



Structured Tandem Repeats in Protein Interactions

Juan Mac Donagh ^{1,2}, Abril Marchesini ^{2,3} , Agostina Spiga ^{1,2}, Maximiliano José Fallico ⁴, Paula Nazarena Arriás ⁵ , Alexander Miguel Monzon ⁶ , Aimilia-Christina Vagiona ⁷ , Mariane Gonçalves-Kulik ⁷ , Pablo Mier ⁷ and Miguel A. Andrade-Navarro ^{7,*}

¹ Science and Technology Department, National University of Quilmes, Bernal B1876, Argentina; macjuan17@gmail.com (J.M.D.); agostina.spiga@becarios.unq.edu.ar (A.S.)

² National Scientific and Technical Research Council (CONICET), Buenos Aires C1033AAJ, Argentina; abrilmarchesini@biol.unlp.edu.ar

³ Biotechnology and Molecular Biology Institute (IBBM, UNLP-CONICET), Faculty of Exact Sciences, University of La Plata, La Plata 1900, Argentina

⁴ Laboratory of Bioactive Compound Research and Development, Faculty of Exact Sciences, University of La Plata, La Plata 1900, Argentina; maxifallico@gmail.com

⁵ Department of Biomedical Sciences, University of Padova, Via U. Bassi 58/b, 35121 Padova, Italy

⁶ Department of Information Engineering, University of Padova, Via Giovanni Gradenigo 6/B, 35131 Padova, Italy; alexander.monzon@unipd.it

⁷ Institute of Organismic and Molecular Evolution, Faculty of Biology, Johannes Gutenberg University, Hans-Dieter-Hüsch-Weg 15, 55128 Mainz, Germany; avagiona@uni-mainz.de (A.-C.V.); magoncal@uni-mainz.de (M.G.-K.); munoz@uni-mainz.de (P.M.)

* Correspondence: andrade@uni-mainz.de

Abstract: Tandem repeats (TRs) in protein sequences are consecutive, highly similar sequence motifs. Some types of TRs fold into structural units that pack together in ensembles, forming either an (open) elongated domain or a (closed) propeller, where the last unit of the ensemble packs against the first one. Here, we examine TR proteins (TRPs) to see how their sequence, structure, and evolutionary properties favor them for a function as mediators of protein interactions. Our observations suggest that TRPs bind other proteins using large, structured surfaces like globular domains; in particular, open-structured TR ensembles are favored by flexible termini and the possibility to tightly coil against their targets. While, intuitively, open ensembles of TRs seem prone to evolve due to their potential to accommodate insertions and deletions of units, these evolutionary events are unexpectedly rare, suggesting that they are advantageous for the emergence of the ancestral sequence but are early fixed. We hypothesize that their flexibility makes it easier for further proteins to adapt to interact with them, which would explain their large number of protein interactions. We provide insight into the properties of open TR ensembles, which make them scaffolds for alternative protein complexes to organize genes, RNA and proteins.

Keywords: tandem repeats; protein–protein interactions; protein structure; protein evolution; protein flexibility



Citation: Mac Donagh, J.; Marchesini, A.; Spiga, A.; Fallico, M.J.; Arriás, P.N.; Monzon, A.M.; Vagiona, A.-C.; Gonçalves-Kulik, M.; Mier, P.; Andrade-Navarro, M.A. Structured Tandem Repeats in Protein Interactions. *Int. J. Mol. Sci.* **2024**, *25*, 2994. <https://doi.org/10.3390/ijms25052994>

Academic Editors: Jessica Holien and Istvan Simon

Received: 9 February 2024

Revised: 28 February 2024

Accepted: 1 March 2024

Published: 5 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tandem repeats (TRs) in protein sequences are approximate consecutive repeats, which can fold into units with similar folding packing together in a structural domain [1–3]. In 2012, Kajava categorized TRs that form structures according to their lengths into class I (length of 1–2 amino acids, forming crystalline structures), class II (length of 3–4 amino acids, fibrous proteins), and longer ones that can form open (class III) or closed (class IV) ensembles [3]. Open ensembles include solenoids, in which the array of TRs is elongated in one direction. Their units can be composed of a pure helical [4] or beta structure [5], or they can be mixed (e.g., Leucine Rich Repeats; LRRs [6]). Solenoids have a large surface and flexibility that allows them to interact with multiple proteins simultaneously as scaffold

proteins [7–9]. As a result, they interact with many proteins for a variety of cellular functions in many species widely distributed across the tree of life [10–12].

Well-known examples are importin alpha (which controls the traffic of proteins into the nucleus and binds proteins with a nuclear localization signal using Armadillo repeats [13]), mammalian ribonuclease inhibitor (which tightly coils around ribonuclease with an ensemble of LRRs [14]) and Huntingtin (a large protein with several ensembles of HEAT repeats, mostly studied for its role in Huntington's disease but with a large list of protein interactors, likely to play a role coordinating multiple cellular processes [15]).

Because each unit in ensembles of repeats has evolved to interact with its neighbours, ensembles of repeats can easily accommodate the insertion or deletion of a unit. This gives them an evolutionary advantage over globular domains because the ensemble of repeats can increase or decrease in size without the need for complex rearrangements of the folding of the domain, which would not be possible for a globular domain [16,17]. This is probably one of the reasons why TRs are so ubiquitous. Accordingly, an enrichment of proteins with conserved TRs in functions related to binding was observed [18]. Because TR proteins (TRPs) have a variety of types and particular properties, we wondered if the association of TRs to an interaction function could be related to general or specific properties of TRs giving them advantages for a function in protein interactions. To address this question, here we evaluate the properties of TRPs in relation to protein interactions.

2. Results

2.1. Comparison of the Properties of TRs and Other Protein Regions

We obtained the complete set of human proteins (a total of 20,600) and identified 11 types of structured TRs (Ankyrin, Armadillo, HAT, HEAT and three variants thereof, Kelch, LRR, PFTA, PFTB, RCC1, TPR and WD40) in these sequences using REP2 [19] (see [20] for details about these TR types). There are other methods for the detection of tandem repeats (e.g., TAPO [21], RepeatsDB-Lite [22], CE-Symm [23], and TRDistiller [24]), but for the purposes of this work, using a single method such as REP2, which detects and assigns TRs in protein sequences to a few well-defined categories of well-known structural repeats, is sufficient.

