

REVIEW ARTICLE

Diversity and structural-functional insights of alpha-solenoid proteins

Paula Nazarena Arrias^{1,2}  | Zarifa Osmanli¹  | Estefanía Peralta³ |
Patricio Manuel Chinestrada⁴ | Alexander Miguel Monzon⁵  |
Silvio C. E. Tosatto^{1,6} 

¹Department of Biomedical Sciences, University of Padova, Padova, Italy

²Department of Protein Science, KTH Royal Institute of Technology, Stockholm, Sweden

³Laboratorio de Investigación y Desarrollo de Bioactivos (LIDeB), Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina

⁴Laboratorio de Farmacología Molecular, Universidad Nacional de Quilmes, Bernal, Buenos Aires, Argentina

⁵Department of Information Engineering, University of Padova, Padova, Italy

⁶Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council (CNR-IBIOM), Bari, Italy

Correspondence

Silvio C. E. Tosatto, Department of Biomedical Sciences, University of Padova, via U. Bassi 58/b, 35121 Padova, Italy.

Email: silvio.tosatto@unipd.it

Funding information

European Union's Horizon 2020, Grant/Award Number: 823886 (H2020 MSCA-RISE "REFRACT"); European Cooperation in Science and Technology

Review Editor: Nir Ben-Tal

Abstract

Alpha-solenoids are a significant and diverse subset of structured tandem repeat proteins (STRPs) that are important in various domains of life. This review examines their structural and functional diversity and highlights their role in critical cellular processes such as signaling, apoptosis, and transcriptional regulation. Alpha-solenoids can be classified into three geometric folds: low curvature, high curvature, and corkscrew, as well as eight subfolds: ankyrin repeats; Huntingtin, elongation factor 3, protein phosphatase 2A, and target of rapamycin; armadillo repeats; tetratricopeptide repeats; pentatricopeptide repeats; Pumilio repeats; transcription activator-like; and Sel-1 and Sel-1-like repeats. These subfolds represent distinct protein families with unique structural properties and functions, highlighting the versatility of alpha-solenoids. The review also discusses their association with disease, highlighting their potential as therapeutic targets and their role in protein design. Advances in state-of-the-art structure prediction methods provide new opportunities and challenges in the functional characterization and classification of this kind of fold, emphasizing the need for continued development of methods for their identification and proper data curation and deposition in the main databases.

KEYWORDS

alpha-solenoids, ankyrins, armadillo repeats, HEAT, structured tandem repeats, tandem repeat classification, TPR repeats

Paula Nazarena Arrias and Zarifa Osmanli contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

1 | INTRODUCTION

Tandem repeat proteins (TRPs) are defined by the presence of periodically repeated amino acid sequences (Andrade, Perez-Iratxeta, & Ponting, 2001; Heringa, 1998; Kajava, 2012) and are the second most abundant class of proteins involved in protein–protein binding after immunoglobulins (Li et al., 2006). They have been reported to be ubiquitous in genomes, with at least 14% of protein-coding sequences containing repeats (Marcotte et al., 1999). Recently, a specific subset of TRPs has been defined: “structured tandem repeat proteins” or STRPs, which are mainly characterized by their defined modular 3D structure, among other characteristics (Monzon et al., 2023). The conformational space of STRPs is dominated by elongated (class III) and closed structures (class IV), according to Kajava's classification (Kajava, 2012). Class III is mostly composed of solenoid structures characterized by repeating structured units of 5–40 residues, forming elongated folds in contrast to globular proteins. The so-called “alpha-solenoids,” mainly composed of alpha-helices, are the most common (Kajava, 2001; Kobe & Kajava, 2000) and have been intensively studied in the last two decades. Solenoidal folds, in general, can also be functionally characterized by their structural properties, such as curvature, handedness, and twist, which contribute to their hierarchical classification (Kobe & Kajava, 2000).

RepeatsDB (Paladin et al., 2021) is a comprehensive, manually curated database focused on the identification, classification, and analysis of STRPs. In its efforts to classify and describe alpha-solenoids, RepeatsDB categorizes them based on their class and topology as elongated repeats/alpha-solenoids, as well as on their fold structural properties, such as low and high curvature, and corkscrew. Moreover, alpha-solenoids are defined not only by their structural differences but also by the presence of different sequence motifs, making it possible to identify diverse subfolds (Figure 1). Here we review the main types of alpha-solenoidal STRPs subfolds, namely ankyrins (ANK); Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and target of rapamycin (TOR) (HEAT); armadillo (ARM); tetratricopeptide (TPR); pentatricopeptide (PPR); Pumilio (PUF) repeats; TAL effector repeat (TAL); and Sel-1 and Sel1-like repeats (SLRs). We provide a brief description of their structure and sequence patterns as well as functionally relevant examples for each type of repeat. We also highlight noteworthy cases of designed alpha-solenoid proteins that are not naturally occurring but are of particular interest to researchers.

2 | ALPHA-SOLENOIDS SUBFOLDS

The distribution of subfolds across the tree of life varies significantly in some particular cases (Figure 2 and Table 1). ANK and TPR are prominent across all domains but with varying proportions. Eukaryotes have a more diverse representation of subfolds, while Bacteria and Archaea are heavily dominated by TPR. Viruses show a strong preference for ANK, mostly due to their presence in poxviruses (Herbert et al., 2015). However, the presence of alpha-solenoid subfolds in viruses is highly depleted compared to the other kingdoms. On the contrary, Eukaryotes display the highest diversity in alpha-solenoid subfolds, which aligns with their complex cellular machinery and functions. The diversity of alpha-solenoids is also reflected in their subcellular localization and associated functions, including cell cycle control, cell signaling, apoptosis, transcriptional regulation, cellular scaffolding, development, and differentiation, among others (Li et al., 2006) (Table S1).

3 | ANKYRIN REPEATS

ANKs repeats are one of the most common protein sequence motifs and, as such, are involved in a myriad of key cellular processes (Figure 1; Tables 1 and S1). Each repeat has a length between 30 and 34 amino acids and presents a helix-turn-helix conformation (Li et al., 2006). Structurally, ANK regions consist of repeats of antiparallel alpha-helices stacked side by side and connected by beta-hairpin motifs, where the extended beta-sheets project away from the helical pairs almost at right angles to them, resulting in a characteristic L-shaped cross-section (Figure 1). Specific residues are preserved in key consensus positions, which are essential for a correct conformation: Gly in positions 4, 15, and 27 of the repeat unit, the Thr-Pro-Leu-His9 tetrapeptide (TPLH) motif (Figure 3). While the interactions between these conserved residues result in the non-globular conformation of the domain, these proteins remain functional both within the cell and in the extracellular space (Michaely, 2002; Sedgwick & Smerdon, 1999). Generally, the structure is mainly stabilized by intra- and inter-repeat hydrophobic interactions and hydrogen bonds (Li et al., 2006). Lastly, a group of six non-conserved residues on the surface of the protein drive peptide recognition, providing each protein with specific binding partners (Magliery & Regan, 2005).

Protein sequences of ANK repeats are highly divergent, with sequence identity falling below 20% in certain cases. Additionally, it has been shown that internal repeats are better conserved than those at the N- and

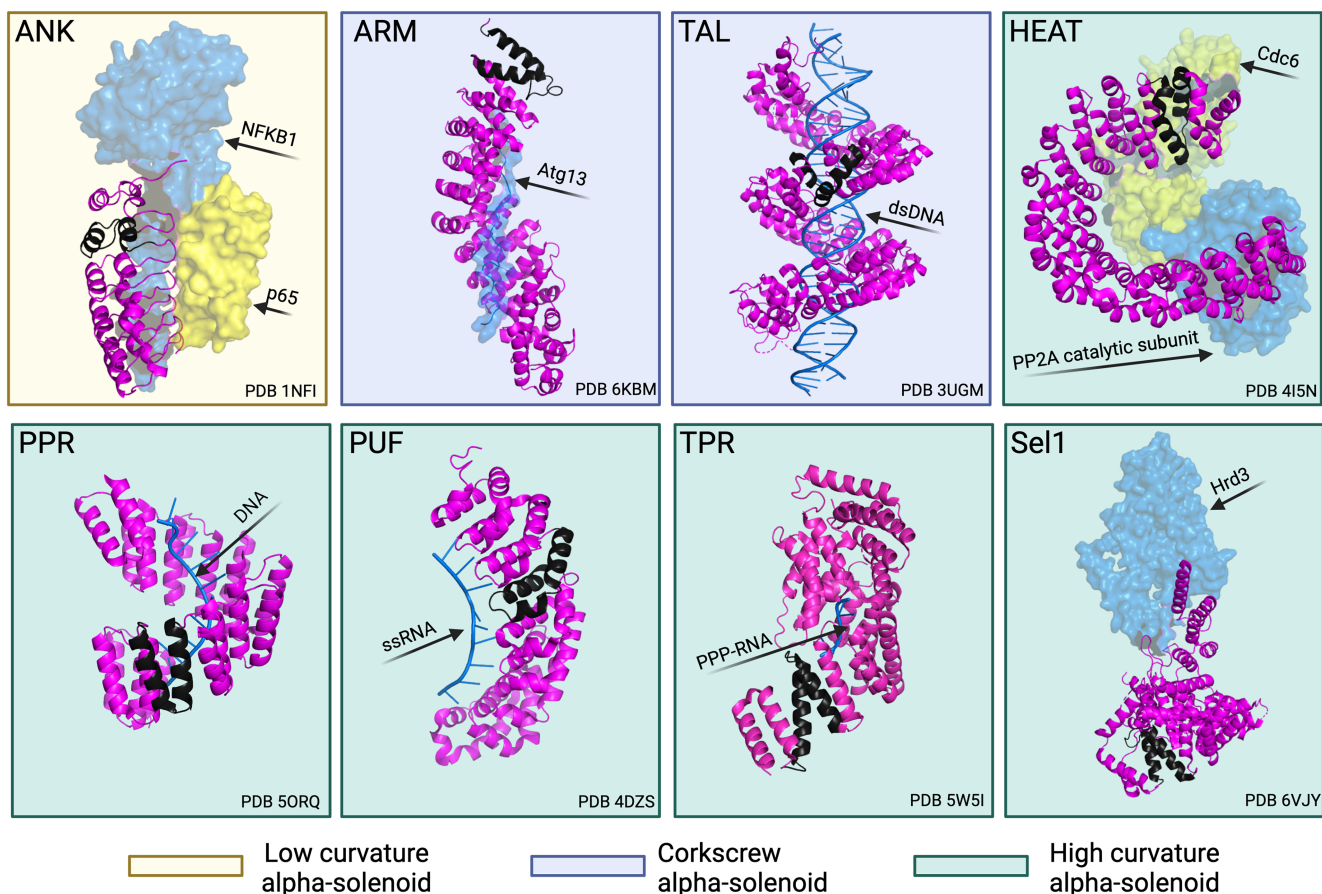


FIGURE 1 Representative structures of the main eight subfolds of alpha-solenoid proteins under three folds and their interaction with proteins/peptides/nucleic acids. Repeat-containing chain is in magenta color and a typical unit is highlighted in dark gray. The interacting proteins/peptides are shown as surfaces. ANK: *Homo sapiens* nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kappa-B) inhibitor alpha in complex with transcription factor p65 and NFKB1 (PDB: 1NFI), ARM: *Saccharomyces cerevisiae* Vac8 in complex with Atg13 (PDB: 6KBM), TAL: *Xanthomonas oryzae* TAL effector protein pthXo1 binds to dsDNA (PDB: 3UGM), HEAT: *H. sapiens* serine/threonine-protein phosphatase 2A in complex with Cdc6 and PP2A catalytic subunit (PDB: 4I5N), PPR: designed cPPR-Telo1 in complex with DNA (PDB: 5ORQ), PUF: *S. cerevisiae* Puf4p bound to ssRNA (PDB: 4DZS), TPR: *H. sapiens* IFIT1 with PPP-RNA (PDB: 5W5I), Sel1: *S. cerevisiae* Hrd1/Hrd3 monomer where Sel1-like Hrd1 interacts with Hrd3 (PDB: 6VJY).

C-terminal regions, with N-terminal repeats being the least conserved (Herbert et al., 2015). Multiple insertions and deletions occurring in specific repeating units often play a role in protein–protein interactions (PPIs) (Utgés et al., 2021). These positions are typically enriched in highly frustrated interactions, with potential impact on the recognition process and folding dynamics (Sanches et al., 2022).

ANK are a family of scaffold proteins that mainly bind other proteins (Figure 1), making them central to several key cellular processes, including but not limited to, cell signaling, cytoskeleton integrity, inflammatory response, and different transport phenomena (Bennett & Chen, 2001; Mosavi et al., 2004; Sedgwick & Smerdon, 1999).

In mammals, there are three canonical ankyrin proteins: ankyrin-1 (or erythrocyte ankyrin, also known as ankyrin-R), ankyrin-2 (or brain ankyrin, also known

as ankyrin-B), and ankyrin-3 (or ankyrin-G), which are encoded by genes *ANK1*, *ANK2*, and *ANK3*, respectively. The canonical isoform of *Homo sapiens* ankyrin-1 has 23 repeats and is part of the ankyrin-1 complex, which is involved in the stability and shape of the erythrocyte membrane (Vallese et al., 2022), while the canonical isoform of ankyrin-2 has 24 repeats and is highly expressed in cardiac muscle, where it has an essential role in the localization and membrane stabilization of ion transporters and channels. Human ankyrin-3 has 23 repeats, and it is specifically localized to the neuromuscular junction and nodes of Ranvier (Kordeli et al., 1995) via the interaction with SCHIP1 (Papandréou et al., 2015).

