



# The Effect of Mutations in the TPR and Ankyrin Families of Alpha Solenoid Repeat Proteins

Matylda Anna Izert, Patrycja Emanuela Szybowska, Maria Wiktoria Górna\* and Matthew Merski\*

Structural Biology Group, Department of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Warsaw, Poland

## OPEN ACCESS

### Edited by:

Tugce Bilgin Sonay,  
Columbia University, United States

### Reviewed by:

Diego U Ferreira,  
University of Buenos Aires, Argentina  
R. Gonzalo Parra,  
European Molecular Biology  
Laboratory Heidelberg, Germany

### \*Correspondence:

Maria Wiktoria Górna  
mgorna@chem.uw.edu.pl  
Matthew Merski  
merski@gmail.com

### Specialty section:

This article was submitted to  
Protein Bioinformatics,  
a section of the journal  
Frontiers in Bioinformatics

**Received:** 16 April 2021

**Accepted:** 22 June 2021

**Published:** 06 July 2021

### Citation:

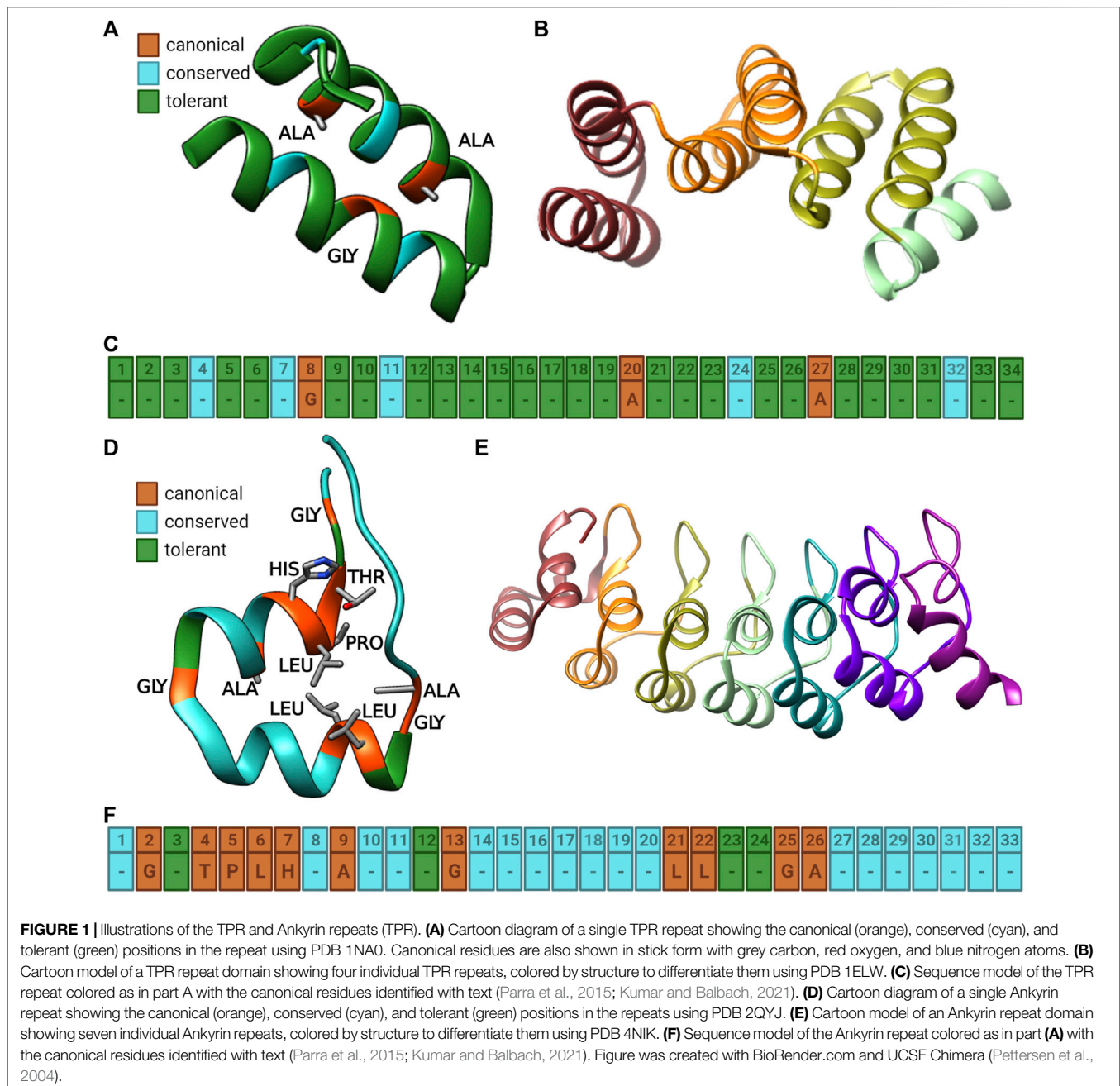
Izert MA, Szybowska PE, Górna MW  
and Merski M (2021) The Effect of  
Mutations in the TPR and Ankyrin  
Families of Alpha Solenoid  
Repeat Proteins.  
*Front. Bioinform.* 1:696368.  
doi: 10.3389/fbinf.2021.696368

