossRef]
22. Toor, S.M.; Saleh, R.; Sasidharan Nair, V.; Taha, R.Z.; Elkord, E. T-cell responses and therapies against SARS-CoV-2 infection. *Immunology* **2021**, *162*, 30–43. [CrossRef]
23. Brockman, M.A.; Kwon, D.S.; Tighe, D.P.; Pavlik, D.F.; Rosato, P.C.; Sela, J.; Porichis, F.; Le Gall, S.; Waring, M.T.; Moss, K.; et al. IL-10 is up-regulated in multiple cell types during viremic HIV infection and reversibly inhibits virus-specific T cells. *Blood* **2009**, *114*, 346–356. [CrossRef] [PubMed]
24. Morens, D.M.; Fauci, A.S. Emerging Pandemic Diseases: How We Got to COVID-19. *Cell* **2020**, *182*, 1077–1092. [CrossRef] [PubMed]
25. Song, D.H.; Garcia, G.; Situ, K.; Chua, B.A.; Hong, M.L.O.; Do, E.A.; Ramirez, C.M.; Harui, A.; Arumugaswami, V.; Morizono, K. Development of a blocker of the universal phosphatidylserine- and phosphatidylethanolamine-dependent viral entry pathways. *Virology* **2021**, *560*, 17–33. [CrossRef]

26. Khanam, A.; Gutiérrez-Barbosa, H.; Lyke, K.E.; Chua, J.V. Immune-Mediated Pathogenesis in Dengue Virus Infection. *Viruses* **2022**, *14*, 2575. [CrossRef] [PubMed]
27. Mellor, J.D.; Brown, M.P.; Irving, H.R.; Zalcberg, J.R.; Dobrovic, A. A critical review of the role of Fc gamma receptor polymorphisms in the response to monoclonal antibodies in cancer. *J. Hematol. Oncol.* **2013**, *6*, 1. [CrossRef]
28. BIOVIA Discovery Studio | Dassault Systèmes n.d. Available online: https://www.3ds.com/products/biovia/discovery-studio (accessed on 6 April 2025).
29. Kaba, B.; Pinet, N.; Lelandais, G.; Sigayret, A.; Berry, A. Clustering Gene Expression Data Using Graph Separators. *In Silico Biol.* **2007**, *7*, 433–452. [CrossRef]
30. Wu, G.; Robertson, D.H.; Brooks, C.L.; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMm-based MD docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549–1562. [CrossRef]
31. Brooks, B.R.; Brooks, C.L.; Mackerell, A.D.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614. [CrossRef]
32. Yang, J.; Roy, A.; Zhang, Y. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* **2013**, *29*, 2588–2595. [CrossRef]
33. Laskowski, R.A.; Jabłońska, J.; Pravda, L.; Vařeková, R.S.; Thornton, J.M. PDBsum: Structural summaries of PDB entries. *Protein Sci.* **2018**, *27*, 129–134. [CrossRef] [PubMed]
34. Pino-Lagos, K.; Benson, M.J.; Noelle, R.J. Retinoic Acid in the Immune System. *Ann. N. Y. Acad. Sci.* **2008**, *1143*, 170–187. [CrossRef]
35. Morita, T.; Miyakawa, K.; Jeremiah, S.S.; Yamaoka, Y.; Sada, M.; Kuniyoshi, T.; Yang, J.; Kimura, H.; Ryo, A. All-trans retinoic acid exhibits antiviral effect against SARS-CoV-2 by inhibiting 3CLpro activity. *Viruses* **2021**, *13*, 1669. [CrossRef]
36. Prado, M.B.; Adiao, K.J.B. Methotrexate in generalized myasthenia gravis: A systematic review. *Acta Neurol. Belg.* **2023**, *123*, 1679–1691. [CrossRef] [PubMed]
37. Chen, Y.T.; Chang, Y.H.; Pathak, N.; Tzou, S.C.; Luo, Y.C.; Hsu, Y.C.; Li, T.N.; Lee, J.Y.; Chen, Y.C.; Huang, Y.W.; et al. Methotrexate inhibition of SARS-CoV-2 entry, infection and inflammation revealed by bioinformatics approach and a hamster model. *Front. Immunol.* **2022**, *13*, 1080897. [CrossRef]
38. Volpe, E.; Sambucci, M.; Battistini, L.; Borsellino, G. Fas-fas ligand: Checkpoint of t cell functions in multiple sclerosis. *Front. Immunol.* **2016**, *7*, 217466. [CrossRef]