We detected a total of 7749 repeat units in 1005 proteins (with some redundancy). WD40, LRR and Ankyrin were the most abundant (Figure 1a; Supplementary Table S1).

Regarding the amount of protein interactions of proteins with these TRs (TRPs), we observed a large variability (Figure 1b). Open TRPs with ARM, HEAT and TPRs had more than 100 interactions on average, well above the average human protein (65, $n = 20,596$), proteins with annotated Pfam domains (70, $n = 17,329$) and TRP (79, $n = 1005$; Supplementary Table S2). However, open TRPs with ANK and LRR have below average numbers of protein–protein interactions (PPIs). Thus, in this respect, while TRPs have more interactions on average than the global human proteome, this is not specific to open or closed TRPs.

To identify properties of TRs relevant to protein interactions and how they compare to regions outside TRs and to globular domains, we defined different sequence ranges and computed the frequency of various annotations inside and outside TRs.

First, we studied the presence of Short Linear Motifs (SLiMs), which are sites for protein interactions [25,26], expecting to find an enrichment of those given the interacting character of TRs. We actually found the opposite result (Figure 1c). These motifs mediate interactions of a peptide with another protein [27]. We observed that globular domains have an even lower frequency of SLiMs than TRs (Figure 1c; see Methods for definition).

For comparison, we also tested the frequency of phosphorylation sites, as these are often present in the interfaces of interacting proteins, resulting in phosphorylation-dependent PPIs [28,29]. These sites were also depleted of TRs, even more than in globular domains (Figure 1d).

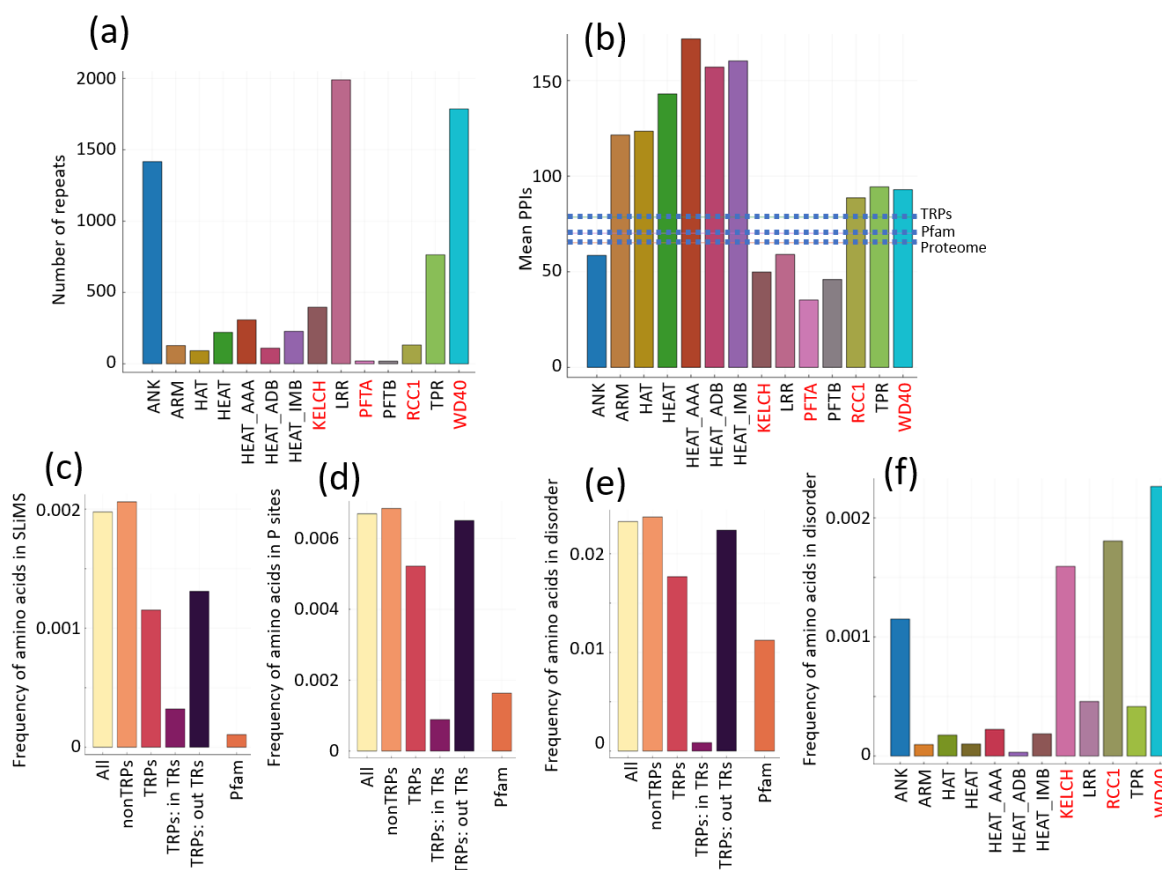


Figure 1. Properties of TRs by type and compared to other protein regions. **(a)** Number of repeat units identified. HEAT_AAA, HEAT_ADB and HEAT_IMB are three variants of HEAT repeats and can be redundant. TRs forming open and closed ensembles are indicated with black and red labels, respectively. **(b)** Average number of protein partners for TRPs. Horizontal dotted lines indicate the values for all human proteins (proteome), for proteins annotated with globular domains (Pfam), and for all TRPs. **(c)** Frequency of amino acids in SLiMs. Values are shown for: the complete human proteome (All), for proteins that do not contain TRs (non TRPs), for TRPs, for residues in TRs of TRPs (TRPs: in TRs), for residues outside TRs in TRPs (TRPs: out TRs), and in annotated globular domains (Pfam). **(d)** Frequency of amino acids in phosphorylation sites. **(e)** Frequency of amino acids in IDRs. **(f)** Disordered content by repeat type (PFTA and PFTB are not displayed because their numbers are too low). To prepare the plots (c–f), we obtained the coordinates of the features (SLiMs, phosphorylation sites, disorder regions) in all human sequences from the corresponding databases (see Methods for details) and then divided the number of residues within the given feature by the total of residues in the corresponding type of sequence.

To further substantiate this observation, since SLiMs and phosphorylation sites are very frequent in disordered regions [30,31], we checked the content of disorder in TRPs. While both IDRs and TRs can have a function in PPIs, they differ in that TRs have a structure, and therefore IDRs impose stronger constraints on the proteins that have them. For example, proteins with long IDRs have a shorter half-life [32].