Genetic alterations such as point mutations, deletions, and insertions in the genes that encode ANK-containing proteins result in dysfunctional proteins and subsequently lead to various pathologies, including

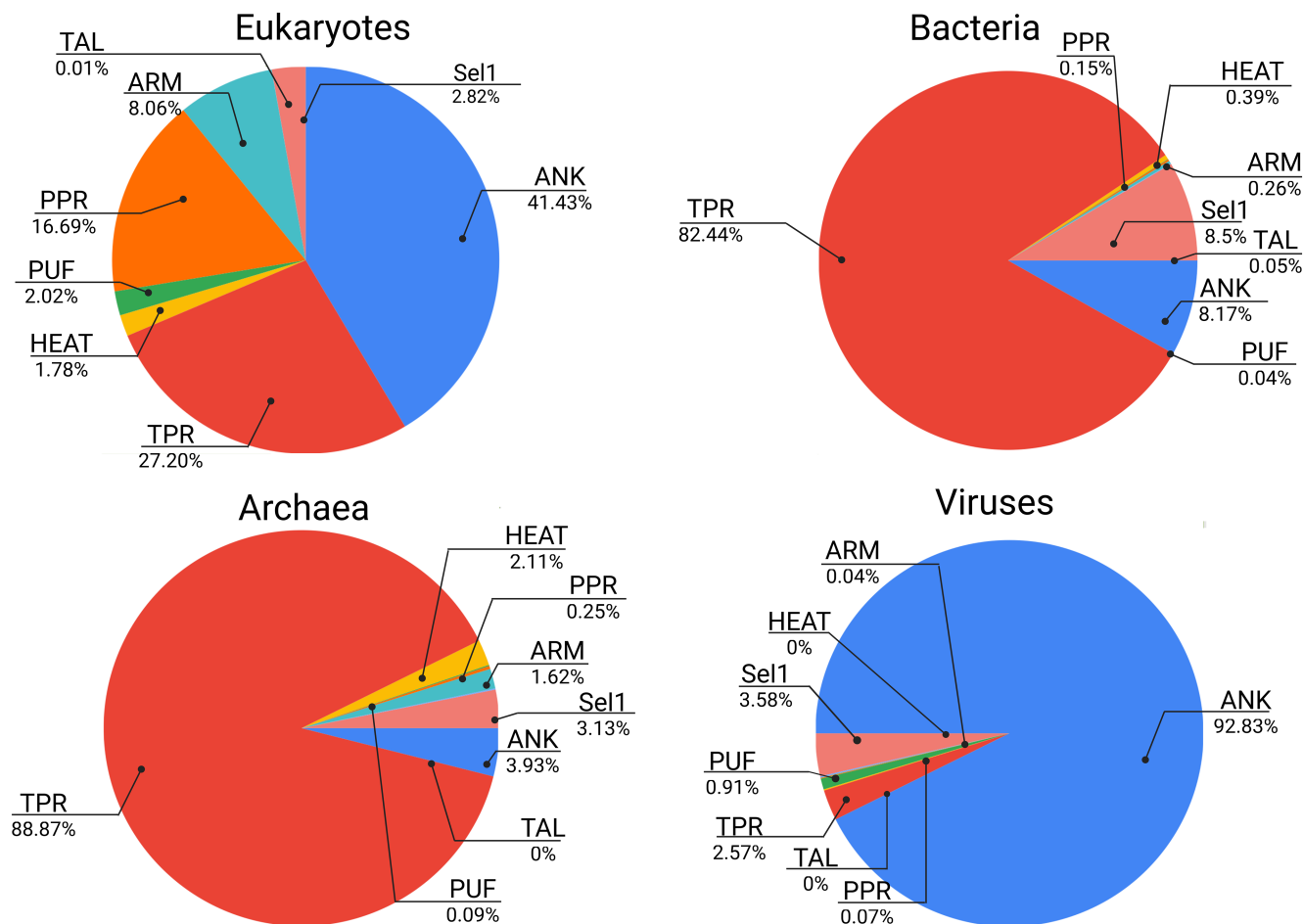


FIGURE 2 Proportion of alpha-solenoid proteins by subfolds in the tree of life. Proportions were obtained by dividing the total number of alpha-solenoid proteins of each subfold by the total number of all alpha-solenoid proteins. ANK, ankyrin repeats; ARM, armadillo repeats; HEAT, Huntingtin; Sel1, Sel-1-like repeats; PPR, pentatricopeptide repeats; PUF, Pumilio repeats; TAL, transcription activator-like; TPR, tetratricopeptide repeats.

cancer and neurological disorder. Some of these mutations are located within repetitive regions of ANK-containing proteins. For example, familial melanoma-associated mutations within ANK repeat regions in p16 interfere with the biological function of this molecule. Notably, residues D74 and R87 directly interact with cyclin-dependent kinases and are conserved among INK4 family members and p16 orthologs (Li et al., 2006; Mosavi et al., 2004).

As stated before, ANK proteins are present in a wide range of organisms. For example, ankyrin repeat protein 2A (AKR2A) of *Arabidopsis thaliana*, which has four ankyrin repeats at the C-terminal end, is involved in disease resistance and antioxidant metabolism (Yan et al., 2002), and also displays chaperone activities toward membrane proteins located in the chloroplast outer envelope membrane (OEM) (Bae et al., 2008). AKR2A is also involved in the flowering process via the interaction with the FLOWERING LOCUS T (FT) protein, as loss-

of-function mutant plants for AKR2A present a delayed flowering phenotype (Tang et al., 2022). In *Triticum aestivum* (common wheat), the Tiller Number1 (TN1) gene encodes an ankyrin protein with a transmembrane domain. Mutations in a single residue of the third conserved ankyrin repeat in TN1 cause a decrease in tillers, which are key elements of the plant that directly contribute to crop yield (Dong et al., 2023).

Some ANK proteins are key for the pathogenesis of intracellular bacteria. For instance, *Legionella pneumophila* modulates the host's processes in order to establish the *Legionella*-containing vacuole (LCV), where it replicates by employing multiple effector proteins. The protein RomA is one of these effectors; it contains six ANK repeats and is required for histone H3 binding and subsequent modification of the host's chromatin (Rolando et al., 2023). Another effector is the ankyrin-containing GTPase LegA15, which hijacks the host vesicular transport by promoting Golgi fragmentation and co-opting

TABLE 1 Subfolds of the alpha-solenoid proteins and their distribution across the tree of life (UniProtKB accessed on June 17, 2024). The highest absolute number of proteins for each alpha-solenoid subfold among the tree of life is underlined.

Abbreviation	Name	Number of repeat proteins (UniProtKB)				
		Eukaryotes	Bacteria	Archaea	Viruses	Total
ANK	Ankyrin	<u>565,710</u>	71,112	786	6853	637,608
ARM	Armadillo	<u>110,109</u>	2261	323	3	112,718
TAL	TAL effector repeat	2	<u>407</u>	N/A	N/A	409
HEAT	HEAT	<u>24,349</u>	3378	422	0	28,202
PPR	Pentatricopeptide	<u>227,968</u>	1288	49	5	229,305
PUF	Pumilio	<u>27,534</u>	382	18	67	28,015
TPR	Tetratricopeptide	371,396	<u>717,207</u>	17,776	190	1,114,754
Sel1	Sel1-like repeats	38,494	<u>73,919</u>	626	264	114,318

p115 of the host (Chen et al., 2022). This induction of the Golgi apparatus fragmentation is not exclusive to *Legionella* ankyrin proteins, as the ANK protein RARP2 of *Rickettsia rickettsii* has also been found to induce selective fragmentation of the trans-Golgi network (Aistleitner et al., 2020). Another ANK protein from *Rickettsia* spp., RARP1, promotes entry, intracellular growth, and host-cell invasion, presumably by interaction with multiple binding partners (Sanderlin et al., 2022).

Although ANK repeats are the most prevalent among eukaryotes, they are also predominant in viruses (Figure 2 and Table 1). The most notable instances of ANK proteins existing in viral genomes belong to poxviruses (Odon et al., 2018). Analyses show that the sequences of viral ANKs are divergent and support extensive gene duplication in chordopoxvirus (Sonnberg et al., 2011). For example, the 282E4 strain of the avipoxvirus, which infects several species of birds (Bolte et al., 1999) and has been extensively used for the development of recombinant vaccines, codes for 57 ankyrin repeat proteins (Deng et al., 2023). The function of these proteins has not been experimentally confirmed, but Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analysis suggest that they might play a role in the transcription of viral genes (Deng et al., 2023).

The oncolytic virus myxoma virus (MV), which causes myxomatosis in rabbits, can infect and kill several human cancer cell lines by a mechanism that involves the protein M-T5, which has seven ANK repeats (Wang et al., 2006). M-T5 can bind and activate cellular Akt, and the susceptibility of the cell lines to MV infection is directly related to the level of Akt activation (Wang et al., 2006; Werden et al., 2007).

Although ANK proteins normally bind other proteins, it has been reported that they can also bind lipids. For example, the protein K1 of vaccinia virus, which is involved in the negative regulation of the NF- κ B

signaling pathway, also promotes infection by binding and deforming the host's membrane through the basic amino acids of its ANK region (Kitamata et al., 2020).

ANK proteins are also found, albeit to a lesser extent, in archaea. For example, a thermophilic archaeal ANK protein, tANK of *Thermoplasma volcanium*, which has five repeats, has the particularity that presents three folding states (Löw et al., 2008).

In recent years, designed ankyrin repeats proteins (DARPin) have gained relevance (Forrer et al., 2003) because they constitute smaller, easier to produce, and high affinity binders that can be used instead of antibodies in multiple applications.

DARPin maintain the primary structure of ANK by retaining consensus sequences from the ANK family, preserving the most frequent residues for each canonical position but introducing changes into specific amino acids (Parra et al., 2015; Plückthun, 2015). These amino acids are located on the surface and involved in the binding interface, and thus are able to recognize a desired target with high specificity and affinity (Walser et al., 2022). Moreover, the introduction of a 19-amino acid flexible loop into the β -turn of DARPin allows for the selection of even higher affinity binding molecules (Schilling et al., 2014). Theoretical combinatorial libraries of DARPin can range from 5.2×10^{15} to 3.8×10^{23} for those containing two and three repeat modules (Plückthun, 2015), respectively. Highly diverse libraries can be produced easily and inexpensively by overlap-extension polymerase chain reaction (OE-PCR) using degenerate oligonucleotides (Morselli et al., 2024).

The broad applicability of DARPin is highlighted by the extensive amount of published work that employs them. One example of the use of DARPin for immunoncology purposes is that of MP0250, a drug candidate that binds and inhibits vascular endothelial growth factor (VEGF) and hepatocyte growth factor (HGF), thus



FIGURE 3 Sequence logos of main subfolds of alpha-solenoid proteins. One representative Pfam family per each subfold is considered. The sequence logos were made by the Skylign tool (Wheeler et al., 2014) with default parameters using seed Pfam alignment as an input. Occupancy is the probability of observing a letter at position k , and every amino acid has a unique color by default parameters. The length of the sequence logos is modified to fit into one repeat unit length of the specific Pfam superfamily. ANK, ankyrin repeats; ARM, armadillo repeats; HEAT, Huntingtin; Sel1, Sel-1-like repeats; PPR, pentatricopeptide repeats; PUF, Pumilio repeats; TAL, transcription activator-like; TPR, tetratricopeptide repeats.

altering the tumor microenvironment (Baird et al., 2021). Antiviral drugs have also been developed based on this technology and tested in clinical trials; Ensivibep, an

experimental therapy for CoronaVirus Disease of 2019 (COVID-19) disease, was designed to engage the three units of the spike protein trimer of Severe Acute

Respiratory Syndrome 2 (SARS-CoV-2) and inhibit ACE2 binding with high potency (Rothenberger et al., 2022). With the aim of improving these therapies, combined in silico and experimental studies have shown that certain residue positions are key to maintaining the general thermostability of these proteins, and DARPins introducing mutations in Asp17 present greater thermostability (Schilling et al., 2022).

4 | ARMADILLO REPEATS

The ARM repeats, which owe their name to *Drosophila melanogaster's* ARM protein where they were first identified (Nüsslein-Volhard & Wieschaus, 1980; Riggelman et al., 1989), are motifs predominantly implicated in PPIs such as those mediating intracellular signaling, cell-adhesion, and cytoskeletal regulation (Coates, 2003). A typical ARM repeat has a length of 42 residues and folds into three helices, denominated H1, H2, and H3, respectively. These helices are organized in a way in which H2 and H3 are packed antiparallel against each other and the shorter H1 is in a perpendicular position (Andrade, Petosa, et al., 2001) (Figures 1 and 3).

The ARM segment polarity protein of the fruit fly (UniprotKB acc. P18824), which is homologous to the highly evolutionary conserved β -catenin (Valenta et al., 2012), contains 12 prototypical repeats and a truncated 13th repeat. It is involved in signal transduction of the Wnt signaling pathway, which plays essential roles in embryonic development as well as in adherens junctions of the central nervous system and epidermis (Cavallo et al., 1997; Loureiro & Peifer, 1998).

Cell-adhesion is one of the well-known biological functions performed by ARM repeat-containing proteins. For example, in adherens junctions between cells, we find the junctional plaque protein plakoglobin. Plakoglobin, which in humans has 12 ARM repeats, is a key component of desmosomes and intermediate junctions and interacts with multiple proteins, including E-cadherin (Aktary et al., 2017). ARM repeats are also present in a multitude of ubiquitin-ligases, which are enzymes central to protein targeted-ubiquitination and degradation. For example, the protein Ufd4 of *Saccharomyces cerevisiae* has four ARM repeats, which are essential for Ufd4 ubiquitination of its substrates (Ju et al., 2007).