39. Coënon, L.; Villalba, M. From CD16a Biology to Antibody-Dependent Cell-Mediated Cytotoxicity Improvement. *Front. Immunol.* **2022**, *13*, 913215. [CrossRef] [PubMed]
40. Nussinov, R.; Tsai, C.J. Allostery in disease and in drug discovery. *Cell* **2013**, *153*, 293–305. [CrossRef]
41. Mobley, D.L.; Dill, K.A. Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get". *Structure* **2009**, *17*, 489–498. [CrossRef]
42. Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964–972. [CrossRef]
43. Hopkins, A.L.; Groom, C.R.; Alex, A. Ligand efficiency: A useful metric for lead selection. *Drug Discov. Today* **2004**, *9*, 430–431. [CrossRef]
44. Cheng, A.C.; Coleman, R.G.; Smyth, K.T.; Cao, Q.; Soulard, P.; Caffrey, D.R.; Salzberg, A.C.; Huang, E.S. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **2007**, *25*, 71–75. [CrossRef] [PubMed]
45. Shoichet, B.K.; Kobilka, B.K. Structure-based drug screening for G-protein-coupled receptors. *Trends Pharmacol. Sci.* **2012**, *33*, 268–272. [CrossRef] [PubMed]
46. Naderi, M.; Lemoine, J.M.; Govindaraj, R.G.; Kana, O.Z.; Feinstein, W.P.; Brylinski, M. Binding site matching in rational drug design: Algorithms and applications. *Brief. Bioinform.* **2019**, *20*, 2167–2184. [CrossRef] [PubMed]
47. Çınaroğlu, S.S.; Timuçin, E. Insights into an alternative benzofuran binding mode and novel scaffolds of polyketide synthase 13 inhibitors. *J. Mol. Model.* **2019**, *25*, 130. [CrossRef]
48. Rai, H.; Barik, A.; Singh, Y.P.; Suresh, A.; Singh, L.; Singh, G.; Nayak, U.Y.; Dubey, V.K.; Modi, G. Molecular Docking, Binding Mode Analysis, Molecular Dynamics, and Prediction of ADMET/Toxicity Properties of Selective Potential Antiviral Agents against SARS-CoV-2 Main Protease: An Effort toward Drug Repurposing to Combat COVID-19. *Mol. Divers.* **2021**, *25*, 1905–1927. [CrossRef]
49. Owliaee, I.; Khaledian, M.; Shojaeian, A.; Madanchi, H.; Yarani, R.; Boroujeni, A.K.; Shoushtari, M. Antimicrobial Peptides Against Arboviruses: Mechanisms, Challenges, and Future Directions. *Probiotics Antimicrob. Proteins* **2025**. [CrossRef]
50. Hamosh, A.; Scott, A.F.; Amberger, J.S.; Bocchini, C.A.; McKusick, V.A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **2005**, *33*, D514–D517. [CrossRef]
51. Darif, D.; Hammi, I.; Kihel, A.; El Idrissi Saik, I.; Guessous, F.; Akarid, K. The pro-inflammatory cytokines in COVID-19 pathogenesis: What goes wrong? *Microb. Pathog.* **2021**, *153*, 104799. [CrossRef]

52. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **2016**, *44*, W90–W97. [CrossRef]
53. Yoo, M.; Shin, J.; Kim, J.; Ryall, K.A.; Lee, K.; Lee, S.; Jeon, M.; Kang, J.; Tan, A.C. DSigDB: Drug signatures database for gene set analysis. *Bioinformatics* **2015**, *31*, 3069–3071. [CrossRef] [PubMed]
54. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717. [CrossRef] [PubMed]
55. Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M.C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **2003**, *31*, 3381–3385. [CrossRef]
56. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; Wiegers, J.; Wiegers, T.C.; Mattingly, C.J. Comparative Toxicogenomics Database (CTD): Update 2021. *Nucleic Acids Res* **2021**, *49*, D1138–D1143. [CrossRef]