Protein repeats are short, highly similar peptide motifs that occur several times within a single protein, for example the TPR and Ankyrin repeats. Understanding the role of mutation in these proteins is complicated by the competing facts that 1) the repeats are much more restricted to a set sequence than non-repeat proteins, so mutations should be harmful much more often because there are more residues that are heavily restricted due to the need of the sequence to repeat and 2) the symmetry of the repeats in allows the distribution of functional contributions over a number of residues so that sometimes no specific site is singularly responsible for function (unlike enzymatic active site catalytic residues). To address this issue, we review the effects of mutations in a number of natural repeat proteins from the tetratricopeptide and Ankyrin repeat families. We find that mutations are context dependent. Some mutations are indeed highly disruptive to the function of the protein repeats while mutations in identical positions in other repeats in the same protein have little to no effect on structure or function.

**Keywords:** protein repeat, Ankyrin repeat, TPR, mutation, sequence function relationship

## INTRODUCTION

Protein repeats are stretches of 20–50 residues that repeat (both sequence and structure) within the larger protein (Marcotte et al., 1999; Andrade et al., 2001; Kajava, 2012), with as few as three (Mosavi et al., 2002) to more than 100 occurrences of the repeat with a theoretical limit of 120 repeat units (Galpern et al., 2020). The repeats usually occur sequentially (e.g. tandem repeats) (Kajava, 2012; Jorda et al., 2010), although insertions between repeats can also occur. Sequence repetition in the repeats can be as high as 100% identity (perfect repeats) (Jorda et al., 2010), although lower degrees of similarity (imperfect repeats) are much more common. In theory, any sequence can be the basis of a set of repeats but most identified repeats are restricted to a small set of sequence pattern families (Pallen et al., 2003; Cushing et al., 2005; Champion et al., 2009; Marte et al., 2019). Most of the sequence positions in a repeat are highly variable with only a smaller subset of residues being highly conserved and even the most conserved (canonical) positions display some variability (Stumpp et al., 2003; Kobayashi et al., 2012), although there are often functional costs to deviation at these positions (D'Andrea and Regan, 2003; Severi et al., 2008). However, there tends to be a significant enough restriction of the allowed sequences that the same type of repeat can be identified in distantly related species (Jernigan and Bordenstein, 2014; Schaper et al., 2014; Jernigan and Bordenstein, 2015), even among essentially unrelated species such as bacteria and humans (although lateral gene transfer can never be truly eliminated in these cases) (Schaper et al., 2012). As expected, this sequence restriction



results in a high degree of conservation of the secondary structure (Tramontano and Cozzetto, 2005) of the repeats giving rise to highly symmetrical, extended protein structures (Figure 1). As such, local interactions (within a repeat or between neighboring repeats) (Stumpp et al., 2003; Main et al., 2005; Geiger-Schuller et al., 2018) tend to dominate over long range interactions that dominate in non-repeat proteins (Sengupta and Kundu, 2012). As such, repeat sequence repetition may be related to secondary and tertiary structural constraints from these local interactions (Yang et al., 2018) or by co-evolution of interacting surfaces, or some combination of selective pressures (Lovell and Robertson, 2010).

It is possible to classify repeat proteins by their secondary structure composition. For example, the WD40 (Vander Kooi et al., 2010) and Kelch repeats (Severi et al., 2008) are comprised of  $\beta$ -strands arranged into a roughly triangular plane which then collect into a circular shape where the repeats form a topology similar to the blades of a propeller (or the slices of a pizza). Leucine rich repeats contains both a  $\beta$ -strand and an  $\alpha$ -helix which in combination stack upon each other to form a solenoid [spiral staircase-like (Magliery and Regan, 2004)] shape, although the strand region is sometimes replaced with a less complex coil structure (Stumpp et al., 2003). Ankyrin repeats (Tripp and Barrick, 2004) also contain  $\alpha$ -helices and a conserved ligand

binding strand region (Desrosiers and Peng, 2005) and stack into a solenoid structure while the HEAT (Urvoas et al., 2010), Armadillo (Zhao et al., 2009), and tetratricopeptide (TPR) (Grenha et al., 2013) repeats form solenoids from repeats composed of two  $\alpha$ -helices. Tetratricopeptide repeats are readily identified by their conserved 34 residue sequence length (Blom et al., 2004) differentiating them from the so-called TPR-like repeats such as the pentatricopeptide (PPR, 35 residues) (Cushing et al., 2005), octotricopeptide (OPR, 38 residues) (Loizeau et al., 2014), and HAT (half a TPR) (Preker and Keller, 1998) repeats. This clustering is more indicative of the structural rather than sequence similarity of these repeats (Paladin et al., 2017; Paladin et al., 2021), although it can be difficult to determine if two classes of repeats are truly separate as there can sometimes be evidence for unexpected relationships [i.e. HEAT and armadillo repeats (Andrade et al., 2001) or LDLreceptorA and LDLreceptorB, or WD40 and PD40 (Turjanski et al., 2016)].