In some cases, a sole ARM-containing protein can participate in multiple biological processes. For example, Vac8 (UniProtKB acc. P39968) of *S. cerevisiae*, which has nine ARM repeats, is a key mediator for various cellular processes such as autophagy (Park et al., 2020), vacuole inheritance, and protein targeting (Wang et al., 2006), among others. Vac8 regulates these

processes by interaction via its ARM repeats with different associated proteins (Kim et al., 2023), like Nvj1 (Jeong et al., 2017), Atg13 (Park et al., 2020), and Vac17 (Tang et al., 2006).

ARM repeats are also ubiquitous in plants, in which they participate not only in the processes mentioned above but also in growth, development, and the stress response (Mandal et al., 2018). For instance, the protein MLKS2 of *Zea mays*, which interacts via its N-terminal ARM-repeat region with F-actin, is required for normal plant development and fertility (Gumber et al., 2019). In addition, both the ARM-containing region as well as the C-terminal non-ARM region of EFR3 OF PLANT 3 (EFOP3) and EFR3 OF PLANT 4 (EFOP4), are required for pollen development in *A. thaliana* (Chen et al., 2022).

Several reports have linked ARM proteins in plants to the response to different stresses. In rice, Sharma and colleagues showed that a subset of ARM genes is differentially expressed in response to abiotic stress, that is, salt, cold, and dehydration (Sharma et al., 2014). Moreover, in *Arabidopsis*, the protein AtPUB2 is induced under oxidative stress and may contribute to the ability of plants to tolerate this type of stress, although the specific mechanism remains to be determined (Saini et al., 2023).

Although ARM repeats are also detected in the proteomes of bacteria and archaea, their numbers are significantly lower than in eukaryotes (Jernigan & Bordenstein, 2014) (Figure 2), and there are not many reports about them and their associated functions. The structure of the hypothetical protein TON1937 of the hyperthermophilic archaea *Thermococcus onnurineus* NA1, reveals a superhelix fold composed of seven alpha-helical repeats, out of which only one, repeat 3, resembles a prototypical ARM (Jeong et al., 2013). In *Helicobacter pylori*, a pathogenic bacteria, the protein FliG, which contains several ARM repeats, forms an essential part of the flagellum. The ARM repeats of FliG are directly involved in the assembly of the MS-C ring and responsible for the superhelical arrangement of the inner part of it (Xue et al., 2018). Another example of a bacterial protein with an ARM region is the protein PgDPP III of *Porphyromonas gingivalis*, a human pathogen responsible for periodontitis. PgDPP II is an atypical dipeptidyl-peptidase (DPP) in which the ARM-repeat domain, which is absent in typical DPPs, contributes to its enzymatic activity and influences its interdomain dynamics (Hromić-Jahjefendić et al., 2017).

5 | TRANSCRIPTION ACTIVATOR-LIKE EFFECTORS

Transcription activator-like effectors are proteins secreted by plant pathogens belonging to the β - and γ -proteobacteria

classes that bind DNA in a sequence-specific way (Schornack et al., 2013) (Figure 1). The DNA-binding domain of these proteins is composed of almost identical, 34 residue-long repeats, each of which folds into a left-handed, two-helix bundle (Mak et al., 2012). The repeats associate with adjacent repeats to form a right-handed superhelix that wraps around DNA. Sequentially, the repeats vary in two amino acid positions, 12 and 13, named “repeat-variable diresidue” (RVD) (Boch & Bonas, 2010; Moscou & Bogdanove, 2009), which are responsible for the specific recognition of a DNA base (Figure 3).

Naturally, Transcription activator-like effectors (TALEs) are virulence factors that support bacterial colonization of the plants by exerting their transcription factor activity (Doyle et al., 2013). Mainly, they are found in *Xanthomonas* spp. bacteria, which are responsible for the bacterial leaf blight and leaf streak diseases, which have devastating effects on rice production (Boch & Bonas, 2010; Verdier et al., 2012). TAL effector-like proteins have also been found in fungal endosymbiotic bacteria and are key factors for the stress tolerance of the phytopathogenic fungal host (Carter et al., 2020) as well as for the maintenance of the endosymbiotic relationship (Richter et al., 2023) (Figure 2 and Table 1).

In addition to their importance for agriculture in terms of improving plant disease resistance, TALEs became interesting to researchers as genetic engineering tools. They have been employed to activate/repress expression of desired genes in different organisms, such as *Solanum lycopersicum*, *A. thaliana*, and also human cells (Cong et al., 2012; Geißler et al., 2011; Morbitzer et al., 2010), and they have also been fused to nuclease domains like FokI to generate TALEN restriction enzymes. TALEN enzymes can be used to edit genomes by introducing double-strand breaks in DNA. Although TALEN are not as popular as CRISPR, they have been utilized in vitro for the correction of mutations associated with different diseases, such as sickle cell disease (Moiani et al., 2024).

6 | HEAT REPEATS

HEAT repeats owe their name to the proteins in which they were first identified; HEAT, eukaryotic translation elongation factor 3 (eEF3), regulatory A subunit of PP2A, and mechanistic target of rapamycin (mTOR) (Andrade & Bork, 1995; Andrade, Petosa, et al., 2001). One HEAT repeat is approximately 40 amino acids long, and its canonical structure consists of two helices (A and B) that form a helical hairpin (Figures 1 and 3). HEAT repeats stack against each other to form an elongated superhelix, in which the B helices form the concave

surface important for interactions with other proteins (Cingolani et al., 1999). HEATs resemble ARM repeats; the helix A of HEAT corresponds to helices H1 and H2 of ARM, and helix B to ARM's helix H3. The similarity between these two types of repeats is suggested to be a consequence of convergent evolution (Andrade, Petosa, et al., 2001; Jernigan & Bordenstein, 2014). Later, in a separate study, analysis of mathematically defined repetitive short fragments of the protein sequences supported the structural homology between those two subfolds (Turjanski et al., 2016). As is also the case for ARM repeats, HEAT repeats are found in eukaryotic organisms as well as in bacteria and archaea. However, no viral proteins have been annotated as having HEAT repeats (Figure 2 and Table 1).

HEAT repeats are involved in signaling, cellular transport, and protein synthesis, among other processes. The different interactions that this type of repeat can carry out are a direct consequence of their structural plasticity, which allows HEAT repeats to adopt multiple tertiary conformations. The highly flexible and elastic conformational changes that occur in these repetitions are due to their characteristic hydrophobic core, which allows them to bind to specific partners as well as respond to intracellular environments (Kappel et al., 2010; Yoshimura & Hirano, 2016).

Huntingtin (HTT) is a 348-kDa protein that is conserved from flies to mammals. Mutant variants (mHTT) generated by expansion of CAG repeats in exon 1 of the HTT gene located on chromosome 4 are the cause of Huntington's disease (HD) (Potkin & Potkin, 2018; Ross et al., 2014). mHTT protein can form β sheets and abnormal protein aggregates, and is involved in neurotoxicity and, subsequently, brain atrophy (Harper et al., 2005; Macdonald, 1993). A myriad of different interactors have been proposed for HTT. HAP40 has been experimentally validated as a HTT interactor, and this interaction has been shown to be conserved throughout evolution, although the specific function of the interaction remains unclear (Seefelder et al., 2020). Structural studies of the HTT-HAP40 complex show that the HEAT domains of HTT wrap around HAP40, and biophysical characterization shows that the HAP40-bound HTT is less prone to aggregation and presents greater structural stability, suggesting that HAP40 may stabilize HTT (Guo et al., 2018). The involvement of the HEAT repeats of HTT in the pathogenesis of HD, if any, has not been described. A recent study employing chimeric constructs of both wild-type and pathogenic lengths of the exon 1 region of HTT and the HEAT region of PR65/A showed that both lengths of the exon 1 polyQ tract disturb the structure of PR65/A HEAT domain to the same extent. These results suggest that the interaction between the exon 1 polyQ

tract and the HEAT repeats of HTT might play a role in both the physiological function of HTT as well as in the pathogenesis of HD, but further studies remain to be conducted (Zhang et al., 2023).

Eukaryotic eEF3 is encoded by YEF3 gene and has been isolated in different fungal species such as *S. cerevisiae*, *Candida albicans*, *Candida guilliermondii*, and *Pneumocystis carinii*. eEF3 has ATPase and GTPase activities, in addition to being involved in EF-1 α -dependent binding of aminoacyl-tRNA by the ribosome (Belfield et al., 1995; Sandbaken et al., 1990).

In eukaryotic cells, PP2A holoenzyme consists of a heterodimeric core enzyme composed of a scaffolding protein called subunit A and a catalytic subunit called subunit C (Xu et al., 2006). Subunit A has 15 HEAT repeats that are arranged in a particular L-shaped fashion (Hemmings et al., 1990), which is necessary for its function: phosphorylation and dephosphorylation of signaling factors (Cho & Xu, 2007; Grinthal et al., 2010). Finally, the last protein responsible for HEAT's acronym is mTOR protein kinase. This protein is an essential component of two multiprotein complexes called MTORC1 and MTORC2, and contains 32 HEAT repeats (Chao & Avruch, 2019). This kinase is involved in both protein synthesis and autophagy, and it has been observed that deregulated mTOR signaling is directly related to the progression of cancer and diabetes (Saxton & Sabatini, 2017).

Tor1 kinase in *C. albicans*, has eight N-terminal HEAT repeats in its protein-protein interaction domain. This less conserved region is important for *C. albicans* response to oxidative stress, cellular respiration, and hyphal growth. Furthermore, this site of the Tor1 protein is a therapeutic target because small drug-type molecules that inhibit PPIs could be possible selective inhibitors of Tor1 kinase for fungi, thus being able to differentiate from human cells, a problem which is found in treatment with rapamycin (Qi et al., 2022).

HEATR1, a protein that contains a HEAT domain at C-terminus, has been a target of extensive studies in recent years due to its involvement in different types of cancers and its potential as a therapeutic target. For example, in oral squamous cell carcinoma (OSCC), HEATR1 acts as a binding factor for the Pontin/Reptin complex, which is overexpressed in tumor cells (Nakamura et al., 2021), while in pancreatic cancer HEATR1 inhibits the activity of the Nrf2 factor that participates in the protection of cells from oxidative stress (Zhou et al., 2020). In the case of hepatocellular carcinoma (HCC) HEATR1 is up-regulated through the IGF1-mTORC1-SP1 axis. It promotes the growth of HCC by enhancing ribosome biogenesis (Yang et al., 2023).

HEATs are also found in bacteria; for example, in *Myxococcus xanthus*, Cpc7 is a HEAT-containing protein

formed by 27 alpha-helices which interacts with single-domain response regulator (SD-RR) CheY7. This protein-protein interaction regulates development in response to stress (Darnell et al., 2014). Gram-positive bacteria such as *Bacillus cereus* and *Streptococcus mutans*, as well as Gram-negative bacteria such as *Pseudomonas fluorescens* have HEAT-like repeat (HLR) proteins, which are alkyl-DNA glycosylases involved in DNA repair. The structure of the members of this superfamily is formed by an N-terminal α -helical bundle (NTB) (α A'- α C) and five HLR motifs (α D- α E, α F- α G, α H- α I, α J- α K, and α L- α M) which make up a C-shaped structure (Shi et al., 2018). In *P. fluorescens*, AlkC excises 7-methylguanine, 3-methyladenine, 3-methylguanine, 3mC, and 1mA, and its residues Glu121 and Glu156, which belong to the HLR domain, are involved in this function. Although the structures of AlkC and AlkD are similar, the strategy to repair damaged DNA is different; AlkD only disturbs the helical axis and bases stacking to excise alkylated base, while AlkC opens the minor groove of the DNA in the injury (Alseth et al., 2006; Shi et al., 2018). In *S. mutans*, AlkD2 has a weak binding to DNA due to the absence of the α B helix in its structure. This protein is involved in purine metabolism, binding to insulin-like growth factor 2 (IGF2) messenger RNA (mRNA)-binding protein (IMP) through the loop that surrounds the α A and α C helices, and the N-terminal of the α G helix (Mullins, Shi, Kotsch, & Eichman, 2015; Mullins, Shi, Parsons, et al., 2015).

7 | PENTATRICOPEPTIDE REPEATS

PPR repeat proteins were discovered in the genome of *A. thaliana* (Aubourg et al., 2000), and they were initially thought to be unique to plants. However, they were soon found in multiple other organisms (Figure 2, Tables 1 and S1), although in very low numbers when compared to the hundreds that are found in terrestrial plants (O'Toole et al., 2008). Similarly to TPRs, which have been evolutionarily linked to (Barkan & Small, 2014), a canonical PPR is a 35 amino acids motif that forms a helix-turn-helix, which together with the adjacent repeats fold into a superhelix (Figures 1 and 3). PPR proteins can be divided into two classes: P-type proteins, which are composed mainly of canonical PPR motifs, and PLS-type proteins, which consist of triplets that include a canonical PPR motif, a longer motif denominated as L, and a shorter one known as S. In general, repeat unit length varies between 31 and 36 residues (Yin et al., 2013). The number of PPRs in a protein is variable and it can go from two PPRs to as many as 28 like in *A. thaliana* PGR3 protein (UniProtKB acc. Q9SZ52).