We observed that the structured TRs analyzed have even less disorder content than globular domains (Figure 1e). TRs forming closed ensembles (KELCH, RCC1 and WD40) had the largest frequency of disorder (Figure 1f).

Together, these results suggest that TRs are involved in interactions using folded surfaces similarly to globular domains and less like disordered regions. In fact, regarding the properties analyzed, TRs are further from IDRs than globular domains.

We note that the regions of TRPs outside the TR ensemble have a similar disorder content and phosphorylation sites to the average protein but a lower content of SLiMs (not

so low as Pfam domains; Figure 1c). Since SLiMs tend to accumulate in IDRs, this is an indication that the IDRs of TRPs are depleted of SLiMs, suggesting that they would be less involved in PPIs mediated via SLiMs. We take this result as an indication that the TRs in TRPs are mainly responsible for their PPI function to the detriment of their IDRs, which would fulfil a more passive function, such as providing flexible linkers.

2.2. Flexibility

While our results indicate that TRPs behave as folded proteins in their interaction properties, they are well known to have a certain degree of flexibility, particularly those that form solenoids [2,33]. We wondered if there could be different properties in terms of the flexibility of TRs in open and closed structures. Restricted mobility of central residues in elongated ensembles of TRs has already been observed and is hypothesized to play a role in their interacting properties [34]. To examine the flexibility of TR ensembles, we studied known structures of human proteins with open and closed ensembles of TRs. As a normalized proxy for flexibility, we used the pLDDT scores in their corresponding AlphaFold predictions [35].

Considering the number of cases available, we were able to obtain enough data for one closed TR type (WD40 $n = 31$) and four open TR types (ANK $n = 25$, ARM $n = 11$, LRR $n = 24$, TRP $n = 20$). The average of pLDDT scores along the TR ensemble have values of around 93 with drops to 90 at the termini (Figure 2a). Individual open-ensemble TR types have some drops in the middle part, but the drops at the termini are found in all of them (Figure 2b). For comparison, the profile for the closed TR (WD40) has no such drops at the termini and displays slightly higher values (Figure 2a).

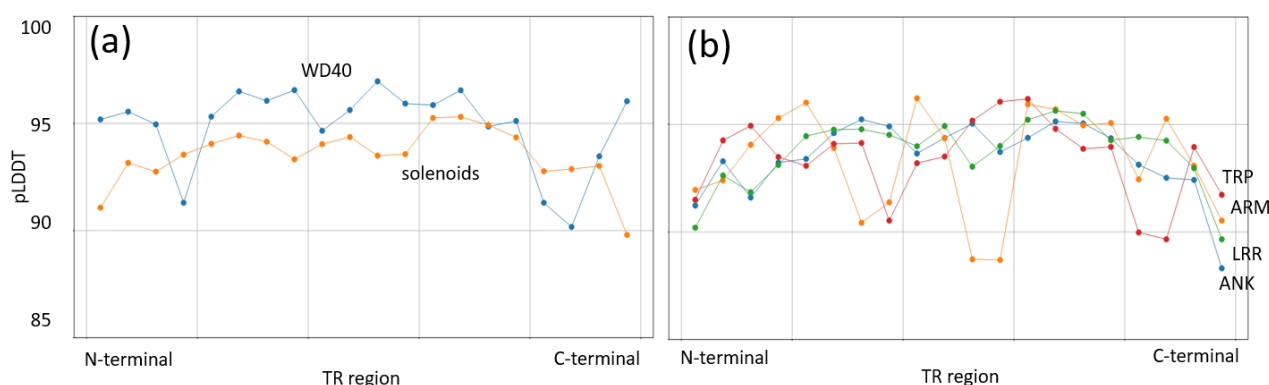


Figure 2. pLDDT scores of AlphaFold predictions along ensembles of TRs. (a) WD40 (closed ensemble) compared to an average of four open ensembles (shown in (b)). (b) Values for the four open ensembles. The x-axis indicates the relative position in the TR ensemble N- to C-terminal.

These results suggest that TRs that form open ensembles might be more flexible than closed ones, particularly at the termini. This could result in their ability to grasp proteins in their center and then coil the termini around. Disordered regions are often found at the termini of TR ensembles, which could contribute to this effect [11].

We illustrate this with examples of TRPs in complex with another protein (as annotated in RepeatsDB [36]; Figure 3). The elongated ensembles of repeats (ARM, LRR, HEAT) can coil around their bound proteins, whereas closed ensembles (RCC1, WD40) form a more compact domain that behaves more like a globular domain. We illustrate the flexibility of elongated TRPs showing structures of the HEAT repeat domain of importin subunit beta-1 (C-terminal at the top of the image) bound to three different proteins (histone H1.0, SNAI1 and Snurportin-1; Figure 3, right side): both the binding location and the shape of the TRP change in each interaction. This property is very specific to elongated TRPs.