Unfortunately, many of the expected identifying characteristics are variant (or missing) in any specific example repeat. There is no hard rule that fractional repeats cannot occur within a protein, suggesting that the entirety of the repeat does not need to be conserved (Andrade et al., 2001; Pekkala et al., 2004; Clarke et al., 2008; Espada et al., 2015). Their sequences repeat several times making the sequence patterns essentially circular, diluting the meaning of a “starting” or “ending” sequence position (Wall et al., 1995; Marcotte et al., 1999). The identification of the starting residue of a repeat (e.g. the “phase” of the repeat) has been the topic of much discussion (Michaely and Bennett, 1993; Sedgwick and Smerdon, 1999; Mosavi and Peng, 2003; Parra et al., 2013; Parra et al., 2015) and there has been a report of a detectable natural starting pattern for Ankyrin repeats (Parra et al., 2015). Repeat proteins usually, but not always, unfold in a two-state manner despite the fact that repeats can often be freely added or deleted (Tripp and Barrick, 2004; Main et al., 2005; Mello et al., 2005). Nor is repeat length totally conserved as most repeat types have some flexibility in their lengths as well. Increases in the length of TPR repeats, defined by their 34 amino acid length, up to 42 residues have also been reported (Marold et al., 2015). On top of this, the repeating sequences hamper phylogenetic analyses (Andrade et al., 2001; Schaper et al., 2012) of repeat proteins with some repeat regions appearing to change more quickly (Cervený et al., 2013; Schüler et al., 2016) and others changing more slowly than non-repeat proteins (Schaper et al., 2012). Protein repeat sequences are short enough that they may have arisen more than once in evolutionary history. Functional differences in prokaryotic and eukaryotic repeats point towards this possibility (Marcotte et al., 1999; Kajava, 2001). Further confounding is the observation that in a number of repeat proteins, the functionality is distributed over the full set of repeats, rather than localized to a single repeat (or residue) as is typical for non-repeat proteins (Wang and Lambert, 2010). Which then raises the question, if repeats do not need to be complete to be maintained over evolutionary time and repeat structure (Stumpp et al., 2003), length, and sequence (including the canonical residues) can vary, are mutations in repeats more or

less disruptive than in non-repeat regions in light of the fact that the canonical repeat positions are, in fact, highly conserved?

To answer this question, it is necessary to identify what it means for a residue to be conserved. To some extent conservation is what is tolerated by the fold as well as function and the physiology of the host organism (Magliery and Regan, 2004). While simple sequence conservation can be employed (Preker and Keller, 1998; D’Andrea and Regan, 2003), differences in amino acid usage frequencies bias this analysis. Instead, one can also examine the specific effects of mutational substitutions in repeat proteins by comparing changes to the stability and function of the mutant to the naturally occurring proteins. Direct measurements of these parameters allow a quantitative assessment of the impacts of a point mutation. Several models have been developed to analyze these perturbations including one dimensional Ising analysis (Marold et al., 2021), positional frustration analysis (Parra et al., 2015; Espada et al., 2017), use of energy functions like Rosetta (Zhu et al., 2016), and others (Hutton et al., 2015). Free energy analysis can also be used to define sequence conservation, perhaps more accurately than sequence consensus and co-variation analysis has also been applied to both Ankyrin and TPR proteins (Mosavi et al., 2002; Magliery and Regan, 2004) using both real and simulated data to improve the statistical parameters of the analysis (Travers and Fares, 2007). Consensus designs have been shown to generate more stable repeat proteins (Magliery and Regan, 2004), although recent work has identified subtypes within several repeat families suggesting that several notably different consensus designs are possible for a given repeat (Marchi et al., 2019). Care must be taken here however, as the consensus sequence is not necessarily the most stable and residues in contact with ligands tend to be the most variable in repeat proteins (Magliery and Regan, 2004), suggesting a function/stability trade-off (Karanicolas et al., 2011; Houlihan et al., 2015).

In this mini-review, we will examine the role of mutations in repeat proteins using TPR and Ankyrin proteins, two well studied classes of alpha solenoid repeat proteins (Main et al., 2005). Due to the differences between natural and laboratory selective pressures, we will largely avoid designed repeat proteins as well as mutations in capping helices (Main et al., 2003; Stumpp et al., 2003) and mutations that indirectly affect the protein (Boisson et al., 2017). Nor do we claim that this will be an exhaustive list of every mutation ever documented but with enough detail to produce a fairly confident overall assessment of the effects of these mutations. We also note positional numberings are not always uncontroversial for every protein (Lubman et al., 2004; Li et al., 2010). We use the positional numberings from the referenced work or UniProtKB (Bateman et al., 2021).

## MUTATIONS IN TPR PROTEINS

Of the 34 positions in the TPR repeat (**Figure 1**), sequence conservation identified canonical positions 8, 20, and 27, which are involved in inter helical contacts are the least variable and are typically occupied by alanine or glycine residues (D’Andrea and Regan, 2003; Pallen et al., 2003; Broms et al., 2006; Iakhiaeva et al., 2009; Wang and Lambert,

**TABLE 1** | Tabular representation of mutational data for TPR. Mutations at a position that notably perturbed the protein function are indicated with an “F” and those that perturbed stability or structure with “S,” and “X” if both function and structure were explicitly reported as affected. Those that were functionally neutral are denoted by “O.” Mutations in positions were specifically noted to have an effect in some instances but not in others are indicated by “#.” The associated literature reference citations are given in grey text under the protein name. At least one mutation was identified for every position except 21 position 4 in the TPRs. A more detailed version of this table is provided as **Supplementary Material**.