The main binding partner of PPRs is RNA and the binding process is done in a sequence-specific manner (Barkan & Small, 2014), which is in contrast to the few TPRs that are able to bind RNA (which do so in a sequence-independent manner). The way PPR proteins recognize their RNA partners is done modularly; amino acid positions 5 and 35 within each repeat are responsible for the recognition and interact directly with specific nucleotides (Yagi et al., 2013; Yin et al., 2013). RNA binding abilities of PPRs make them central to multiple processes regarding organellar RNA metabolism, such as transcription, RNA-processing (including 5'-tRNA and 3'-tRNA processing), polyadenylation, stability, splicing, translation, and editing. For instance, the protein PPR756 of *Oryza sativa* (rice) acts as an editing factor required for the editing of mitochondrial genes essential for development (Zhang et al., 2020). The structure of PPR756 has not yet been experimentally determined, but PPRs have been sequentially detected, and the AlphaFold2 (Jumper et al., 2021) model (accessible with the UniProtKB acc. Q2QTL4) predicts the characteristic α -solenoid that is expected for PPR-containing proteins (Jumper et al., 2021; Varadi et al., 2022). The editing function is not limited to mitochondrial PPRs modifying mtRNAs. For example, the product of the *GmPGL2* gene of *Glycine max* (soybean) participates in C-to-U RNA-editing in chloroplasts (Feng et al., 2021).

PPR proteins can also play a role in RNA stabilization. In *A. thaliana*, the protein PPR287 is crucial for chloroplast function, presumably via the stabilization of chloroplast ribosomal RNAs (Page et al., 1987) (rRNAs). The protozoan *Trypanosoma brucei*, which causes acute trypanosomiasis in humans, codes for multiple PPR proteins. Among these, six of them, TbPPR2 to TbPPR7, are key for mitochondrial rRNA stabilization (Pusnik et al., 2007), and another one, TbPPR9, selectively stabilizes the mRNAs of both cytochrome oxidase subunits 1 and 2 (COX1 and COX2, respectively) (Pusnik & Schneider, 2012). The protein CCM1 of *S. cerevisiae* also stabilizes mitochondrial rRNA by binding to a specific AU-rich motif of the ribosome small subunit (Piątkowski & Golik, 2020).

Another key process in which PPRs are involved is protein translation, particularly of organelle-encoded genes. The Pet309 protein of *S. cerevisiae* activates translation of COX1 by binding to the 5' untranslated region (UTR) of its mRNA (Manthey & McEwen, 1995; Zamudio-Ochoa et al., 2014). Its human ortholog LRPPRC (also known as Lrp130) plays a similar role, and mutations in the gene cause the French Canadian variant of Leigh syndrome, a neurodegenerative disorder (Cooper et al., 2006; Mootha et al., 2003; Xu et al., 2004). Recently, a study identified two PPR proteins in *A. thaliana* missannotated as adenylyl cyclases, which are

required for translation initiation specifically in mitochondria (Tran et al., 2023). These two proteins, denominated mTRAN1 (AT4G15640) and mTRAN2 (AT3G21465), are part of the small subunit of the ribosome. The authors demonstrated that mTRAN1 binds to specific AU-rich motifs in the 5'-UTRs of mitochondrial genes, giving insight into the mechanism by which plant mitoribosomes recognize mRNAs even though they lack bacterial-like Shine-Dalgarno sequences (Tran et al., 2023).

Group-II introns in plant organelles are introns that do not have the ability to self-splice and they require the assistance of proteins to do so (Lambowitz & Zimmerly, 2011). PPR-containing proteins can promote group-II intron splicing thanks to their RNA binding capability. For example, the protein OsPPR939 of *O. sativa* is required for the splicing of introns 1, 2, and 3 of the *nad5* transcript, and knockdown or deletion of OsPPR939 severely affects development (Zheng et al., 2021). Something similar happens in maize, where the protein PPR4 is a trans-splicing factor of *rsp12* (Schmitz-Linneweber et al., 2006) and PPR5 promotes *trnG* splicing (Beick et al., 2008) and *rpl16* splicing (Rojas et al., 2018).

An interesting and recently described PPR protein is OsCPPR1. This rice protein has 16 PPR motifs and is essential for tapetal plastid development via regulation of *OsGLK1* transcript levels, and what makes OsCPPR1 unique is the fact that it is the first cytoplasmic PPR to be reported (Zheng et al., 2022).

It is worth mentioning that the characteristics of PPR proteins, such as their modularity and sequence-specific RNA binding ability, have made them attractive as biotechnological tools. Specifically designed PPRs can be used for controlling the expression of specific genes in organelles (Rojas et al., 2019), controlling splicing (Yagi et al., 2022) and to edit RNA directly without the use of cofactors (Bernath-Levin et al., 2022).

8 | PUMILIO REPEATS

PUF repeats can be found in the evolutionary conserved PUF family of RNA-binding proteins. Typically, each PUF-containing protein has eight repeats of around 36 amino acids, although some proteins, such as Puf-A of *H. sapiens* and Nop9 of *S. cerevisiae* have 11 and Pum23 of *A. thaliana* has 10 repeats (Qiu et al., 2014; Tam et al., 2010; Zhang et al., 2016) (Figure 3). Each repeat folds into three alpha-helices, which align with the helices of the following repeat, forming a crescent shape (Edwards et al., 2001; Wang et al., 2001) (Figure 1). The repeats present a degenerated sequence pattern, but three amino acid positions in the middle are conserved (Wang et al., 2002). These conserved positions are important because

they make up the RNA-binding surface of the PUF proteins; each repeat interacts directly with a RNA base (Figure 3) via the 3-conserved residues in a sequence-specific way (Wang et al., 2002). However, PUFs can also bind non-cognate sequences (Gupta et al., 2008; Miller et al., 2008).

As post-transcriptional regulators, PUF proteins are involved in a plethora of biological functions. They are usually repressors of gene expression, normally by binding to the 3'-UTR of mRNAs and subsequently downregulating them, although mRNA stabilizing effects have also been reported (Archer et al., 2009; Kaye et al., 2009; Naudin et al., 2017). The specific mechanisms by which PUF proteins exert their role are diverse. For instance, PUF, one of the canonical and founding members of the PUF family, is involved in *D. melanogaster's* early embryos development by binding hunchback (hb) mRNA, producing its translational arrest via deadenylation (Barker et al., 1992; Murata & Wharton, 1995) (removal of the poly-A tail), and it also represses translation of eIF4E in the neuromuscular junctions of fruit fly larvae, modulating synaptic function (Menon et al., 2004).

Human PUF proteins PUM1 and PUM2 are important repressors that lower the levels of their target mRNAs. The mechanism can be both deadenylation-dependent, for example, by directly binding the CCR4-NOT deadenylase complex (Enwerem et al., 2021; Van Etten et al., 2012), and deadenylation-independent, by promoting accessibility to miRNAs (Galgano et al., 2008; Kedde et al., 2010). *H. sapiens* also possesses non-canonical PUF proteins; Puf-A (also known as PUM3) has 11 repeats and is highly expressed in primordial germ cells (Kuo et al., 2009). Although it can bind ds-RNA and DNA without sequence specificity (Qiu et al., 2014), its biological function is not clear. Its ortholog in yeast, Puf6, with which it shares less than 25% sequence identity but a similar structure, is involved in ribosome biogenesis (Gerhardy et al., 2021; Li et al., 2009). A recent work from Cho and collaborators (Cho & Xu, 2007) reported that Puf-A promotes non-small cell lung cancer (NSCLC) progression by the interaction with nucleophosmin (NPM1) in the nucleolus. Up-regulated Puf-A is also involved in breast cancer progression (Fan et al., 2013). This involvement in cancer is not unheard for PUF proteins; canonical PUM1 and PUM2 are altered in 17 types of cancer tissues (Smialek et al., 2021), have been shown to sustain cell growth in acute myeloid leukemia by promoting the expression of FOXP1 (Naudin et al., 2017) and the protein Nop9 has been suggested as a prognostic marker for renal cancer (Silva et al., 2022).

Proteins containing PUF repeats have also been identified and studied in fungi. For instance, *S. cerevisiae* Puf3p regulates downregulation of mitochondrial

biogenesis (García-Rodríguez et al., 2007; Saint-Georges et al., 2008), and the proteins Puf3 and Puf4 play multiple functions in *Schizosaccharomyces pombe*, such as protection from toxic iron concentrations by repression of Frp1 expression (Beaudoin et al., 2021) and regulation of PI4P5K by suppression of its3-1 (Satoh et al., 2023). Nop9, a nucleolar PUF with 10 repeats, is required for 18S rRNA synthesis by binding its central pseudoknot region and possibly regulating its folding (Wang & Ye, 2017).

As a probable consequence of gene duplication in plants, *A. thaliana* codes for 25 PUF proteins. Some of them are involved in stem cell maintenance and differentiation (Francischini & Quaggio, 2009), while others are required for normal growth and development (Abbasi et al., 2011; Zhang & Muench, 2015). Moreover, some are involved in responses to both abiotic and biotic stress, like APUM5 of *Arabidopsis* (Huh & Paek, 2014).

As it is the case with PPRs, given the ability of PUF proteins to regulate gene expression in a sequence-specific manner, designer PUF proteins are interesting to the scientific community as biotechnological tools. Different approaches can be employed to, for example, enhance and suppress translation of target mRNAs (Campbell et al., 2014; Choudhury et al., 2012). For instance, one could employ natural variants of PUF proteins and engineer them in order to repurpose their biological function (Bhat et al., 2019) or generate different repeats that recognize different RNA bases (Zhao et al., 2018; Zhou et al., 2021). Adamala et al. (2016) generated four different PUF modules, which they have called “Pumby” (PUF-based assembly), each of which recognizes specifically one RNA base and that can be combined with each other in order to bind desired target RNAs.

9 | TETRATRICOPEPTIDE REPEATS

Initially described in yeast (Hirano et al., 1990; Sikorski et al., 1990); but now found to be widespread across organisms. They are more prevalent in bacteria and archaea (Figure 2, Tables 1 and S1). Each TPR repeat is approximately 34 amino acids in length and folds into a helix-turn-helix motif (Figure 1), which packs against the adjacent repeats, forming solenoid structures. Although TPRs present a degenerate sequence pattern, a consensus sequence has been defined for a group of amino acids (D'Andrea, 2003; Lamb et al., 1995) (such as small or large hydrophobic). Certain TPRs have 42 residue repeat units and more than 90% of sequence identity found, like those identified in *Podospora anserina* (Lamb et al., 1995) (Figure 3). The number of TPRs per protein varies; for example, *H. sapiens* TPR repeats protein 21B (UniProtKB

acc. Q7Z4L5) and TPR repeat protein 28 (UniProtKB acc. Q96AY4) have 19 and 28 repeats, respectively.

TPRs are a highly diverse fold involved in PPIs and protein–peptide interactions for which several different binding modes have been described (Perez-Riba & Itzhaki, 2019). This diversity of TPRs, both in structure and binding modes, is reflected in their associated functions, as they have been implicated in a plethora of processes that require PPIs, such as cargo recognition by kinesins (Pernigo et al., 2013), cell division (Yu et al., 2003; Zhu et al., 2011), peroxisomal import (Harper et al., 2003), and protein translocation across membranes (Schlegel et al., 2007), to mention a few.

Among these processes, an interesting case is the role of TPR-containing proteins in the immune responses of different organisms. For instance, the interferon-induced proteins with tetratricopeptide repeats (IFITs), which belong to the interferon-stimulated genes (ISGs), play a critical role in the defense against viral pathogens in both humans and mice (Fensterl & Sen, 2011; Sen & Sarkar, 2007). Four IFIT proteins have been described in *H. sapiens* (IFIT1, IFIT2, IFIT3, and IFIT5) and three in *Mus musculus* (IFIT1, IFIT2, and IFIT3), all of which are composed mostly of TPR motifs. None of these proteins are expressed in basal conditions, but instead their levels increase rapidly as a response to type I interferons (IFN-I) (Daffis et al., 2007) or by signals produced after PRRs-mediated (pattern recognition receptors) recognition of pathogen-associated molecular patterns (PAMPs) (Sarkar & Sen, 2004). The mechanisms by which these proteins exert their antiviral functions are different. In the case of both IFIT1 and IFIT2 of *H. sapiens* and *M. musculus*, it has been shown that they can inhibit viral protein translation by binding eIF3 (eukaryotic initiation factor 3) and preventing the formation of the pre-initiation complex (Diamond & Farzan, 2013; Guo et al., 2000; Vladimer et al., 2014). Regarding human IFIT3, its interaction via one of its TPRs with the serine/threonine-protein kinase TBK1 enhances the IFN-I antiviral signaling pathway. This effect is attributed to IFIT3 acting as a “bridge” between TBK1 and the mitochondrial antiviral signaling protein (MAVS) (Liu et al., 2011). Interestingly, although IFIT5 has been shown to positively regulate NF- κ B signaling via its interaction with I κ B kinase (IKK) and the transforming growth factor- β -activated kinase 1 (TAK1) (Zheng et al., 2015), its antiviral activity is associated with its ability to specifically bind single-stranded 5'-PPP-RNA (Abbas et al., 2013; Feng et al., 2013) (non-capped RNAs that are specific to viruses). The RNA-binding ability of TPRs is not unique to IFIT5, as the product of the *bsr-k1* gene of *Oryza sativa* (rice) is a TPR-containing protein that binds mRNA of multiple genes (Zhou et al., 2018). This demonstrates that while TPRs

typically engage in PPIs, they are also capable of binding to nucleic acids (Figure 1).

TPR-containing proteins are also involved in the tolerance of plants to different types of stress. For example, in *S. lycopersicum* (tomato), the protein SIREC2 enhances the cold tolerance of plants by interaction with SIBCH1b (Zhang et al., 2023), while the proteins DJC31 and DJC62 are involved in the abiotic stress responses of *A. thaliana*, presumably through the interaction with HSP80.1 and HSP90.2 (Dittmer et al., 2021). TPR proteins in plants are also important for development; *Z. mays* SMALL KERNEL 11 (*smk11*) is essential for the assembly of cytochrome c oxidase complex (the mitochondrial respiratory chain's complex IV) via its interaction with other four proteins, and this function requires all TPRs, as truncated forms are not sufficient for correct assembly (Ren et al., 2023).