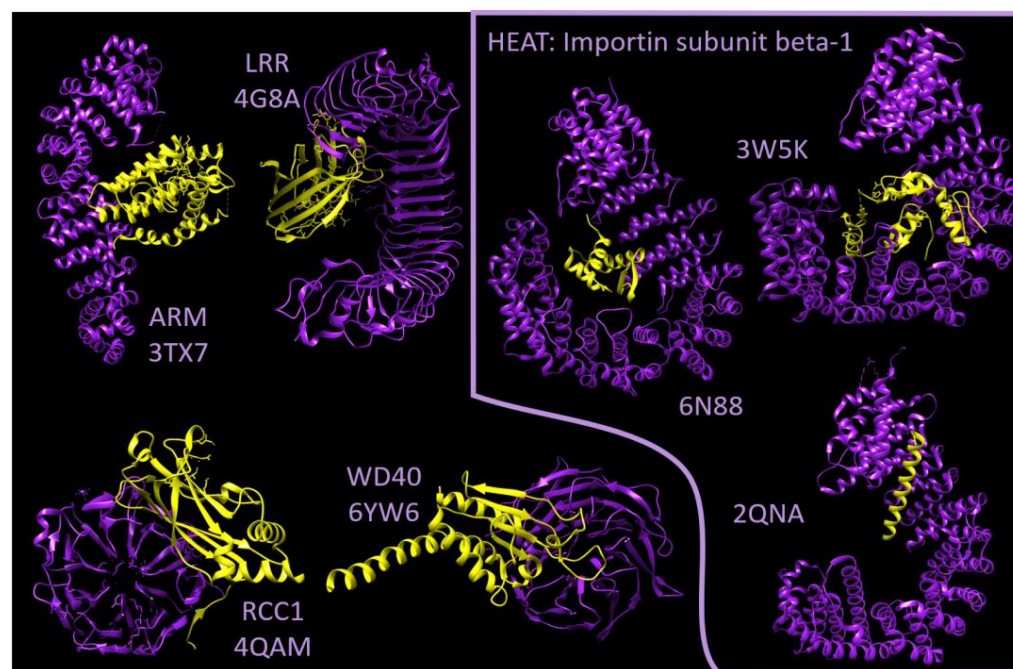


Figure 3. Flexibility of interacting TRPs. Structures of protein complexes with TRPs by repeat type. Elongated: ARM repeats in human catenin beta-1 binding NR5A2 (PDB:3TX7); LRR repeats in Toll-like receptor 4 binding LY96 (PDB:4G8A); HEAT repeats in importin subunit beta-1 shown in the same orientation, forming three complexes with histone H1.0 (PDB:6N88), Zinc finger protein SNAI1 (PDB:3W5K) and the IBB domain of Snurportin-1 (PDB:2QNA). Cyclic: RCC1 in RPGR repeats binding the interacting domain of RPGRIP1 (PDB:4QAM) and KELCH repeats in ARPC1B binding ARPC4 (PDB:6YW6). TRP in purple and bound protein in yellow.

2.3. Length Variability

The evolutionary adaptability of TRPs for changes in the number of TR units is yet another property of open ensembles of TRs that distinguishes them from close TRs and makes them appealing molecules for protein interactions [7,37].

Changes in the number of units of ensembles of TRs have been observed by examination of orthologs from species at long divergence times (see e.g., for the mineralocorticoid receptor in human and fish [38]). With the sequencing of many complete genomes for species, strains and populations, it has also become possible to assess the fast evolution of proteins by comparing sequences from closely related organisms; one recent example in relation to sequence repeats is the examination of the variation of short TRs across different populations of the plant *Arabidopsis thaliana* [39].

To add examples specific to the structured TRs studied here and to illustrate the evolutionary variation between closely related species, we used the complete proteomes of six species of *Plasmodium*. The study of TRs in these species is particularly interesting because of their involvement in malaria. Clarifying features that could impart *Plasmodium* with evolutionary variability may contribute to an understanding of its pathogenicity in terms of adaptability to the host. In *Plasmodium*, short tandem repeats are variable [40], and here we examine whether this is the case for structured TRs.

We grouped their complete proteomes by sequence similarity (see Methods for details), predicted TRs in their sequences and then found cases where the number of predicted TRs was different between orthologs.

We illustrate this fast evolution of TRs with two examples (Figure 4). The two cases regard open ensemble types (LRR and TPR, respectively), which is not surprising as it is easier to add or remove units to open ensembles than to closed ones, which have a more restricted folding [7]. Regarding their taxonomic distribution, LRRs are particularly predominant in plants, and TPRs are particularly predominant in Archaea and Bacteria,

while in *Plasmodium falciparum* WD40 repeats (which form a closed ensemble) are actually more frequent [19].

In both cases, we have the insertion of a new unit in the ortholog of the *P. malariae* protein, and the region seems to be surrounded by disorder (Figure 4). The AlphaFold structure predictions [41] for the *P. malariae* sequences are available in UniProt (versions AF-A0A1D3JKH0-F1 and AF-A0A1D3TFE6-F1, respectively). None of them predict the insertion as a TR unit, and therefore our sequence-based prediction could be incorrect. The AlphaFold prediction for our extra predicted LRR in A0A1D3JKH0_PLAMA positions 595–618 does not find this LRR, but it has difficulties in predicting other repeats (Figure 4a; bottom left). However, the prediction for the A0A1D3TFE6_PLAMA sequence with the extra TPR in positions 660–693 suggests some helical structure (Figure 4b; bottom left).

We also obtained predictions using the Robetta structure prediction server (using RoseTTAFold [42]). These predictions contain less disorder. Interestingly, in both cases, unlike AlphaFold, Robetta predicts a compact unit, formed by two secondary structure elements packed in an anti-parallel way against each other, with the start and the end of the unit in close proximity in 3D (Figure 4a,b; bottom right). However, the predicted secondary structures (alpha-alpha and beta-alpha, respectively) are not those of the repeats in the ensemble (LRR and TPR, respectively).

These two examples have common features that suggest that insertions in open ensembles of TRs could contain extra TRs. Structural predictions by AlphaFold do not confirm them, but alternative structural predictions by Robetta suggest that the insertions might contain units with the size and anti-parallel packing of the TRs in the ensemble. Therefore, while these extra TRs could be false positives from the sequence detection method, it is not possible to completely rule out that disordered insertions in TRs could be a seed for the evolution of new TRs.