Positon	SycD Büttner et al. (2008)	T3S Broms et al. (2006), Edqvist et al. (2006)	Peroxin 5 recept. Sampathkumar et al. (2008)	FANCG Blom et al. (2004), Wang and Lambert (2010)	$\gamma$ -secretase Zhang et al. (2012)	PilF Koo et al. (2013)	ISG54 Yang et al. (2012)	O-GlcNAc transferase Rafie et al. (2017)	TRIP8b Han et al. (2011)	SRP72 Iakhaieva et al. (2009)	HOP Kajander et al. (2009)	Prolyl 4-hydroxylase Pekkala et al. (2004)	Cyclophilin 40 Ward et al. (2002)
1		S								F			
2	X	#				O			F		F		F
3	F	X				F						F	
4	X	X				O							
5	X	X							F		X		F
6	X	X			O	O					#		F
7	#	#		F		#							
8		X		F		O				X			
9		X			#	O	#				X	F	#
10	F	#				F	F						O
11		#				O				X			
12	X	#				O	F				O	F	#
13						O	#		F		#		O
14						O			O				
15						O			F		O		
16				O		O							
17	O	#				O						F	
18						O						O	
19						O							
20		X		O						X			
21													
22	O					O					F		
23						O							F
24	F	#				O							
25	O					O							O
26	#				O	O							
27		X				O							
28						O							
29			S		O	O							
30	X		S			O		F					
31	F	O	S			O							F
32	F	#											O
33	#		S					F			S		
34													

**TABLE 2 |** Ankyrin repeats. Mutations at a position that notably perturbed the protein function are indicated with an “F” and those that perturbed stability or structure with “S,” and “X” if both function and structure were explicitly reported as affected. Those that were functionally neutral are denoted by “O.” Mutations in positions were specifically noted to have an effect in some instances but not in others are indicated by “#.” The associated literature reference citations are given in grey text under the protein name. At least one mutation was identified for every position except 21 position 22 in the Ankyrin repeats. A more detailed version of this table is provided as **Supplementary Material**.

Position	Integrin-linked kinase Chiswell et al. (2010)	RFXANK Nekrep et al. (2001)	Ankyrin B Mohler et al. (2004), Abdi et al. (2006)	Notch Lubman et al. (2004), Del Bianco et al. (2008)	K1 Li et al. (2010)	VPS9 Tamura et al. (2011)	DHHC17 Verardi et al. (2017)	KIF21A Guo et al. (2018)	IκBα Inoue et al. (1992)	p16 Yarbrough et al. (1999)	Gankrin Hutton et al. (2015)	TRPV4 Kang (2012)
1	F	#			F		O	F				
2		#							#			
3					#			F				
4				#	#				#	O		
5				O	#				#	F	S	
6				O					#		S	
7		F		O	#				#			
8		F	F	O	#		F		F	F	S	
9				#						F	#	
10		F		S	F					O	S	
11		F	F	S				F		O		
12	O			#	#	O	F			F		F
13				#	#							
14				#	#	F						
15				O	#	F				O		
16		O		O	#					O		
17				O	F	O	F					
18		O									#	
19		O			#	O						
20					#						#	
21											S	
22												
23		O			F	F						
24		O			F		F					
25							F					
26				X								
27					F		O					
28				X	F		F			F		
29							F					
30				X	F	O						
31					F					F		
32		#			F		O					F
33		#		X				F				F

2010). Residues at conserved positions 4, 7, 11, 24, and 32 are also often restricted to a subset of amino acids, although several of these favor large, aromatic side chains (e.g. positions 4, 11, and 24) and they are also involved in the interaction between the two helices of the repeat. Other positions are more tolerant to substitution; here we use a notation of canonical (the most conserved), conserved (highly conserved) and tolerant as previously established (D'Andrea and Regan, 2003) (**Figure 1**). This metric has some notable similarity to measurements of protein frustration (Ferreiro et al., 2007) while statistical free energy analysis suggests a slightly different sequence set (Magliery and Regan, 2004). Many TPR proteins have an additional C-terminal helix which does not follow the canonical TPR pattern (D'Andrea and Regan, 2003; Kajander et al., 2009) while others have a divergent N-terminal repeat (Yuzawa et al., 2011). Functionally, TPR domains tend to be protein-protein interaction domains (Blom et al., 2004; Sampathkumar et al., 2008; Wittwer and Dames, 2015; Bidlingmaier et al., 2016) often as an auto-inhibition module (Wu et al., 2001; Yuzawa et al., 2011) but have also been identified as ceramide binders (Bidlingmaier et al., 2016), or involved in chloroplast development (Stanley et al., 2020),  $\gamma$ -secretase activity (Zhang et al., 2012), outer membrane targeting (Koo et al., 2013), and RNA binding (Katibah et al., 2014), among others (Kang et al., 2001). Like many repeat proteins, TPR proteins often have redundant functions (Sampathkumar et al., 2008; Koo et al., 2013). Several TPR proteins have been the targets of significant research interest including protein phosphatase 5 (PP5) (Kang et al., 2001), the C-terminus of hsc70-interacting protein (CHIP) (Wu et al., 2001), and the type III secretion (T3S) system (Broms et al., 2006) which can serve as good models of TPR protein behavior in general.