57. Apweiler, R.; Bairoch, A.; Bougueleret, L.; Altairac, S.; Amendolia, V.; Auchincloss, A.; Argoud-Puy, G.; Axelsen, K.; Baratin, D.; Blatter, M.C.; et al. The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.* **2009**, *37*, D169–D174. [CrossRef]
58. Dogan, B.; Gumusoglu, E.; Ulgen, E.; Sezerman, O.U.; Gunel, T. Integrated bioinformatics analysis of validated and circulating miRNAs in ovarian cancer. *Genomics Inform.* **2022**, *20*, e20. [CrossRef] [PubMed]
59. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S.; et al. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646. [CrossRef]
60. Chin, C.H.; Chen, S.H.; Wu, H.H.; Ho, C.W.; Ko, M.T.; Lin, C.Y. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **2014**, *8*, S11. [CrossRef]
61. Bader, G.D.; Hogue, C.W.V. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **2003**, *4*, 2. [CrossRef]
62. 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]
63. Lex, A.; Gehlenborg, N.; Strobelt, H.; Vuillemot, R.; Pfister, H. UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 1983–1992. [CrossRef] [PubMed]
64. Elkhalifa, A.E.O.; Al-Shammari, E.; Kuddus, M.; Adnan, M.; Sachidanandan, M.; Awadelkareem, A.M.; Qattan, M.Y.; Khan, M.I.; Abduljabbar, S.I.; Sarwar Baig, M.; et al. Structure-Based Multi-Targeted Molecular Docking and Dynamic Simulation of Soybean-Derived Isoflavone Genistin as a Potential Breast Cancer Signaling Proteins Inhibitor. *Life* **2023**, *13*, 1739. [CrossRef] [PubMed]
65. Bhatt, R.; Wang, A.; Durrant, J.D. Teaching old docks new tricks with machine learning enhanced ensemble docking. *Sci. Rep.* **2024**, *14*, 20722. [CrossRef]
66. Torres, P.H.M.; Sodero, A.C.R.; Jofily, P.; Silva, F.P., Jr. Key Topics in Molecular Docking for Drug Design. *Int. J. Mol. Sci.* **2019**, *20*, 4574. [CrossRef]
67. Krishnan, A.; Kocab, A.J.; Zacks, D.N.; Marshak-Rothstein, A.; Gregory-Ksander, M. A small peptide antagonist of the Fas receptor inhibits neuroinflammation and prevents axon degeneration and retinal ganglion cell death in an inducible mouse model of glaucoma. *J. Neuroinflam.* **2019**, *16*, 184. [CrossRef]
68. Jiang, V.C.; Liu, Y.; Jordan, A.; Leeming, A.; McIntosh, J.; Huang, S.; Zhang, R.; Cai, Q.; Chen, Z.; Li, Y.; et al. Targeting FcγRIIB by antagonistic antibody BI-1206 improves the efficacy of rituximab-based therapies in aggressive mantle cell lymphoma. *J. Hematol. Oncol.* **2022**, *15*, 42. [CrossRef] [PubMed]
69. Kakutani, M.; Takeuchi, K.; Waga, I.; Iwamura, H.; Wakitani, K. JTE-607, a novel inflammatory cytokine synthesis inhibitor without immunosuppression, protects from endotoxin shock in mice. *Inflamm. Res.* **1999**, *48*, 461–468. [CrossRef]
70. Perron, M.D.; Chowdhury, S.; Aubry, I.; Purisima, E.; Tremblay, M.L.; Uri Saragovi, H. Allosteric noncompetitive small molecule selective inhibitors of CD45 tyrosine phosphatase suppress T-cell receptor signals and inflammation in vivo. *Mol. Pharmacol.* **2014**, *85*, 553–563. [CrossRef]
71. Goddard, T.D.; Huang, C.C.; Meng, E.C.; Pettersen, E.F.; Couch, G.S.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **2018**, *27*, 14–25. [CrossRef]
72. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Meng, E.C.; Couch, G.S.; Croll, T.I.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*, 70–82. [CrossRef]