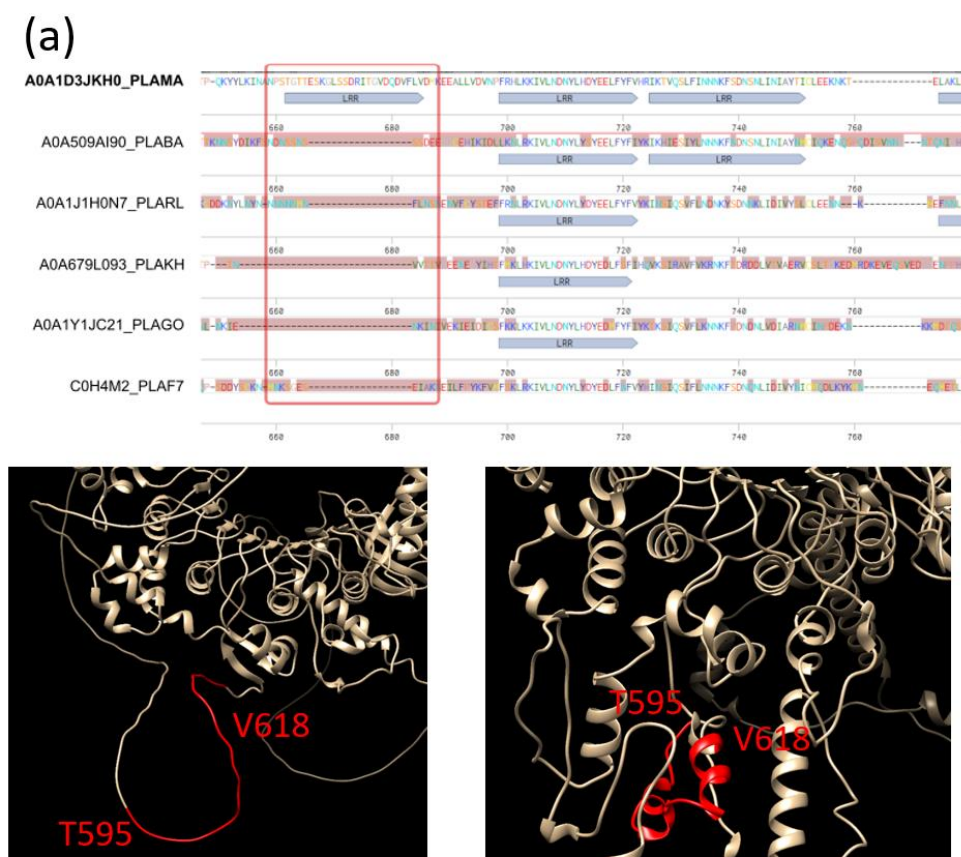


Figure 4. Cont.

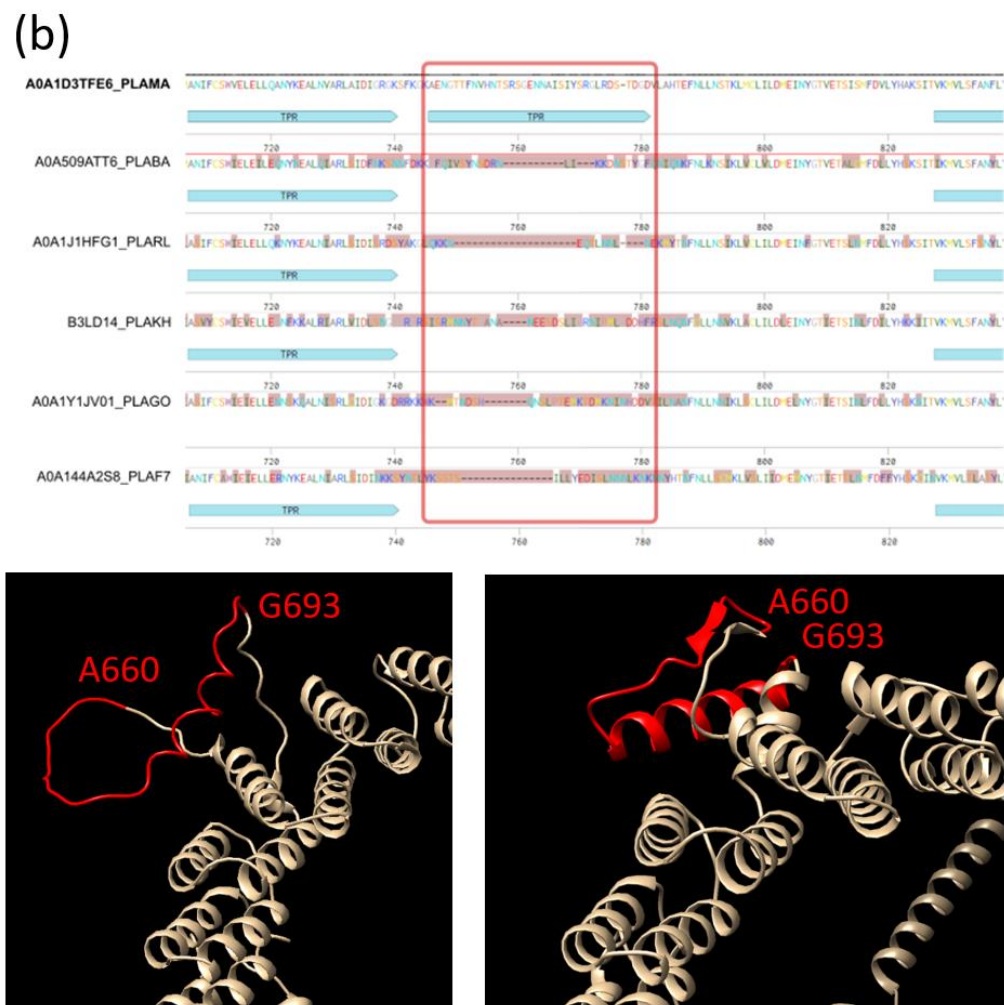


Figure 4. Cases of fast evolution in *Plasmodium* species. (a) Gain of an LRR unit in A0A1D3JKH0_PLAMA positions 595–618. (b) Gain of a TPR unit in A0A1D3TFE6_PLAMA positions 660–693 (colored in red in the structure). Sequence identifiers from UniProtKB. In order, species are: *Plasmodium berghei*, *Plasmodium relictum*, *Plasmodium malariae*, *Plasmodium knowlesi*, *Plasmodium gonderi* and *Plasmodium falciparum*. No alternative spliced isoforms for A0A1D3JKH0_PLAMA or A0A1D3TFE6_PLAMA are given in UniProt (February 2024). The structures shown are models from AlphaFold [41] and Robetta [42] (left and right, respectively).

3. Discussion

The capacity for TRPs to interact with other proteins favors their participation in functions that require the formation of protein complexes. The examination of the functional enrichment of proteins that interact with human TRPs (see Methods for details) indicates that they often operate in processes in the nucleus involving RNA binding (Figure 5). This is in agreement with the observation that most TRPs bind RNA, but through disordered regions and not directly with structural TRs [43]. As a result, functions in the regulation of transcription, expression and splicing are found. Protein regulation functions are associated with ARM (regulation of apoptosis) and TPR (phosphorylation).