We will now examine the effects of some mutational changes in TPR proteins (summarized in **Table 1**). In the *Y. enterocolitica* T3S protein SycD, mutations in several but not all tolerant positions affected binding of one protein ligand but not the other due to their different binding sites, while others and those at conserved positions appeared to affect both (Büttner et al., 2008). Additionally, introduction of charged residues in tolerant positions disrupted dimer formation in some instances but not others (Büttner et al., 2008). In the *P. aeruginosa* T3S protein PcrH, mutations in canonical residues and a double mutant at tolerant positions were found to be greatly detrimental to the protein although all resulted in stable proteins, while mutations in other tolerant and conserved positions did not always destroy the phenotype (Broms et al., 2006). In the *Y. pestis* T3S protein LrcH mutations at tolerant positions conserved, and canonical positions eliminated ligand binding of one or both natural ligands in a yeast two hybrid assay (Edqvist et al., 2006), however, other mutations did not and two mutations in a tolerant positions and several in conserved positions resulted in a negative growth phenotype. Double mutations at the C-terminal end of some repeats allowed crystallization of the peroxin 5 receptor from *T. brucei* by modifying the protein surface to allow the formation of a crystal contact (Sampathkumar et al., 2008). In the nicastrin subunit of human  $\gamma$ -secretase, mutations in tolerant positions were found to be detrimental to activity (Zhang et al., 2012). For the *P. aeruginosa* pilus protein PilF, deletions of repeat 5 or 6 did not reduce

phenotypic activity, nor did mutations at a canonical position or several conserved or tolerant positions but mutations at two tolerant and one conserved position did (Koo et al., 2013). In the human RNA binding protein ISG54, mutation of positively charged residues to negative ones at some tolerant positions abolished RNA binding while those at other tolerant positions did not, but did when combined as a double mutant (Yang et al., 2012). In the nucleotide gated channel, TRIP8b, mutations at several (but not all) tolerant positions disrupted the interaction between the TPR domain and the channel (Han et al., 2011). In human SRP72, deletion of repeat 1 or 4 destroyed complex formation (but not protein solubility/stability) while mutations at a set of canonical and conserved positions heavily reduced solubility and also eliminated complex formation although a mutation at one tolerant position also moderately reduced it (Iakhiaeva et al., 2009). In a very thorough report, Kajander et al. examined the role of the TPR repeats in Hsp-organizing protein both computationally and experimentally (Kajander et al., 2009). A mutation in a tolerant position in TPR domain 1 greatly reduced affinity for Hsp70, but other tolerant positions had expectedly neutral effects, while in domain 2 several (but not all) tolerant mutations greatly reduced affinity for Hsp90 (Kajander et al., 2009). In collagen prolyl 4-hydroxylase, mutations in four tolerant positions were detrimental to activity while another tolerant mutation was neutral (Pekkala et al., 2004). Finally, in cyclophilin 40, mutations in several tolerant positions eliminated ligand binding while those in other tolerant positions and one conserved position did not (Ward et al., 2002).

While a plethora of mutations have been made in a number of TPR proteins, and changes to some specific residues can result in near complete loss of protein function, many mutations in positions that are highly sequence conserved, (or even elimination of one or more full repeats) can have little or even neutral effects on function, while conversely, some mutations in tolerant positions can be highly detrimental to the function of the TPR protein, although this is expected to be due to loss of specific, localized binding interactions rather than protein unfolding (**Table 1**). While there was a bit of a bias towards mutational investigation of the N-terminal half of TPR repeats, overall it is clear that context matters for TPR repeats and mutations at canonical positions and even complete deletion of a repeat may have little or no effect on protein function or stability.

## MUTATIONS IN ANKYRIN REPEATS

The Ankyrin repeat is a 33 amino acid repeat consisting of two helices and a ligand-binding loop region (Parra et al., 2015; Kumar and Balbach, 2021). While most of the repeat is fairly well defined, positions 2, 4–7, 9, 13, 21, 22, 25, and 26 are the most conserved while only positions 3, 12, 23, and 24 can be considered tolerant. (Mosavi et al., 2002). (**Figure 1**, note that there are several numbering systems used for Ankyrin repeats present in the literature (Sedgwick and Smerdon, 1999)). On average, 20% of the residues of an Ankyrin repeat are involved in ligand binding, with the majority (80%) of these being canonical residues and the positions that are most intolerant to mutation tend to be interface interactions between the repeats (Parra et al., 2015). Mutations in

Ankyrin repeats are often targeted at positions conserved within a family rather than at the generic repeat structure due to the broader sequence conservation of these repeats compared to TPR proteins (**Figure 1**) (Tamura et al., 2011). Here we also focus more on mutations in natural proteins rather than designed Ankyrin proteins (Karanicolas et al., 2011) partly because natural proteins are subject to evolutionary pressures that do not affect lab evolved ones although analysis sometimes finds differences between natural and designed repeat sequences (Parra et al., 2015) but not always (Espada et al., 2017). Many designed Ankyrin repeat proteins (DARPs) are generated by random mutation followed by screening so mutations that disrupt protein folding are effectively invisible by virtue of not appearing in the screens (Urvoas et al., 2010; Kummer et al., 2013; Pluckthun, 2015; Schütz et al., 2016).