73. O'Boyle, N.M.; Morley, C.; Hutchison, G.R. Pybel: A Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.* **2008**, *2*, 5. [CrossRef] [PubMed]

74. Vázquez-Jiménez, L.K.; Juárez-Saldivar, A.; Gómez-Escobedo, R.; Delgado-Maldonado, T.; Méndez-Álvarez, D.; Palos, I.; Bandyopadhyay, D.; Gaona-Lopez, C.; Ortiz-Pérez, E.; Nogueda-Torres, B.; et al. Ligand-Based Virtual Screening and Molecular Docking of Benzimidazoles as Potential Inhibitors of Triosephosphate Isomerase Identified New Trypanocidal Agents. *Int. J. Mol. Sci.* **2022**, *23*, 10047. [CrossRef]
75. Al Noman, A.; Dev Sharma, P.; Jahin Mim, T.; Al Azad, M.; Sharma, H. Molecular docking and ADMET analysis of coenzyme Q10 as a potential therapeutic agent for Alzheimer's disease. *Aging Pathobiol. Ther.* **2024**, *6*, 170–182. [CrossRef]
76. Srinivasan, S.; Sadasivam, S.K.; Gunalan, S.; Shanmugam, G.; Kothandan, G. Application of docking and active site analysis for enzyme linked biodegradation of textile dyes. *Environ. Pollut.* **2019**, *248*, 599–608. [CrossRef] [PubMed]
77. Naveed, M.; Salah Ud Din, M.; Aziz, T.; Javed, T.; Miraj Khan, S.; Naveed, R.; Ali Khan, A.; Alharbi, M. Comparative analysis among the degradation potential of enzymes obtained from Escherichia coli against the toxicity of sulfur dyes through molecular docking. *Z. Naturforschung* **2024**, *79*, 221–234. [CrossRef]
78. Wang, S.; Jiang, J.H.; Li, R.Y.; Deng, P. Docking-based virtual screening of TβR1 inhibitors: Evaluation of pose prediction and scoring functions. *BMC Chem.* **2020**, *14*, 52. [CrossRef] [PubMed]
79. Alhawarri, M.B.; Dianita, R.; Rawa, M.S.A.; Nogawa, T.; Wahab, H.A. Potential Anti-Cholinesterase Activity of Bioactive Compounds Extracted from *Cassia grandis* L.f. and *Cassia timoriensis* DC. *Plants* **2023**, *12*, 344. [CrossRef]
80. Grünberg, R.; Nilges, M.; Leckner, J. Flexibility and Conformational Entropy in Protein-Protein Binding. *Structure* **2006**, *14*, 683–693. [CrossRef]
81. Spassov, D.S. Binding Affinity Determination in Drug Design: Insights from Lock and Key, Induced Fit, Conformational Selection, and Inhibitor Trapping Models. *Int. J. Mol. Sci.* **2024**, *25*, 7124. [CrossRef]
82. Kim, T.; Zhen, J.; Lee, J.; Bauer, R.; Lee, C.; Kwon, B.O.; Chae, K.H.; Hong, S.; Giesy, J.P.; Chang, G.S.; et al. Influence of ligand's directional configuration, chrysenes as model compounds, on the binding activity with aryl hydrocarbon receptor. *Sci. Rep.* **2020**, *10*, 13821. [CrossRef]
83. Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **2011**, *39*, W270–W277. [CrossRef] [PubMed]
84. Bugnon, M.; Röhrig, U.F.; Goullieux, M.; Perez, M.A.S.; Daina, A.; Michielin, O.; Zoete, V. SwissDock 2024: Major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. *Nucleic Acids Res.* **2024**, *52*, W324–W332. [CrossRef] [PubMed]

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.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Pharmaceuticals Editorial Office
E-mail: pharmaceuticals@mdpi.com
www.mdpi.com/journal/pharmaceuticals



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editor. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editor and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.





Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-9002-6