In contrast to their roles in the immune defense of eukaryotes, some bacterial-encoded TPR proteins play key roles in pathogenesis. In the opportunistic pathogen *P. aeruginosa*, the gene *pscG* encodes for a 115-residue TPR-containing chaperone, which is essential for the assembly of the type-III secretion system commonly known as the injectisome, and this chaperone activity is mediated by its TPR motifs (Plé et al., 2010). Other virulence-related functions of bacterial TPR proteins include immune evasion. The serine/threonine-protein kinase PknG of *Mycobacterium tuberculosis* inhibits phagosome-lysosome fusion, enabling the pathogen's survival inside macrophages (Walburger et al., 2004), while the protein BB0238 of *Borrelia burgdorferi*, the causative agent of Lyme disease, is required for various PPIs indispensable for evading the host's early immune responses (Foor et al., 2023), and specifically, its TPR motifs are essential for BB0238 stability in the environment of the host.

Other worth mentioning examples of functional bacterial TPR proteins are those of magnetotactic bacteria such as *Magnetospirillum gryphiswaldense* and *Magnetospirillum magneticum*. Both organisms have a specialized organelle called the “magnetosome” which allows their navigation along Earth's geomagnetic fields. The magnetosome biosynthesis is highly complex, but the MamA TPR protein is essential for its assembly (Lohße et al., 2014; Zeytuni et al., 2011). MamA contains five TPR repeats, and it interacts both with itself (Zeytuni et al., 2011) as well as with Mms6, a membrane-bound magnetosome protein (Nguyen et al., 2016) to form a scaffold inside the magnetosome. Although TPRs are also present in archaea (Kyrpides & Woese, 1998) not much is known about their functions in these organisms.

In the case of viruses, a TPR-containing protein of *Acanthamoeba polyphaga* mimivirus R856 (UniProtKB

acc. Q5UQQ7) has been reported to be involved in fiber-formation. The specific role of R856 and the importance of its TPR domain have not been established, but gene silencing experiments result in viruses with shorter and abnormal fibers, which suggest that R856 is either an essential component of fibers or a key factor in their biosynthesis (Sobhy et al., 2015).

Recent works have described a small-communication system in SPbeta phages that regulates the lysis-to-lysogeny transition (Erez et al., 2017; Gallego Del Sol et al., 2022; Guan et al., 2019; Stokar-Avihail et al., 2019). The system is composed of three genes, one of which codes for a transcription factor called AimR. AimR, which presents nine degenerated-TPRs that are directly responsible for binding another of the system's gene products, a short peptide called AimP. AimP binding impedes AimR's transcription factor activity, thus promoting lysogeny (Gallego Del Sol et al., 2022).

10 | SEL-1 AND SEL-1-LIKE REPEATS

The SLRs are related to TPRs but their consensus sequence, in which conserved Ala and Gly residues can be found, is different (Mittl & Schneider-Brachert, 2007) (Figures 1 and 3). First identified in the gene *sel-1* (suppressor-enhancer of *lin 1*) of *Caenorhabditis elegans* (Grant & Greenwald, 1996), SLRs have a length between 36 and 44 residues and fold into a helix-turn-helix motif, which together with the neighboring repeats form superhelical structures (Figure 3). The repeat number in SLR-proteins varies; for example, *H. sapiens'* eukaryotic elongation factor 2 kinase (eEF-2K) contains four SLRs, while *Escherichia coli's* secretory immunoglobulin A-binding protein (EsiB) contains 11.

In terms of function, SLRs are involved in multiple processes such as signal transduction (Ji et al., 2023; Mittl & Schneider-Brachert, 2007), endoplasmic reticulum protein quality control (Cattaneo et al., 2009; Jeong et al., 2017) assembly of mitochondrial respiratory chain complexes (Kozjak-Pavlovic et al., 2014) and activation of the integrated stress response (Fessler et al., 2020; Guo et al., 2020), among others. In particular, mutations in SEL1L have been linked to progressive early-onset cerebellar ataxia in dogs (Kyöstiä et al., 2012) and to neurodevelopmental disorders in humans (Weis et al., 2024).

SLRs have also been identified in bacteria and have been implicated in several processes such as flagellar biogenesis (Sakuma et al., 2019), antibiotic resistance (Lüthy et al., 2002), and alginate biosynthesis (Gheorghita et al., 2022), among others. SLR-proteins are also important for pathogenesis, for instance, EsiB, which contains 11 SLRs

contributes to *E. coli* immune evasion by impairing neutrophil activation (Pastorello et al., 2013), while DipA of *Francisella tularensis* is a virulence factor required for intracellular replication (Chong et al., 2013). Moreover, the protein LpnE of *L. pneumophila* plays an important role in host-cell invasion, phagosome acidification, and the prevention of the LCV fusion to the lysosome (Voth et al., 2019).

11 | CONCLUDING REMARKS

Alpha-solenoids represent a significant and diverse subset of STRPs. This review highlights the remarkable structural and functional diversity of alpha-solenoids, which are crucial across various domains of life. Their essential roles in fundamental cellular processes, such as signaling, apoptosis, and transcriptional regulation, underscore their biological importance and evolutionary relevance.

We have classified alpha-solenoids into three main folds based on their geometric properties: low and high curvature, and corkscrew. Additionally, we identified eight main subfolds: ANK, HEAT, ARM, TPR, PPR, PUF, TAL, and SLRs. Each subfold showcases unique structural properties and functions, reflecting the versatile nature of alpha-solenoids. This classification aligns with the evolutionary ensembles of naturally occurring proteins and provides a deeper understanding of how sequence motifs and structural configurations contribute to the specific functions of these proteins.

The association of alpha-solenoids with various diseases further emphasizes their significance as potential targets for therapeutic research (Llabrés et al., 2020). The increasing availability of accurate structural protein models from advanced machine learning methods, such as AlphaFold (Varadi et al., 2022) and ESMFold (Lin et al., 2023), presents new opportunities and challenges for the large-scale application of STRP structure-based predictors (Do Viet et al., 2015; Hirsh et al., 2018; Mozaffari et al., 2024; Qiu et al., 2024). A key resource in this field is RepeatsDB, which offers comprehensive and curated information on alpha-solenoid proteins and other STRPs. RepeatsDB serves as a reference repository for researchers, providing detailed data on the structural classification and insights that facilitate further study on STRPs' potential in therapeutic interventions and protein design.

AUTHOR CONTRIBUTIONS

Paula Nazarena Arrías: Writing – original draft; writing – review and editing; methodology; visualization; data curation; formal analysis. **Zarifa Osmanli:**

Writing – original draft; writing – review and editing; visualization; methodology; formal analysis; data curation. **Estefanía Peralta**: Data curation. **Patricio Manuel Chinestrada**: Data curation. **Alexander Miguel Monzon**: Writing – review and editing; project administration; supervision. **Silvio C. E. Tosatto**: Writing – review and editing; project administration; resources; supervision.

ACKNOWLEDGMENTS

This work was supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 823886 (H2020 MSCA-RISE “REFRACT”) and based upon work from COST Action ML4NGP, CA21160, supported by COST (European Cooperation in Science and Technology).

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data used in this review is available in publicly accessible repositories.

ORCID

Paula Nazarena Arrías  <https://orcid.org/0000-0001-5668-768X>

Zarifa Osmanli  <https://orcid.org/0000-0003-2460-1989>

Alexander Miguel Monzon  <https://orcid.org/0000-0003-0362-8218>

Silvio C. E. Tosatto  <https://orcid.org/0000-0003-4525-7793>

REFERENCES

- Abbas YM, Pichlmair A, Górna MW, Superti-Furga G, Nagar B. Structural basis for viral 5'-PPP-RNA recognition by human IFIT proteins. *Nature*. 2013;494:60–4.
- Abbasi N, Park Y-I, Choi S-B. Pumilio Puf domain RNA-binding proteins in Arabidopsis. *Plant Signal Behav*. 2011;6:364–8.
- Adamala KP, Martin-Alarcon DA, Boyden ES. Programmable RNA-binding protein composed of repeats of a single modular unit. *Proc Natl Acad Sci U S A*. 2016;113:E2579–88. <https://doi.org/10.1073/pnas.1519368113>
- Aistleitner K, Clark T, Dooley C, Hackstadt T. Selective fragmentation of the trans-Golgi apparatus by *Rickettsia rickettsii*. *PLoS Pathog*. 2020;16:e1008582.
- Aktary Z, Alaei M, Pasdar M. Beyond cell-cell adhesion: plakoglobin and the regulation of tumorigenesis and metastasis. *Oncotarget*. 2017;8:32270–91.
- Alseth I, Rognes T, Lindbäck T, Solberg I, Robertsen K, Kristiansen KI, et al. A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol Microbiol*. 2006;59:1602–9.
- Andrade MA, Bork P. HEAT repeats in the Huntington's disease protein. *Nat Genet*. 1995;11:115–6.
- Andrade MA, Perez-Iratxeta C, Ponting CP. Protein repeats: structures, functions, and evolution. *J Struct Biol*. 2001;134:117–31.
- Andrade MA, Petosa C, O'Donoghue SI, Müller CW, Bork P. Comparison of ARM and HEAT protein repeats. *J Mol Biol*. 2001;309:1–18.
- Archer SK, Luu V-D, De Queiroz RA, Brems S, Clayton C. Trypanosoma brucei PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. *PLoS Pathog*. 2009;5:e1000565.
- Aubourg S, Boudet N, Kreis M, Lecharny A. In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol Biol*. 2000;42:603–13.
- Bae W, Lee YJ, Kim DH, Lee J, Kim S, Sohn EJ, et al. AKR2A-mediated import of chloroplast outer membrane proteins is essential for chloroplast biogenesis. *Nat Cell Biol*. 2008;10:220–7.
- Baird RD, Linossi C, Middleton M, Lord S, Harris A, Rodón J, et al. First-in-human phase I study of MP0250, a first-in-class DARP-in drug candidate targeting VEGF and HGF, in patients with advanced solid tumors. *J Clin Oncol*. 2021;39:145–54.
- Barkan A, Small I. Pentatricopeptide repeat proteins in plants. *Annu Rev Plant Biol*. 2014;65:415–42.
- Barker DD, Wang C, Moore J, Dickinson LK, Lehmann R. Pumilio is essential for function but not for distribution of the drosophila abdominal determinant Nanos. *Genes Dev*. 1992;6:2312–26.
- Beaudoin J, Normant V, Brault A, Henry DJ, Bachand F, Massé É, et al. Fission yeast RNA-binding proteins Puf2 and Puf4 are involved in repression of ferrireductase Frp1 expression in response to iron. *Mol Microbiol*. 2021;116:1361–77.
- Beick S, Schmitz-Linneweber C, Williams-Carrier R, Jensen B, Barkan A. The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol Cell Biol*. 2008;28:5337–47.
- Belfield GP, Ross-Smith NJ, Tuite MF. Translation elongation factor-3 (EF-3): an evolving eukaryotic ribosomal protein? *J Mol Evol*. 1995;41:376–87.
- Bennett V, Chen L. Ankyrins and cellular targeting of diverse membrane proteins to physiological sites. *Curr Opin Cell Biol*. 2001;13:61–7.
- Bernath-Levin K, Schmidberger J, Honkanen S, Gutmann B, Sun YK, Pullakhandam A, et al. Cofactor-independent RNA editing by a synthetic S-type PPR protein. *Synth Biol*. 2022;7:ysab034.
- Bhat VD, McCann KL, Wang Y, Fonseca DR, Shukla T, Alexander JC, et al. Engineering a conserved RNA regulatory protein repurposes its biological function in vivo. *Elife*. 2019;8:e43788.
- Boch J, Bonas U. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol*. 2010;48:419–36.
- Bolte AL, Meurer J, Kaleta EF. Avian host spectrum of avipoxviruses. *Avian Pathol*. 1999;28:415–32.
- Campbell ZT, Valley CT, Wickens M. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat Struct Mol Biol*. 2014;21:732–8.
- Carter ME, Carpenter SCD, Dubrow ZE, Sabol MR, Rinaldi FC, Lastovetsky OA, et al. A TAL effector-like protein of an endofungal bacterium increases the stress tolerance and alters the transcriptome of the host. *Proc Natl Acad Sci U S A*. 2020;117:17122–9.