For integrin-linked kinase, a mutation in a conserved position was detrimental to binding in a pull down assay but one in a tolerant position was not, as expected (Chiswell et al., 2010; **Table 2**). In the regulatory protein RFXANK, group mutation of the hairpin positions (1, 2, 32, and 33) eliminated glutathione S-transferase binding in the first three but not the fourth repeat, and had no effect on binding to the class II transactivator (Nekrep et al., 2001). Some mutations in conserved positions did not hamper function but mutations in canonical positions in repeat 3 did. Double alanine hairpin loop mutations on any single one of the 24 Ankyrin B repeats did not significantly harm binding to the inositol 1,4,5-trisphosphate receptor, but these mutations in repeat 24 in combination with those in any one of repeats 19–22 did, although the exact positions of these mutations were not fully specified (Mohler et al., 2004). Double mutation of a pair of charged conserved positions to alanine in Ankyrin B abolished its interaction with its own C-terminal membrane binding domain (Abdi et al., 2006). A review of Notch mutations by Lubman et al. (2004) reports that a set of mutations at both canonical and conserved positions as well as a few in the terminal seventh, poorly folded repeat were completely disruptive to function, while other mutations in all three positional classes had a milder effect. In the viral K1 protein, mutations at 20 conserved positions in repeat 2, and both single and combination mutations covering much of the repeat length stopped viral replication in HeLa cells (Li et al., 2010). In the VPS9-ankyrin-repeat protein, mutations at several conserved and a tolerant position eliminated activity, but other similar mutations did not (Tamura et al., 2011). Mutations in the human DHH17 palmitoyl transferase at conserved and tolerant positions prevented ligand binding but two in hairpin positions did not (Verardi et al., 2017). Mutations in the kinesin family protein 21A (KIF21A) AK1 or ANK2 domains in hairpin positions had a moderate (up to 30 fold) effect on binding affinity (as well as some in tolerant and conserved positions) as measured by ITC. These interactions were confirmed by mutation in the peptide ligand (Guo et al., 2018). Simultaneous mutations to alanine at canonical positions of human IκBα always gave soluble protein but did affect complex formation and were most detrimental to activity when in the third repeat (Inoue et al., 1992). Analysis of several cancer-associated mutations in human p16 showed that mutations in all positional three position classes had detectable kinase binding defects and failed to inhibit cell

growth (Yarbrough et al., 1999). In gankyrin, almost every mutation tried noticeably affected protein folding by altering a complex folding pathway in which the first three repeats were likely to fold before the four C-terminal repeats. Non-classical  $\Phi$ -values were also observed for mutations in repeats 1, 2, 3, and 7 in some cases (Hutton et al., 2015). Mutations in the Ankyrin domain of TRPV4 that were known to cause genetic disorders were collected in a book chapter by Kang and were distributed among conserved and tolerant positions (Kang, 2012)

Ankyrin repeat proteins have more canonically-defined, highly conserved positions than TPR proteins (**Figure 1**). Curiously, they seem to also exhibit functional delocalization similarly to the TPR repeats (Parra et al., 2015). Functional delocalization is observed when mutations at any single position are compensated by a distributed set of other functional interactions present in the protein [e.g. single point mutations are non-debilitating; in one study 83% of randomly picked in-frame DARPs could be purified as monomers (Seeger et al., 2013) while another found a much weaker sequence restriction in TPR than Ankyrin repeat proteins (Parra et al., 2015; Kumar and Balbach, 2021)]. This delocalization would be in contrast to the majority of functional contribution being generally localized in the catalytic residues in an enzyme active site (Carter and Wells, 1988). For example, it was found that phosphorylation at no single residue was responsible for activity in the yeast Ankyrin repeat protein Pho81p and a greater number of mutations increased activity in some instances (Knight et al., 2004). Distributed weak interactions were also suggested by mutations in Notch when complex formation was eliminated by charge reversing mutations although mutation of other charged conserved residues was neutral, as was the removal of a tryptophan (Del Bianco et al., 2008). Additionally, in the TPR-containing Fanconi anemia FANCG protein, mutations at canonical position 8 were detrimental to activity in three repeats and a mutation at position 7 was so in a fourth. (Blom et al., 2004). A mutation at a tolerant position did not harm activity although the binding interaction was suggested to be widely distributed over the entire TPR domain rather than linked to any specific repeat (Wang and Lambert, 2010). Likewise, in human O-GlcNAc transferase, a combined set of mutations at tolerant positions 30 and 33 greatly reduced but did not eliminate activity despite accounting for the bulk of interactions between the TPR domain and substrate (Rafie et al., 2017).