The lack of evolutionary variation in the number of TR units we exemplified in *Plasmodium* has also been observed at long evolutionary distances in eukaryotes [18] and in plants [44]. While unit number variation might be advantageous for the establishment of the ancestral interaction, further evolution of TRPs seems to be rare. This is likely related to the mode of interaction of TRPs, which we found shares some aspects with that of globular domains. While it might be simple to modify the TRP side of a protein–protein interaction, the globular non-TRP interacting partner is not free to change, and

thus the interacting interface is frozen, meaning that none of the partners will evolve. If both interacting partners are TRPs (or TRPs interact with an IDR), we could expect the evolutionary constraints to be much lower. Fast evolutionary rates of TRPs have been observed in giant viruses (e.g., ANK in the *Acanthamoeba polyphaga mimivirus*; [45]); we would hypothesize that the function of these proteins is to interact with other TRPs or IDRs, or maybe not with any protein.

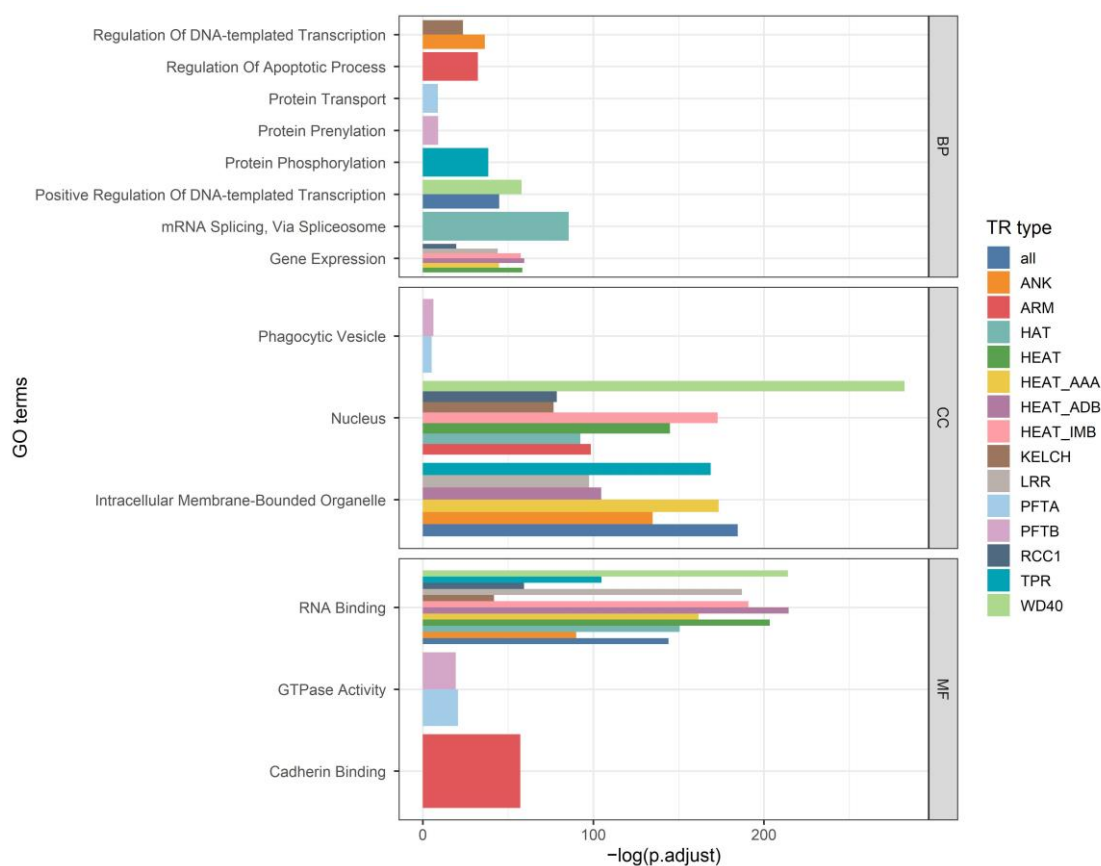


Figure 5. Functional enrichment of TRP interactors. **Top:** Biological Process. **Middle:** Cellular Component. **Bottom:** Molecular Function. Gene Ontology (GO) enrichment analysis was carried out for a set of TRP interactors (All) and then separately for the interactors of each TRP type (see Methods for details). Enriched GO Biological Process (BP), Molecular Function (MF) and Cellular Component (CC) terms with the lowest adjusted p -value were kept.

Our observations provide a two-sided view of TRs. On the one hand, they provide modes of interaction by complementarity to large-structured surfaces (similar to globular domains) and by not using motifs and sites of phosphorylation (as disordered regions do). This results in a lack of evolutionary capacity. While inherent evolutionary plasticity exists, this is restricted to the generation of an ancestral surface for interaction that remains fixed.

On the other hand, elongated TRs have flexibility at long range, which allows them to interact with many partners and adopt different shapes depending on the interacting partners. Although the TRP is evolutionarily fixed, TRP flexibility can be exploited by novel partners adapting to interact with existing TRPs. The advantage here is that even if the length of the TRP cannot change, the conformational variability offers more possibilities for those new interactors that can bend the TRPs to their interacting needs. This would explain the fact that TRPs in general have more interactors than non-TRPs. Our results and interpretations support a bilateral view of open TRs, explaining their success and the multiplicity of TR types that have emerged in evolution.

4. Materials and Methods

A total of 20,600 human proteins were obtained from the human reference proteome from the UniProtKb release 2022_04 [46]. We identified 11 types of TRs with structural information (Ankyrin, Armadillo, HAT, HEAT and three variants thereof, Kelch, LRR, PFTA, PFTB, RCC1, TPR and WD40) in these sequences using REP2 [19]. A total of 7749 repeat units were detected in 1005 proteins.

Protein interactions between human proteins were obtained from the HIPPIE database [47]. SLiMs were evaluated using the ELM database [48], and phosphorylation sites were evaluated using the Phospho ELM database [49]. Globular domains were obtained from the InterPro annotations of Pfam domains in sequences [50]. Disordered regions were obtained using a consensus of predictors as defined in MobiDB [51]. Known structures of human TRPs and annotated domains of TRs were obtained from RepeatsDB [36]. Multiple sequence alignments were produced using Clustal Omega [52].

For the study of TRs in *Plasmodium* species, we obtained the complete reference proteome (not considering isoforms) of the following six species from UniProtKB release 2022_04: *Plasmodium berghei* (strain Anka), *Plasmodium falciparum* (isolate 3D7), *Plasmodium gonderi* (ATCC 30045), *Plasmodium knowlesi* (strain H), *Plasmodium malariae*, and *Plasmodium relictum* (SGS1). A total of 37,588 sequences were retrieved. They were grouped in groups of orthologs using OrthoFinder [53], and TRs were predicted in them using REP2 [19].