- Cattaneo M, Lotti LV, Martino S, Cardano M, Orlandi R, Mariani-Costantini R, et al. Functional characterization of two secreted SEL1L isoforms capable of exporting unassembled substrate. *J Biol Chem*. 2009;284:11405–15.
- Cavallo R, Rubenstein D, Peifer M. Armadillo and dTCF: a marriage made in the nucleus. *Curr Opin Genet Dev*. 1997;7:459–66.
- Chao LH, Avruch J. Cryo-EM insight into the structure of mTOR complex 1 and its interactions with Rheb and substrates. *F1000Res*. 2019;8:14.
- Chen T-T, Lin Y, Zhang S, Liu S, Song L, Zhong W, et al. Atypical *Legionella* GTPase effector hijacks host vesicular transport factor p115 to regulate host lipid droplet. *Sci Adv*. 2022;8:eadd7945.
- Cho US, Xu W. Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. *Nature*. 2007;445:53–7.
- Chong A, Child R, Wehrly TD, Rockx-Brouwer D, Qin A, Mann BJ, et al. Structure-function analysis of DipA, a *Francisella tularensis* virulence factor required for intracellular replication. *PLoS One*. 2013;8:e67965.
- Choudhury R, Tsai YS, Dominguez D, Wang Y, Wang Z. Engineering RNA endonucleases with customized sequence specificities. *Nat Commun*. 2012;3:1147.
- Cingolani G, Petosa C, Weis K, Müller CW. Structure of importin- β bound to the IBB domain of importin- α . *Nature*. 1999;399:221–9.
- Coates J. Armadillo repeat proteins: beyond the animal kingdom. *Trends Cell Biol*. 2003;13:463–71.
- Cong L, Zhou R, Kuo Y, Cunniff M, Zhang F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun*. 2012;3:968.
- Cooper MP, Qu L, Rohas LM, Lin J, Yang W, Erdjument-Bromage H, et al. Defects in energy homeostasis in Leigh syndrome French Canadian variant through PGC-1 α /LRP130 complex. *Genes Dev*. 2006;20:2996–3009.
- Daffis S, Samuel MA, Keller BC, Gale M, Diamond MS. Cell-specific IRF-3 responses protect against West Nile virus infection by interferon-dependent and -independent mechanisms Seb G, editor. *PLoS Pathog*. 2007;3:e106.
- D'Andrea L. TPR proteins: the versatile helix. *Trends Biochem Sci*. 2003;28:655–62.
- Darnell CL, Wilson JM, Tiwari N, Fuentes EJ, Kirby JR. Chemosensory regulation of a HEAT-repeat protein couples aggregation and sporulation in *Myxococcus xanthus*. *J Bacteriol*. 2014;196:3160–8.
- Deng L, Liu C, Li L, Hao P, Wang M, Jin N, et al. Genomic characteristics of an avipoxvirus 282E4 strain. *Virus Res*. 2023;336:199218.
- Diamond MS, Farzan M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat Rev Immunol*. 2013;13:46–57.
- Dittmer S, Kleine T, Schwenkert S. The TPR- and J-domain-containing proteins DJC31 and DJC62 are involved in abiotic stress responses in *Arabidopsis thaliana*. *J Cell Sci*. 2021;134:jcs259032.
- Do Viet P, Roche DB, Kajava AV. TAPO: a combined method for the identification of tandem repeats in protein structures. *FEBS Lett*. 2015;589:2611–9.
- Dong C, Zhang L, Zhang Q, Yang Y, Li D, Xie Z, et al. Tiller Number1 encodes an ankyrin repeat protein that controls tillering in bread wheat. *Nat Commun*. 2023;14:836.
- Doyle EL, Stoddard BL, Voytas DF, Bogdanove AJ. TAL effectors: highly adaptable phytobacterial virulence factors and readily engineered DNA-targeting proteins. *Trends Cell Biol*. 2013;23:390–8.
- Edwards TA, Pyle SE, Wharton RP, Aggarwal AK. Structure of Pumilio reveals similarity between RNA and peptide binding motifs. *Cell*. 2001;105:281–9.
- Enwerem III, Elrod ND, Chang C-T, Lin A, Ji P, Bohn JA, et al. Human Pumilio proteins directly bind the CCR4-NOT deadenylase complex to regulate the transcriptome. *RNA*. 2021;27:445–64.
- Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, et al. Communication between viruses guides lysis-lysogeny decisions. *Nature*. 2017;541:488–93.
- Fan C-C, Lee L-Y, Yu M-Y, Tzen C-Y, Chou C, Chang M-S. Upregulated hPuf-A promotes breast cancer tumorigenesis. *Tumor Biol*. 2013;34:2557–64.
- Feng X, Yang S, Zhang Y, Zhiyuan C, Tang K, Li G, et al. GmPGL2, encoding a pentatricopeptide repeat protein, is essential for chloroplast RNA editing and biogenesis in soybean. *Front Plant Sci*. 2021;12:690973.
- Feng F, Yuan L, Wang YE, Crowley C, Lv Z, Li J, et al. Crystal structure and nucleotide selectivity of human IFIT5/ISG58. *Cell Res*. 2013;23:1055–8.
- Fensterl V, Sen GC. The *ISG56/IFIT1* gene family. *J Interferon Cytokine Res*. 2011;31:71–8.
- Fessler E, Eckl E-M, Schmitt S, Mancilla IA, Meyer-Bender MF, Hanf M, et al. A pathway coordinated by DELE1 relays mitochondrial stress to the cytosol. *Nature*. 2020;579:433–7.
- Foor SD, Brangulis K, Shakya AK, Rana VS, Bista S, Kitsou C, et al. A unique borrelial protein facilitates microbial immune evasion. *MBio*. 2023;14:e0213523.
- Forrer P, Stumpp MT, Binz HK, Plückthun A. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett*. 2003;539:2–6.
- Francischini CW, Quaggio RB. Molecular characterization of *Arabidopsis thaliana* PUF proteins – binding specificity and target candidates. *FEBS J*. 2009;276:5456–70.
- Galgano A, Forrer M, Jaskiewicz L, Kanitz A, Zavolan M, Gerber AP. Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLoS One*. 2008;3:e3164.
- Gallego Del Sol F, Quiles-Puchalt N, Brady A, Penadés JR, Marina A. Insights into the mechanism of action of the arbitrium communication system in SPbeta phages. *Nat Commun*. 2022;13:3627.
- García-Rodríguez LJ, Gay AC, Pon LA. Puf3p, a Pumilio family RNA binding protein, localizes to mitochondria and regulates mitochondrial biogenesis and motility in budding yeast. *J Cell Biol*. 2007;176:197–207.
- Geißler R, Scholze H, Hahn S, Streubel J, Bonas U, Behrens S-E, et al. Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One*. 2011;6:e19509.
- Gerhardy S, Oborská-Oplová M, Gillet L, Börner R, Van Nues R, Leitner A, et al. Puf6 primes 60S pre-ribosome nuclear export at low temperature. *Nat Commun*. 2021;12:4696.
- Gheorghita AA, Li YE, Kitova EN, Bui DT, Pfoh R, Low KE, et al. Structure of the AlgKX modification and secretion complex required for alginate production and biofilm attachment in *Pseudomonas aeruginosa*. *Nat Commun*. 2022;13:7631.

- Grant B, Greenwald I. The *Caenorhabditis elegans* sel-1 gene, a negative regulator of lin-12 and glp-1, encodes a predicted extracellular protein. *Genetics*. 1996;143:237–47.
- Grinthal A, Adamovic I, Weiner B, Karplus M, Kleckner N. PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. *Proc Natl Acad Sci U S A*. 2010;107:2467–72.
- Guan Z, Pei K, Wang J, Cui Y, Zhu X, Su X, et al. Structural insights into DNA recognition by AimR of the arbitrium communication system in the SPbeta phage. *Cell Discov*. 2019;5:29.
- Gumber HK, McKenna JF, Tolmie AF, Jalovec AM, Kartick AC, Graumann K, et al. MLKS2 is an ARM domain and F-actin-associated KASH protein that functions in stomatal complex development and meiotic chromosome segregation. *Nucleus*. 2019;10:144–66.
- Guo X, Aviles G, Liu Y, Tian R, Unger BA, Lin Y-HT, et al. Mitochondrial stress is relayed to the cytosol by an OMA1–DELE1–HRI pathway. *Nature*. 2020;579:427–32.
- Guo Q, Huang B, Cheng J, Seefeldler M, Engler T, Pfeifer G, et al. The cryo-electron microscopy structure of huntingtin. *Nature*. 2018;555:117–20.
- Guo J, Peters KL, Sen GC. Induction of the human protein P56 by interferon, double-stranded RNA, or virus infection. *Virology*. 2000;267:209–19.
- Gupta YK, Nair DT, Wharton RP, Aggarwal AK. Structures of human Pumilio with noncognate RNAs reveal molecular mechanisms for binding promiscuity. *Structure*. 2008;16:549–57.
- Harper CC, Berg JM, Gould SJ. PEX5 binds the PTS1 independently of Hsp70 and the peroxin PEX12. *J Biol Chem*. 2003;278:7897–901.
- Harper SQ, Staber PD, He X, Eliason SL, Martins IH, Mao Q, et al. RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. *Proc Natl Acad Sci U S A*. 2005;102:5820–5.
- Hemmings BA, Adams-Pearson C, Maurer F, Mueller P, Goris J, Merlevede W, et al. Alpha- and beta-forms of the 65-kDa subunit of protein phosphatase 2A have a similar 39 amino acid repeating structure. *Biochemistry*. 1990;29:3166–73.
- Herbert M, Squire C, Mercer A. Poxviral ankyrin proteins. *Viruses*. 2015;7:709–38.
- Heringa J. Detection of internal repeats: how common are they? *Curr Opin Struct Biol*. 1998;8:338–45.
- Hirano T, Kinoshita N, Morikawa K, Yanagida M. Snap helix with knob and hole: essential repeats in *S. pombe* nuclear protein nuc2+. *Cell*. 1990;60:319–28.
- Hirsh L, Paladin L, Piovesan D, Tosatto SCE. RepeatsDB-lite: a web server for unit annotation of tandem repeat proteins. *Nucleic Acids Res*. 2018;46:W402–7.
- Hromić-Jahjefendić A, Jajčanin Jozić N, Kazazić S, Grabar Branilović M, Karačić Z, Schrittwieser JH, et al. A novel *Porphyromonas gingivalis* enzyme: an atypical dipeptidyl peptidase III with an ARM repeat domain. *PLoS One*. 2017;12:e0188915.
- Huh S, Paek K-H. APUM5, encoding a Pumilio RNA binding protein, negatively regulates abiotic stress responsive gene expression. *BMC Plant Biol*. 2014;14:75.
- Jeong J-H, Kim Y-S, Rojviriyi C, Cha HJ, Ha S-C, Kim Y-G. The structure of TON1937 from archaeon *Thermococcus onnurineus* NA1 reveals a eukaryotic HEAT-like architecture. *Int J Biol Macromol*. 2013;61:433–8.
- Jeong H, Park J, Kim H-I, Lee M, Ko Y-J, Lee S, et al. Mechanistic insight into the nucleus–vacuole junction based on the Vac8p–Nvj1p crystal structure. *Proc Natl Acad Sci U S A*. 2017;114:E4539–48. <https://doi.org/10.1073/pnas.1701030114>
- Jernigan KK, Bordenstein SR. Ankyrin domains across the tree of life. *PeerJ*. 2014;2:e264.
- Ji Y, Luo Y, Wu Y, Sun Y, Zhao L, Xue Z, et al. SEL1L–HRD1 endoplasmic reticulum-associated degradation controls STING-mediated innate immunity by limiting the size of the activable STING pool. *Nat Cell Biol*. 2023;25:726–39.
- Ju D, Wang X, Xu H, Xie Y. The armadillo repeats of the Ufd4 ubiquitin ligase recognize ubiquitin-fusion proteins. *FEBS Lett*. 2007;581:265–70.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–9.
- Kajava AV. Review: proteins with repeated sequence—structural prediction and modeling. *J Struct Biol*. 2001;134:132–44.
- Kajava AV. Tandem repeats in proteins: from sequence to structure. *J Struct Biol*. 2012;179:279–88.
- Kappel C, Zachariae U, Dölker N, Grubmüller H. An unusual hydrophobic core confers extreme flexibility to HEAT repeat proteins. *Biophys J*. 2010;99:1596–603.
- Kaye JA, Rose NC, Goldsworthy B, Goga A, L'Etoile ND. A 3'UTR Pumilio-binding element directs translational activation in olfactory sensory neurons. *Neuron*. 2009;61:57–70.
- Kedde M, Van Kouwenhove M, Zwart W, Oude Vrielink JAF, Elkon R, Agami R. A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nat Cell Biol*. 2010;12:1014–20.
- Kim H, Park J, Kim H, Ko N, Park J, Jang E, et al. Structures of Vac8-containing protein complexes reveal the underlying mechanism by which Vac8 regulates multiple cellular processes. *Proc Natl Acad Sci U S A*. 2023;120:e2211501120.
- Kitamata M, Hotta M, Hamada-Nakahara S, Suetsugu S. The membrane binding and deformation property of vaccinia virus K1 ankyrin repeat domain protein. *Genes Cells*. 2020;25:187–96.
- Kobe B, Kajava AV. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem Sci*. 2000;25:509–15.
- Kordeli E, Lambert S, Bennett V. Ankyrin. *J Biol Chem*. 1995;270:2352–9.
- Kozjak-Pavlovic V, Prell F, Thiede B, Götz M, Wosiek D, Ott C, et al. C1orf163/RESA1 is a novel mitochondrial intermembrane space protein connected to respiratory chain assembly. *J Mol Biol*. 2014;426:908–20.
- Kuo M-W, Wang S-H, Chang J-C, Chang C-H, Huang L-J, Lin H-H, et al. A novel puf-a gene predicted from evolutionary analysis is involved in the development of eyes and primordial germ-cells. *PLoS One*. 2009;4:e4980.
- Kyöstilä K, Cizinauskas S, Seppälä EH, Suhonen E, Jeserevics J, Sukura A, et al. A SEL1L mutation links a canine progressive early-onset cerebellar ataxia to the endoplasmic reticulum-associated protein degradation (ERAD) machinery. *PLoS Genet*. 2012;8:e1002759.
- Kyrpides NC, Woese CR. Tetratrico-peptide-repeat proteins in the archaeon *Methanococcus jannaschii*. *Trends Biochem Sci*. 1998;23:245–7.
- Lamb JR, Tugendreich S, Hieter P. Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci*. 1995;20:257–9.

- Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol.* 2011;3:a003616.
- Li Z, Lee I, Moradi E, Hung N-J, Johnson AW, Marcotte EM. Rational extension of the ribosome biogenesis pathway using network-guided genetics. *PLoS Biol.* 2009;7:e1000213.
- Li J, Mahajan A, Tsai M-D. Ankyrin repeat: a unique motif mediating protein–protein interactions. *Biochemistry.* 2006;45:15168–78.
- Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science.* 2023;379:1123–30.
- Liu X-Y, Chen W, Wei B, Shan Y-F, Wang C. IFN-induced TPR protein IFIT3 potentiates antiviral signaling by bridging MAVS and TBK1. *J Immunol.* 2011;187:2559–68.
- Llabrés S, Tsenkov MI, MacGowan SA, Barton GJ, Zachariae U. Disease related single point mutations alter the global dynamics of a tetratricopeptide (TPR) α -solenoid domain. *J Struct Biol.* 2020;209:107405.
- Lohße A, Borg S, Raschdorf O, Kolinko I, Tompa É, Pósfai M, et al. Genetic dissection of the *mamAB* and *mms6* operons reveals a gene set essential for magnetosome biogenesis in *Magnetospirillum gryphiswaldense*. *J Bacteriol.* 2014;196:2658–69.
- Loureiro J, Peifer M. Roles of armadillo, a drosophila catenin, during central nervous system development. *Curr Biol.* 1998;8:622–33.
- Löw C, Weininger U, Neumann P, Klepsch M, Lilie H, Stubbs MT, et al. Structural insights into an equilibrium folding intermediate of an archaeal ankyrin repeat protein. *Proc Natl Acad Sci U S A.* 2008;105:3779–84.
- Lüthy L, Grütter MG, Mittl PRE. The crystal structure of *Helicobacter pylori* cysteine-rich protein B reveals a novel fold for a penicillin-binding protein. *J Biol Chem.* 2002;277:10187–93.
- Macdonald M. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell.* 1993;72:971–83.
- Magliery TJ, Regan L. Sequence variation in ligand binding sites in proteins. *BMC Bioinform.* 2005;6:240.
- Mak AN-S, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science.* 2012;335:716–9.
- Mandal A, Mishra AK, Dulani P, Muthamilarasan M, Shweta S, Prasad M. Identification, characterization, expression profiling, and virus-induced gene silencing of armadillo repeat-containing proteins in tomato suggest their involvement in tomato leaf curl New Delhi virus resistance. *Funct Integr Genomics.* 2018;18:101–11.
- Manthey GM, McEwen JE. The product of the nuclear gene PET309 is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial COX1 locus of *Saccharomyces cerevisiae*. *EMBO J.* 1995;14:4031–43.
- Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D. A census of protein repeats. *J Mol Biol.* 1999;293:151–60.
- Menon KP, Sanyal S, Habara Y, Sanchez R, Wharton RP, Ramaswami M, et al. The translational repressor Pumilio regulates presynaptic morphology and controls postsynaptic accumulation of translation factor eIF-4E. *Neuron.* 2004;44:663–76.
- Michaely P. Crystal structure of a 12 ANK repeat stack from human ankyrinR. *EMBO J.* 2002;21:6387–96.
- Miller MT, Higgin JJ, Tanaka Hall TM. Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nat Struct Mol Biol.* 2008;15:397–402.
- Mittl PRE, Schneider-Brachert W. Sel1-like repeat proteins in signal transduction. *Cell Signal.* 2007;19:20–31.
- Moiani A, Letort G, Lizot S, Chalumeau A, Foray C, Felix T, et al. Non-viral DNA delivery and TALEN editing correct the sickle cell mutation in hematopoietic stem cells. *Nat Commun.* 2024;15:4965.
- Monzon AM, Arriás PN, Elofsson A, Mier P, Andrade-Navarro MA, Bevilacqua M, et al. A STRP-ed definition of structured tandem repeats in proteins. *J Struct Biol.* 2023;215:108023.
- Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, et al. Identification of a gene causing human cytochrome *c* oxidase deficiency by integrative genomics. *Proc Natl Acad Sci U S A.* 2003;100:605–10.
- Morbiter R, Römer P, Boch J, Lahaye T. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc Natl Acad Sci U S A.* 2010;107:21617–22.
- Morselli M, Holton TR, Pellegrini M, Yeates TO, Arbing MA. Design and construction of a designed ankyrin repeat protein (DARPin) display library. *Curr Protoc.* 2024;4:e960.
- Mosavi LK, Cammett TJ, Desrosiers DC, Peng Z. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* 2004;13:1435–48.
- Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science.* 2009;326:1501.
- Mozaffari S, Arriás PN, Clementel D, Piovesan D, Ferrari C, Tosatto SCE, et al. STRPsearch: fast detection of structured tandem repeat proteins. 2024 <https://doi.org/10.1101/2024.07.10.602726>
- Mullins EA, Shi R, Kotsch LA, Eichman BF. A new family of HEAT-like repeat proteins lacking a critical substrate recognition motif present in related DNA glycosylases. *PLoS One.* 2015;10:e0127733.
- Mullins EA, Shi R, Parsons ZD, Yuen PK, David SS, Igarashi Y, et al. The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. *Nature.* 2015;527:254–8.
- Murata Y, Wharton RP. Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in drosophila embryos. *Cell.* 1995;80:747–56.
- Nakamura A, Kakihara Y, Funayama A, Haga K, Mikami T, Kobayashi D, et al. HEATR1, a novel interactor of pontin/Reptin, stabilizes pontin/reptin and promotes cell proliferation of oral squamous cell carcinoma. *Biochem Biophys Res Commun.* 2021;557:294–301.
- Naudin C, Hattabi A, Michelet F, Miri-Nezhad A, Benyoucef A, Pflumio F, et al. PUMILIO/FOXP1 signaling drives expansion of hematopoietic stem/progenitor and leukemia cells. *Blood.* 2017;129:2493–506.
- Nguyen HV, Suzuki E, Oestreicher Z, Minamide H, Endoh H, Fukumori Y, et al. A protein-protein interaction in magnetosomes: TPR protein MamA interacts with an Mms6 protein. *Biochem Biophys Rep.* 2016;7:39–44.
- Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in drosophila. *Nature.* 1980;287:795–801.
- Odon V, Georgana I, Holley J, Morata J, Maluquer De Motes C. Novel class of viral ankyrin proteins targeting the host E3 ubiquitin ligase Cullin-2. *J Virol.* 2018;92:e0137418.

- O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Linneweber C, et al. On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol.* 2008;25:1120–8.
- Page MD, Koppeschaar HPF, Edwards CA, Dieguez C, Scanlon MF. Additive effects of growth hormone releasing factor and insulin hypoglycaemia on growth hormone release in man. *Clin Endocrinol.* 1987;26:589–95.
- Paladin L, Bevilacqua M, Errigo S, Piovesan D, Mičetić I, Necci M, et al. RepeatsDB in 2021: improved data and extended classification for protein tandem repeat structures. *Nucleic Acids Res.* 2021;49:D452–7.
- Papandréou M, Vacher H, Fache M, Klingler E, Rueda-Boroni F, Ferracci G, et al. CK 2-regulated schwannomin-interacting protein IQCJ-SCHIP-1 association with AnkG contributes to the maintenance of the axon initial segment. *J Neurochem.* 2015;134:527–37.
- Park J, Kim H-I, Jeong H, Lee M, Jang SH, Yoon SY, et al. Quaternary structures of Vac8 differentially regulate the Cvt and PMN pathways. *Autophagy.* 2020;16:991–1006.
- Parra RG, Espada R, Verstraete N, Ferreira DU. Structural and energetic characterization of the ankyrin repeat protein family. *PLoS Comput Biol.* 2015;11:e1004659.
- Pastorello I, Rossi Paccani S, Rosini R, Mattera R, Ferrer Navarro M, Urosev D, et al. EsiB, a novel pathogenic *Escherichia coli* secretory immunoglobulin A-binding protein impairing neutrophil activation. *MBio.* 2013;4:e0020613.
- Perez-Riba A, Itzhaki LS. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. *Curr Opin Struct Biol.* 2019;54:43–9.
- Pernig S, Lamprecht A, Steiner RA, Dodding MP. Structural basis for Kinesin-1: cargo recognition. *Science.* 2013;340:356–9.
- Piątkowski J, Golik P. Yeast pentatricopeptide protein Dmr1 (Ccm1) binds a repetitive AU-rich motif in the small subunit mitochondrial ribosomal RNA. *RNA.* 2020;26:1268–82.
- Plé S, Job V, Dessen A, Attree I. Cochaperone interactions in export of the type III needle component PscF of *Pseudomonas aeruginosa*. *J Bacteriol.* 2010;192:3801–8.
- Plückthun A. Designed ankyrin repeat proteins (DARPin): binding proteins for research, diagnostics, and therapy. *Annu Rev Pharmacol Toxicol.* 2015;55:489–511.
- Potkin KT, Potkin SG. New directions in therapeutics for Huntington disease. *Future Neurol.* 2018;13:101–21.
- Pusnik M, Schneider A. A trypanosomal pentatricopeptide repeat protein stabilizes the mitochondrial mRNAs of cytochrome oxidase subunits 1 and 2. *Eukaryot Cell.* 2012;11:79–87.
- Pusnik M, Small I, Read LK, Fabbro T, Schneider A. Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. *Mol Cell Biol.* 2007;27:6876–88.
- Qi W, Acosta-Zaldivar M, Flanagan PR, Liu N-N, Jani N, Fierro JF, et al. Stress- and metabolic responses of *Candida albicans* require Tor1 kinase N-terminal HEAT repeats. *PLoS Pathog.* 2022;18:e1010089.
- Qiu K, Dunin-Horkawicz S, Lupas A. Exploiting protein language model sequence representations for repeat detection. 2024 <https://doi.org/10.1101/2024.06.07.596093>
- Qiu C, McCann KL, Wine RN, Baserga SJ, Hall TMT. A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. *Proc Natl Acad Sci U S A.* 2014;111:18554–9.
- Ren Z, Fan K, Zhen S, Zhang J, Liu Y, Fu J, et al. Tetratricopeptide-containing SMALL KERNEL 11 is essential for the assembly of cytochrome *c* oxidase in maize mitochondria. *Plant Physiol.* 2023;192:170–87.
- Richter I, Uzum Z, Wein P, Molloy EM, Moebius N, Stinear TP, et al. Transcription activator-like effectors from endosymbiotic bacteria control the reproduction of their fungal host. *MBio.* 2023;14:e0182423.
- Riggleman B, Wieschaus E, Schedl P. Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a drosophila segment polarity gene. *Genes Dev.* 1989;3:96–113.
- Rojas M, Ruwe H, Miranda RG, Zoschke R, Hase N, Schmitz-Linneweber C, et al. Unexpected functional versatility of the pentatricopeptide repeat proteins PGR3, PPR5 and PPR10. *Nucleic Acids Res.* 2018;46:10448–59. <https://doi.org/10.1093/nar/gky737/5075029>
- Rojas M, Yu Q, Williams-Carrier R, Maliga P, Barkan A. Engineered PPR proteins as inducible switches to activate the expression of chloroplast transgenes. *Nat Plants.* 2019;5:505–11.
- Rolando M, Wah Chung IY, Xu C, Gomez-Valero L, England P, Cygler M, et al. The SET and ankyrin domains of the secreted *Legionella pneumophila* histone methyltransferase work together to modify host chromatin. *MBio.* 2023;14:e0165523.
- Ross CA, Aylward EH, Wild EJ, Langbehn DR, Long JD, Warner JH, et al. Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nat Rev Neurol.* 2014;10:204–16.
- Rothenberger S, Hurdiss DL, Walser M, Malvezzi F, Mayor J, Rytter S, et al. The trispecific DARPin ensovibep inhibits diverse SARS-CoV-2 variants. *Nat Biotechnol.* 2022;40:1845–54.
- Saini LK, Sharma M, Ravi B, Ghosh S, Pahuja S, Singh N, et al. Overexpression of ARM repeat/U-box containing E3 ligase, PUB2 positively regulates growth and oxidative stress response in Arabidopsis. *Biochem J.* 2023;480:555–71.
- Saint-Georges Y, Garcia M, Delaveau T, Jourden L, Le Crom S, Lemoine S, et al. Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. *PLoS One.* 2008;3:e2293.
- Sakuma M, Nishikawa S, Inaba S, Nishigaki T, Kojima S, Homma M, et al. Structure of the periplasmic domain of SflA involved in spatial regulation of the flagellar biogenesis of vibrio reveals a TPR/SLR-like fold. *J Biochem.* 2019;166:197–204.
- Sanches MN, Parra RG, Viegas RG, Oliveira AB, Wolyne PG, Ferreira DU, et al. Resolving the fine structure in the energy landscapes of repeat proteins. *QRB Discov.* 2022;3:e7.
- Sandbaken MG, Lupisella JA, DiDomenico B, Chakraborty K. Protein synthesis in yeast. Structural and functional analysis of the gene encoding elongation factor 3. *J Biol Chem.* 1990;265:15838–44.
- Sanderlin AG, Hanna RE, Lamason RL. The ankyrin repeat protein RARP-1 is a periplasmic factor that supports *Rickettsia parkeri* growth and host cell invasion. *J Bacteriol.* 2022;204:e0018222.
- Sarkar SN, Sen GC. Novel functions of proteins encoded by viral stress-inducible genes. *Pharmacol Ther.* 2004;103:245–59.
- Satoh R, Tanaka T, Yoshida N, Tanaka C, Takasaki T, Sugiura R. Fission yeast PUF proteins Puf3 and Puf4 are novel regulators of PI4P5K signaling. *Biol Pharm Bull.* 2023;46:163–9.
- Saxton RA, Sabatini DM. mTOR signaling in growth, metabolism, and disease. *Cell.* 2017;168:960–76.
- Schilling J, Jost C, Ilie IM, Schnabl J, Buechi O, Eapen RS, et al. Thermostable designed ankyrin repeat proteins (DARPin) as