## CONCLUSION

Repeat proteins tend to be organized by local interactions rather than long distance ones (Main et al., 2005) and positions involved in inter-repeat interactions tend to be more sequence restricted (Parra et al., 2015). Despite the obvious sequence and structural similarities between individual copies of repeats, they can be functionally distinct in that mutations at the same position in different repeats within the same protein are not equivalent. Much like non-repeat proteins, mutations that perturb the structure (the canonical positions which are involved in inter-repeat interactions) are often highly

disruptive, but due to the delocalization of structural and functional residues in repeat proteins, many of these mutations do not destroy the function of the protein and it is possible to mutate the canonical residues or even delete entire repeats without disrupting protein function. Studies of mutations in repeats will also be aided by adoption of a standardized positional numbering system of some sort (Han et al., 2011) and we recommend that future reports attempt to do so.

## AUTHOR CONTRIBUTIONS

MM wrote the manuscript, MI and PS designed figures and edited the manuscript, MG obtained funding and edited the manuscript.

## REFERENCES

- Abdi, K. M., Mohler, P. J., Davis, J. Q., and Bennett, V. (2006). Isoform Specificity of Ankyrin-B. *J. Biol. Chem.* 281 (9), 5741–5749. doi:10.1074/jbc.m506697200
- Andrade, M. A., Perez-Iratxeta, C., and Ponting, C. P. (2001). Protein Repeats: Structures, Functions, and Evolution. *J. Struct. Biol.* 134 (2-3), 117–131. doi:10.1006/jsbi.2001.4392
- Andrade, M. A., Petosa, C., O'Donoghue, S. I., Müller, C. W., and Bork, P. (2001). Comparison of ARM and HEAT Protein Repeats. *J. Mol. Biol.* 309 (1), 1–18. doi:10.1006/jmbi.2001.4624
- Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., et al. (2021). UniProt: the Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* 49 (D1), D480–D489. doi:10.1093/nar/gkaa1100
- Bidlingmaier, S., Ha, K., Lee, N.-K., Su, Y., and Liu, B. (2016). Proteome-wide Identification of Novel Ceramide-Binding Proteins by Yeast Surface cDNA Display and Deep Sequencing. *Mol. Cell Proteomics* 15 (4), 1232–1245. doi:10.1074/mcp.m115.055954
- Blom, E., van de Vrugt, H. J., Vries, Y. d., de Winter, J. P., Arwert, F., and Joenje, H. (2004). Multiple TPR Motifs Characterize the Fanconi Anemia FANCG Protein. *DNA Repair* 3 (1), 77–84. doi:10.1016/j.dnarep.2003.09.007
- Boisson, B., Puel, A., Picard, C., and Casanova, J.-L. (2017). Human IκBα Gain of Function: a Severe and Syndromic Immunodeficiency. *J. Clin. Immunol.* 37 (5), 397–412. doi:10.1007/s10875-017-0400-z
- Broms, J. E., Edqvist, P. J., Forsberg, A., and Francis, M. S. (2006). Tetratricopeptide Repeats Are Essential for PchH Chaperone Function in *Pseudomonas aeruginosa* Type III Secretion. *Fems Microbiol. Lett.* 256 (1), 57–66. doi:10.1111/j.1574-6968.2005.00099.x
- Büttner, C. R., Sorg, I., Cornelis, G. R., Heinz, D. W., and Niemann, H. H. (2008). Structure of the *Yersinia enterocolitica* Type III Secretion Translocator Chaperone SycD. *J. Mol. Biol.* 375 (4), 997–1012. doi:10.1016/j.jmb.2007.11.009
- Carter, P., and Wells, J. A. (1988). Dissecting the Catalytic Triad of a Serine Protease. *Nature* 332 (6164), 564–568. doi:10.1038/332564a0
- Cervený, L., Strasková, A., Danková, V., Hartlova, A., Cecková, M., Staud, F., et al. (2013). Tetratricopeptide Repeat Motifs in the World of Bacterial Pathogens: Role in Virulence Mechanisms. *Infect. Immun.* 81 (3), 629–635. doi:10.1128/iai.01035-12
- Champion, E. A., Kundrat, L., Regan, L., and Baserga, S. J. (2009). A Structural Model for the HAT Domain of Utp6 Incorporating Bioinformatics and Genetics. *Protein Eng. Des. Selection* 22 (7), 431–439. doi:10.1093/protein/gzp022
- Chiswell, B. P., Stiegler, A. L., Razinia, Z., Nalibotski, E., Boggon, T. J., and Calderwood, D. A. (2010). Structural Basis of Competition between PINCH1 and PINCH2 for Binding to the Ankyrin Repeat Domain of Integrin-Linked Kinase. *J. Struct. Biol.* 170 (1), 157–163. doi:10.1016/j.jsb.2009.12.002
- Clarke, A. J., Hurtado-Guerrero, R., Pathak, S., Schüttelkopf, A. W., Borodkin, V., Shepherd, S. M., et al. (2008). Structural Insights into Mechanism and Specificity of O-GlcNAc Transferase. *Embo J.* 27 (20), 2780–2788. doi:10.1038/emboj.2008.186