Interactors with human TRPs were obtained from the HIPPIE database [47]. Functional enrichment of protein sets was computed using Enricher [54].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms25052994/s1>.

Author Contributions: Conceptualization, M.A.A.-N.; methodology, investigation and visualization, all authors; writing—original draft preparation, M.A.A.-N.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union’s Horizon 2020 research and innovation programme under Marie Skłodowska-Curie [823886].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article and Supplementary Materials.

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

References

1. Kobe, B.; Kajava, A.V. When protein folding is simplified to protein coiling: The continuum of solenoid protein structures. *Trends Biochem. Sci.* **2000**, *25*, 509–515. [\[CrossRef\]](#)
2. Monzon, A.M.; Arriás, P.N.; Elofsson, A.; Mier, P.; Andrade-Navarro, M.A.; Bevilacqua, M.; Clementel, D.; Bateman, A.; Hirsh, L.; Fornasari, M.S.; et al. A STRP-ed definition of Structured Tandem Repeats in Proteins. *J. Struct. Biol.* **2023**, *215*, 108023. [\[CrossRef\]](#)
3. Kajava, A.V. Tandem repeats in proteins: From sequence to structure. *J. Struct. Biol.* **2012**, *179*, 279–288. [\[CrossRef\]](#)
4. Groves, M.R.; Barford, D. Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.* **1999**, *9*, 383–389. [\[CrossRef\]](#)
5. Kajava, A.V.; Steven, A.C. Beta-rolls, beta-helices, and other beta-solenoid proteins. *Adv. Protein Chem.* **2006**, *73*, 55–96. [\[CrossRef\]](#)
6. Kobe, B.; Deisenhofer, J. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **1993**, *366*, 751–756. [\[CrossRef\]](#)
7. Andrade, M.A.; Perez-Iratxeta, C.; Ponting, C.P. Protein repeats: Structures, functions, and evolution. *J. Struct. Biol.* **2001**, *134*, 117–131. [\[CrossRef\]](#)
8. Grove, T.Z.; Cortajarena, A.L.; Regan, L. Ligand binding by repeat proteins: Natural and designed. *Curr. Opin. Struct. Biol.* **2008**, *18*, 507–515. [\[CrossRef\]](#)
9. Linke, W.A. Sense and stretchability: The role of titin and titin-associated proteins in myocardial stress-sensing and mechanical dysfunction. *Cardiovasc. Res.* **2008**, *77*, 637–648. [\[CrossRef\]](#)
10. Dosztanyi, Z.; Chen, J.; Dunker, A.K.; Simon, I.; Tompa, P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J. Proteome Res.* **2006**, *5*, 2985–2995. [\[CrossRef\]](#)
11. Delucchi, M.; Schaper, E.; Sachenkova, O.; Elofsson, A.; Anisimova, M. A New Census of Protein Tandem Repeats and Their Relationship with Intrinsic Disorder. *Genes* **2020**, *11*, 407. [\[CrossRef\]](#)