- building blocks for innovative drugs. *J Biol Chem.* 2022;298:101403.
- Schilling J, Schöppe J, Plückthun A. From DARPins to LoopDARPins: novel LoopDARPin design allows the selection of low picomolar binders in a single round of ribosome display. *J Mol Biol.* 2014;426:691–721.
- Schlegel T, Mirus O, Von Haeseler A, Schleiff E. The tetratricopeptide repeats of receptors involved in protein translocation across membranes. *Mol Biol Evol.* 2007;24:2763–74.
- Schmitz-Linneweber C, Williams-Carrier RE, Williams-Voelker PM, Kroeger TS, Vichas A, Barkan A. A pentatricopeptide repeat protein facilitates the *trans*-splicing of the maize chloroplast *rps12* pre-mRNA. *Plant Cell.* 2006;18:2650–63.
- Schornack S, Moscou MJ, Ward ER, Horvath DM. Engineering plant disease resistance based on TAL effectors. *Annu Rev Phytopathol.* 2013;51:383–406.
- Sedgwick SG, Smerdon SJ. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem Sci.* 1999;24:311–6.
- Seefelder M, Alva V, Huang B, Engler T, Baumeister W, Guo Q, et al. The evolution of the huntingtin-associated protein 40 (HAP40) in conjunction with huntingtin. *BMC Evol Biol.* 2020;20:162.
- Sen GC, Sarkar SN. The interferon-stimulated genes: targets of direct signaling by interferons, double-stranded RNA, and viruses. In: Pitha PM, editor. *Interferon: the 50th anniversary.* Vol. 316. Current topics in microbiology and immunology. Berlin; Heidelberg: Springer; 2007. p. 233–50. https://doi.org/10.1007/978-3-540-71329-6_12
- Sharma M, Singh A, Shankar A, Pandey A, Baranwal V, Kapoor S, et al. Comprehensive expression analysis of rice armadillo gene family during abiotic stress and development. *DNA Res.* 2014; 21:267–83.
- Shi R, Shen X, Rokas A, Eichman BF. Structural biology of the HEAT-like repeat family of DNA glycosylases. *Bioessays.* 2018; 40:1800133.
- Sikorski RS, Boguski MS, Goebel M, Hieter P. A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell.* 1990;60:307–17.
- Silva ILZ, Kohata AA, Shigunov P. Modulation and function of Pumilio proteins in cancer. *Semin Cancer Biol.* 2022;86: 298–309.
- Smialek MJ, Ilaslan E, Sajek MP, Jaruzelska J. Role of PUM RNA-binding proteins in cancer. *Cancer.* 2021;13:129.
- Sobhy H, Scola BL, Pagnier I, Raoult D, Colson P. Identification of giant mimivirus protein functions using RNA interference. *Front Microbiol.* 2015;6:345. <https://doi.org/10.3389/fmicb.2015.00345/abstract>
- Sonnberg S, Fleming SB, Mercer AA. Phylogenetic analysis of the large family of poxvirus ankyrin-repeat proteins reveals orthologue groups within and across Chordopoxvirus genera. *J Gen Virol.* 2011;92:2596–607.
- Stokar-Avihail A, Tal N, Erez Z, Lopatina A, Sorek R. Widespread utilization of peptide communication in phages infecting soil and pathogenic bacteria. *Cell Host Microbe.* 2019;25:746–755.e5.
- Tam PP, Barrette-Ng IH, Simon DM, Tam MW, Ang AL, Muench DG. The Puf family of RNA-binding proteins in plants: phylogeny, structural modeling, activity and subcellular localization. *BMC Plant Biol.* 2010;10:44.
- Tang F, Peng Y, Nau JJ, Kauffman EJ, Weisman LS. Vac8p, an armadillo repeat protein, coordinates vacuole inheritance with multiple vacuolar processes. *Traffic.* 2006;7:1368–77.
- Tang Q, Zhao Y-N, Luo S, Lu S. AKR2A is involved in the flowering process of *Arabidopsis thaliana*. *Plant Signal Behav.* 2022;17: 2100685.
- Tran HC, Schmitt V, Lama S, Wang C, Launay-Avon A, Bernfur K, et al. An mTRAN-mRNA interaction mediates mitochondrial translation initiation in plants. *Science.* 2023;381:eadg0995.
- Turjanski P, Parra RG, Espada R, Becher V, Ferreiro DU. Protein repeats from first principles. *Sci Rep.* 2016;6:23959.
- Utégés JS, Tsenkov MI, Dietrich NJM, MacGowan SA, Barton GJ. Ankyrin repeats in context with human population variation. *PLoS Comput Biol.* 2021;17:e1009335.
- Valenta T, Hausmann G, Basler K. The many faces and functions of β -catenin: β -catenin: a life by, beyond, and against the Wnt canon. *EMBO J.* 2012;31:2714–36.
- Vallese F, Kim K, Yen LY, Johnston JD, Noble AJ, Cali T, et al. Architecture of the human erythrocyte ankyrin-1 complex. *Nat Struct Mol Biol.* 2022;29:706–18.
- Van Etten J, Schagat TL, Hrit J, Weidmann CA, Brumbaugh J, Coon JJ, et al. Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. *J Biol Chem.* 2012;287:36370–83.
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2022;50: D439–44.
- Verdier V, Vera Cruz C, Leach JE. Controlling rice bacterial blight in Africa: needs and prospects. *J Biotechnol.* 2012;159: 320–8.
- Vladimer GI, Górna MW, Superti-Furga G. IFITs: emerging roles as key anti-viral proteins. *Front Immunol.* 2014;10(5):94. <https://doi.org/10.3389/fimmu.2014.00094/abstract>
- Voth KA, Chung IYW, Van Straaten K, Li L, Boniecki MT, Cygler M. The structure of *legionella* effector protein LpnE provides insights into its interaction with oculocerebrorenal syndrome of Lowe (OCRL) protein. *FEBS J.* 2019;286:710–25.
- Walburger A, Koul A, Ferrari G, Nguyen L, Prescianotto-Baschong C, Huygen K, et al. Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science.* 2004;304:1800–4.
- Walser M, Mayor J, Rothenberger S. Designed ankyrin repeat proteins: a new class of viral entry inhibitors. *Viruses.* 2022;14: 2242.
- Wang G, Barrett JW, Stanford M, Werden SJ, Johnston JB, Gao X, et al. Infection of human cancer cells with myxoma virus requires Akt activation via interaction with a viral ankyrin-repeat host range factor. *Proc Natl Acad Sci U S A.* 2006;103: 4640–5.
- Wang X, McLachlan J, Zamore PD, Hall TMT. Modular recognition of RNA by a human Pumilio-homology domain. *Cell.* 2002;110:501–12.
- Wang B, Ye K. Nop9 binds the central pseudoknot region of 18S rRNA. *Nucleic Acids Res.* 2017;45:gkw1323.
- Wang X, Zamore PD, Hall TMT. Crystal structure of a Pumilio homology domain. *Mol Cell.* 2001;7:855–65.

- Weis D, Lin LL, Wang HH, Li ZJ, Kusikova K, Ciznar P, et al. Biallelic Cys141Tyr variant of SEL1L is associated with neurodevelopmental disorders, agammaglobulinemia, and premature death. *J Clin Investig*. 2024;134:e170882.
- Werden SJ, Barrett JW, Wang G, Stanford MM, McFadden G. M-T5, the ankyrin repeat, host range protein of myxoma virus, activates Akt and can be functionally replaced by cellular PIKE-A. *J Virol*. 2007;81:2340–8.
- Wheeler TJ, Clements J, Finn RD. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinform*. 2014;15:7.
- Xu F, Morin C, Mitchell G, Ackerley C, Robinson BH. The role of the *LRPPRC* (leucine-rich pentatricopeptide repeat cassette) gene in cytochrome oxidase assembly: mutation causes lowered levels of COX (cytochrome *c* oxidase) I and COX III mRNA. *Biochem J*. 2004;382:331–6.
- Xu Y, Xing Y, Chen Y, Chao Y, Lin Z, Fan E, et al. Structure of the protein phosphatase 2A holoenzyme. *Cell*. 2006;127:1239–51.
- Xue C, Lam KH, Zhang H, Sun K, Lee SH, Chen X, et al. Crystal structure of the FliF–FliG complex from helicobacter pylori yields insight into the assembly of the motor MS–C ring in the bacterial flagellum. *J Biol Chem*. 2018;293:2066–78.
- Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One*. 2013;8:e57286.
- Yagi Y, Teramoto T, Kaieda S, Imai T, Sasaki T, Yagi M, et al. Construction of a versatile, programmable RNA-binding protein using designer PPR proteins and its application for splicing control in mammalian cells. *Cells*. 2022;11:3529.
- Yan J, Wang J, Zhang H. An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism. *Plant J*. 2002;29:193–202.
- Yang X-M, Wang X-Q, Hu L-P, Feng M-X, Zhou Y-Q, Li D-X, et al. Nucleolar HEAT repeat containing 1 up-regulated by the mechanistic target of rapamycin complex 1 signaling promotes hepatocellular carcinoma growth by dominating ribosome biogenesis and proteome homeostasis. *Gastroenterology*. 2023;165:629–46.
- Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, et al. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*. 2013;504:168–71.
- Yoshimura SH, Hirano T. HEAT repeats – versatile arrays of amphiphilic helices working in crowded environments? *J Cell Sci*. 2016;129:3963–70.
- Yu F, Morin X, Kaushik R, Bahri S, Yang X, Chia W. A mouse homologue of *Drosophila pins* can asymmetrically localize and substitute for *pins* function in *Drosophila* neuroblasts. *J Cell Sci*. 2003;116:887–96.
- Zamudio-Ochoa A, Camacho-Villasana Y, García-Guerrero AE, Pérez-Martínez X. The Pet309 pentatricopeptide repeat motifs mediate efficient binding to the mitochondrial *COX1* transcript in yeast. *RNA Biol*. 2014;11:953–67.
- Zeytuni N, Ozyamak E, Ben-Harush K, Davidov G, Levin M, Gat Y, et al. Self-recognition mechanism of MamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. *Proc Natl Acad Sci U S A*. 2011;108:E480–7. <https://doi.org/10.1073/pnas.1103367108>
- Zhang J, McCann KL, Qiu C, Gonzalez LE, Baserga SJ, Hall TMT. Nop9 is a PUF-like protein that prevents premature cleavage to correctly process pre-18S rRNA. *Nat Commun*. 2016;7:13085.
- Zhang C, Muench DG. A nucleolar PUF RNA-binding protein with specificity for a unique RNA sequence. *J Biol Chem*. 2015;290:30108–18.
- Zhang H, Wu S, Itzhaki LS, Perrett S. Interaction between huntingtin exon 1 and HEAT repeat structure probed by chimeric model proteins. *Protein Sci*. 2023;32:e4810.
- Zhang Q, Xu Y, Huang J, Zhang K, Xiao H, Qin X, et al. The Rice pentatricopeptide repeat protein PPR756 is involved in pollen development by affecting multiple RNA editing in mitochondria. *Front Plant Sci*. 2020;11:749.
- Zhao Y-Y, Mao M-W, Zhang W-J, Wang J, Li H-T, Yang Y, et al. Expanding RNA binding specificity and affinity of engineered PUF domains. *Nucleic Acids Res*. 2018;46:4771–82.
- Zheng S, Dong J, Lu J, Li J, Jiang D, Yu H, et al. A cytosolic pentatricopeptide repeat protein is essential for tapetal plastid development by regulating *OsGLK1* transcript levels in rice. *New Phytol*. 2022;234:1678–95.
- Zheng P, Liu Y, Liu X, Huang Y, Sun F, Wang W, et al. OsPPR939, a nad5 splicing factor, is essential for plant growth and pollen development in rice. *Theor Appl Genet*. 2021;134:923–40.
- Zheng C, Zheng Z, Zhang Z, Meng J, Liu Y, Ke X, et al. IFIT5 positively regulates NF- κ B signaling through synergizing the recruitment of I κ B kinase (IKK) to TGF- β -activated kinase 1 (TAK1). *Cell Signal*. 2015;27:2343–54.
- Zhou X, Liao H, Chern M, Yin J, Chen Y, Wang J, et al. Loss of function of a rice TPR-domain RNA-binding protein confers broad-spectrum disease resistance. *Proc Natl Acad Sci U S A*. 2018;115:3174–9.
- Zhou W, Melamed D, Banyai G, Meyer C, Tuschl T, Wickens M, et al. Expanding the binding specificity for RNA recognition by a PUF domain. *Nat Commun*. 2021;12:5107.
- Zhou Y, Wang K, Zhou Y, Li T, Yang M, Wang R, et al. HEATR1 deficiency promotes pancreatic cancer proliferation and gemcitabine resistance by up-regulating Nrf2 signaling. *Redox Biol*. 2020;29:101390.
- Zhu J, Wen W, Zheng Z, Shang Y, Wei Z, Xiao Z, et al. LGN/mInsc and LGN/NuMA complex structures suggest distinct functions in asymmetric cell division for the Par3/mInsc/LGN and Gai/LGN/NuMA pathways. *Mol Cell*. 2011;43:418–31.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Arrías PN, Osmanli Z, Peralta E, Chinestrada PM, Monzon AM, Tosatto SCE. Diversity and structural-functional insights of alpha-solenoid proteins. *Protein Science*. 2024;33(11):e5189. <https://doi.org/10.1002/pro.5189>