12. D'Andrea, L.D.; Regan, L. TPR proteins: The versatile helix. *Trends Biochem. Sci.* **2003**, *28*, 655–662. [\[CrossRef\]](#)
13. Oka, M.; Yoneda, Y. Importin alpha: Functions as a nuclear transport factor and beyond. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **2018**, *94*, 259–274. [\[CrossRef\]](#)
14. Dickson, K.A.; Haigis, M.C.; Raines, R.T. Ribonuclease inhibitor: Structure and function. *Prog. Nucleic Acid. Res. Mol. Biol.* **2005**, *80*, 349–374. [\[CrossRef\]](#)
15. Saudou, F.; Humbert, S. The Biology of Huntingtin. *Neuron* **2016**, *89*, 910–926. [\[CrossRef\]](#)
16. Marcotte, E.M.; Pellegrini, M.; Yeates, T.O.; Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **1999**, *293*, 151–160. [\[CrossRef\]](#)
17. Bjorklund, A.K.; Ekman, D.; Elofsson, A. Expansion of protein domain repeats. *PLoS Comput. Biol.* **2006**, *2*, e114. [\[CrossRef\]](#)
18. Schaper, E.; Gascuel, O.; Anisimova, M. Deep conservation of human protein tandem repeats within the eukaryotes. *Mol. Biol. Evol.* **2014**, *31*, 1132–1148. [\[CrossRef\]](#)
19. Kamel, M.; Kastano, K.; Mier, P.; Andrade-Navarro, M.A. REP2: A Web Server to Detect Common Tandem Repeats in Protein Sequences. *J. Mol. Biol.* **2021**, *433*, 166895. [\[CrossRef\]](#)
20. Andrade, M.A.; Ponting, C.P.; Gibson, T.J.; Bork, P. Homology-based method for identification of protein repeats using statistical significance estimates. *J. Mol. Biol.* **2000**, *298*, 521–537. [\[CrossRef\]](#)
21. Do Viet, P.; Roche, D.B.; Kajava, A.V. TAPPO: A combined method for the identification of tandem repeats in protein structures. *FEBS Lett.* **2015**, *589*, 2611–2619. [\[CrossRef\]](#)
22. Hirsh, L.; Paladin, L.; Piovesan, D.; Tosatto, S.C.E. RepeatsDB-lite: A web server for unit annotation of tandem repeat proteins. *Nucleic Acids Res.* **2018**, *46*, W402–W407. [\[CrossRef\]](#)
23. Bliven, S.E.; Lafita, A.; Rose, P.W.; Capitani, G.; Prlic, A.; Bourne, P.E. Analyzing the symmetrical arrangement of structural repeats in proteins with CE-Symm. *PLoS Comput. Biol.* **2019**, *15*, e1006842. [\[CrossRef\]](#)
24. Richard, F.D.; Kajava, A.V. TRDistiller: A rapid filter for enrichment of sequence datasets with proteins containing tandem repeats. *J. Struct. Biol.* **2014**, *186*, 386–391. [\[CrossRef\]](#)
25. Tompa, P.; Davey, N.E.; Gibson, T.J.; Babu, M.M. A million peptide motifs for the molecular biologist. *Mol. Cell.* **2014**, *55*, 161–169. [\[CrossRef\]](#)
26. Van Roey, K.; Dinkel, H.; Weatheritt, R.J.; Gibson, T.J.; Davey, N.E. The switches.ELM resource: A compendium of conditional regulatory interaction interfaces. *Sci. Signal.* **2013**, *6*, rs7. [\[CrossRef\]](#)
27. Neduva, V.; Russell, R.B. Peptides mediating interaction networks: New leads at last. *Curr. Opin. Biotechnol.* **2006**, *17*, 465–471. [\[CrossRef\]](#)
28. Yaffe, M.B. Phosphotyrosine-binding domains in signal transduction. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 177–186. [\[CrossRef\]](#)
29. Jin, J.; Pawson, T. Modular evolution of phosphorylation-based signalling systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2012**, *367*, 2540–2555. [\[CrossRef\]](#)
30. Davey, N.E.; Cyert, M.S.; Moses, A.M. Short linear motifs—Ex nihilo evolution of protein regulation. *Cell Commun. Signal.* **2015**, *13*, 43. [\[CrossRef\]](#)
31. 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. [\[CrossRef\]](#)
32. Van der Lee, R.; Lang, B.; Kruse, K.; Gsponer, J.; de Groot, N.S.; Huynen, M.A.; Matouschek, A.; Fuxreiter, M.; Babu, M.M. Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep.* **2014**, *8*, 1832–1844. [\[CrossRef\]](#)
33. Burgi, J.; Ekal, L.; Wilmanns, M. Versatile allosteric properties in Pex5-like tetratricopeptide repeat proteins to induce diverse downstream function. *Traffic* **2021**, *22*, 140–152. [\[CrossRef\]](#)
34. Ramya, L.; Gautham, N.; Chaloin, L.; Kajava, A.V. Restricted mobility of side chains on concave surfaces of solenoid proteins may impart heightened potential for intermolecular interactions. *Proteins* **2015**, *83*, 1654–1664. [\[CrossRef\]](#)
35. Ma, P.; Li, D.W.; Bruschweiler, R. Predicting protein flexibility with AlphaFold. *Proteins* **2023**, *91*, 847–855. [\[CrossRef\]](#)
36. Paladin, L.; Bevilacqua, M.; Errigo, S.; Piovesan, D.; Mičetić, I.; Necci, M.; Monzon, A.M.; Fabre, M.L.; Lopez, J.L.; Nilsson, J.F.; et al. RepeatsDB in 2021: Improved data and extended classification for protein tandem repeat structures. *Nucleic Acids Res.* **2021**, *49*, D452–D457. [\[CrossRef\]](#)
37. Li, J.; Mahajan, A.; Tsai, M.D. Ankyrin repeat: A unique motif mediating protein-protein interactions. *Biochemistry* **2006**, *45*, 15168–15178. [\[CrossRef\]](#)
38. Vlassi, M.; Brauns, K.; Andrade-Navarro, M.A. Short tandem repeats in the inhibitory domain of the mineralocorticoid receptor: Prediction of a beta-solenoid structure. *BMC Struct. Biol.* **2013**, *13*, 17. [\[CrossRef\]](#)
39. Reinart, W.B.; Greulich, A.; Stø, I.M.; Knutsen, J.B.; Reitan, T.; Tørresen, O.K.; Jentoft, S.; Butenko, M.A.; Jakobsen, K.S. Adaptive protein evolution through length variation of short tandem repeats in Arabidopsis. *Sci. Adv.* **2023**, *9*, eadd6960. [\[CrossRef\]](#)
40. Davies, H.M.; Nofal, S.D.; McLaughlin, E.J.; Osborne, A.R. Repetitive sequences in malaria parasite proteins. *FEMS Microbiol. Rev.* **2017**, *41*, 923–940. [\[CrossRef\]](#)
41. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [\[CrossRef\]](#)
42. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876. [\[CrossRef\]](#)
43. Ormazábal, A.; Carletti, M.S.; Saldaña, T.E.; Gonzalez Buitron, M.; Marchetti, J.; Palopoli, N.; Bateman, A. Expanding the repertoire of human tandem repeat RNA-binding proteins. *PLoS ONE* **2023**, *18*, e0290890. [\[CrossRef\]](#)

44. Schaper, E.; Anisimova, M. The evolution and function of protein tandem repeats in plants. *New Phytol.* **2015**, *206*, 397–410. [[CrossRef](#)]
45. Erdozain, S.; Barrionuevo, E.; Ripoll, L.; Mier, P.; Andrade-Navarro, M.A. Protein repeats evolve and emerge in giant viruses. *J. Struct. Biol.* **2023**, *215*, 107962. [[CrossRef](#)]
46. The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523–D531. [[CrossRef](#)]
47. Alanis-Lobato, G.; Andrade-Navarro, M.A.; Schaefer, M.H. HIPPIE v2.0: Enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res.* **2017**, *45*, D408–D414. [[CrossRef](#)]
48. Kumar, M.; Michael, S.; Alvarado-Valverde, J.; Zeke, A.; Lazar, T.; Glavina, J.; Nagy-Kanta, E.; Donagh, J.M.; Kalman, Z.E.; Pascarelli, S.; et al. ELM-the Eukaryotic Linear Motif resource-2024 update. *Nucleic Acids Res.* **2023**, *55*, D442–D455. [[CrossRef](#)]
49. Dinkel, H.; Chica, C.; Via, A.; Gould, C.M.; Jensen, L.J.; Gibson, T.J.; Diella, F. Phospho.ELM: A database of phosphorylation sites—update 2011. *Nucleic Acids Res.* **2011**, *39*, D261–D267. [[CrossRef](#)]
50. Paysan-Lafosse, T.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B.L.; Salazar, G.A.; Bileschi, M.L.; Bork, P.; Bridge, A.; Colwell, L.; et al. InterPro in 2022. *Nucleic Acids Res.* **2023**, *51*, D418–D427. [[CrossRef](#)]
51. Piovesan, D.; Necci, M.; Escobedo, N.; Monzon, A.M.; Hatos, A.; Mičetić, I.; Quaglia, F.; Paladin, L.; Ramasamy, P.; Dosztányi, Z.; et al. MobiDB: Intrinsically disordered proteins in 2021. *Nucleic Acids Res.* **2021**, *49*, D361–D367. [[CrossRef](#)]
52. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [[CrossRef](#)]
53. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [[CrossRef](#)]
54. 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](#)]

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.