## FUNDING

The work was supported by the National Science Centre, Poland (grant agreement 2014/15/D/NZ1/00968) and EMBO Installation Grant to MG, who is the recipient of a L'Oréal-UNESCO For Women in Science scholarship from L'Oréal Poland.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2021.696368/full#supplementary-material>

- Cushing, D. A., Forsthoefel, N. R., Gestaut, D. R., and Vernon, D. M. (2005). Arabidopsis Emb175 and Other Ppr Knockout Mutants Reveal Essential Roles for Pentatricopeptide Repeat (PPR) Proteins in Plant Embryogenesis. *Planta* 221 (3), 424–436. doi:10.1007/s00425-004-1452-x
- D'Andrea, L., and Regan, L. (2003). TPR Proteins: the Versatile helix. *Trends Biochem. Sci.* 28 (12), 655–662. doi:10.1016/j.tibs.2003.10.007
- Del Bianco, C., Aster, J. C., and Blacklow, S. C. (2008). Mutational and Energetic Studies of Notch1 Transcription Complexes. *J. Mol. Biol.* 376 (1), 131–140. doi:10.1016/j.jmb.2007.11.061
- Desrosiers, D. C., and Peng, Z.-y. (2005). A Binding Free Energy Hot Spot in the Ankyrin Repeat Protein GABPβ Mediated Protein-Protein Interaction. *J. Mol. Biol.* 354 (2), 375–384. doi:10.1016/j.jmb.2005.09.045
- Edqvist, P. J., Broms, J. E., Betts, H. J., Forsberg, A., Pallen, M. J., and Francis, M. S. (2006). Tetratricopeptide Repeats in the Type III Secretion Chaperone, LcrH: Their Role in Substrate Binding and Secretion. *Mol. Microbiol.* 59 (1), 31–44. doi:10.1111/j.1365-2958.2005.04923.x
- Espada, R., Parra, R. G., Mora, T., Walczak, A. M., and Ferreira, D. U. (2017). Inferring Repeat-Protein Energetics from Evolutionary Information. *Plos Comput. Biol.* 13 (6), e1005584. doi:10.1371/journal.pcbi.1005584
- Espada, R., Parra, R. G., Sippl, M. J., Mora, T., Walczak, A. M., and Ferreira, D. U. (2015). Repeat Proteins challenge the Concept of Structural Domains. *Biochem. Soc. Trans.* 43, 844–849. doi:10.1042/bst20150083
- Ferreiro, D. U., Hegler, J. A., Komives, E. A., and Wolynes, P. G. (2007). Localizing Frustration in Native Proteins and Protein Assemblies. *Proc. Natl. Acad. Sci.* 104 (50), 19819–19824. doi:10.1073/pnas.0709915104
- Galpern, E. A., Freiburger, M. I., and Ferreira, D. U. (2020). Large Ankyrin Repeat Proteins Are Formed with Similar and Energetically Favorable Units. *Plos One* 15 (6), e0233865. doi:10.1371/journal.pone.0233865
- Geiger-Schuller, K., Sforza, K., Yuhas, M., Parmeggiani, F., Baker, D., and Barrick, D. (2018). Extreme Stability in De Novo-designed Repeat Arrays Is Determined by Unusually Stable Short-Range Interactions. *Proc. Natl. Acad. Sci. USA* 115 (29), 7539–7544. doi:10.1073/pnas.1800283115
- Grenha, R., Slamti, L., Nicaise, M., Refes, Y., Lereclus, D., and Nessler, S. (2013). Structural Basis for the Activation Mechanism of the PlcR Virulence Regulator by the Quorum-sensing Signal Peptide PapR. *Proc. Natl. Acad. Sci.* 110 (3), 1047–1052. doi:10.1073/pnas.1213770110
- Guo, Q., Liao, S., Zhu, Z., Li, Y., Li, F., and Xu, C. (2018). Structural Basis for the Recognition of Kinesin Family Member 21A (KIF21A) by the Ankyrin Domains of KANK1 and KANK2 Proteins. *J. Biol. Chem.* 293 (2), 557–566. doi:10.1074/jbc.m117.817494
- Han, Y., Noam, Y., Lewis, A. S., Gallagher, J. J., Wadman, W. J., Baram, T. Z., et al. (2011). Trafficking and Gating of Hyperpolarization-Activated Cyclic Nucleotide-Gated Channels Are Regulated by Interaction with Tetratricopeptide Repeat-Containing Rab8b-Interacting Protein (TRIP8b) and Cyclic AMP at Distinct Sites. *J. Biol. Chem.* 286 (23), 20823–20834. doi:10.1074/jbc.m111.236125
- Houlihan, G., Gatti-Lafranconi, P., Lowe, D., and Hoffelder, F. (2015). Directed Evolution of Anti-HER2 DARPins by SNAP Display Reveals Stability/function



- Trade-Offs in the Selection Process. *Protein Eng. Des. Selection* 28 (9), 269–279. doi:10.1093/protein/gzv029
- Hutton, R. D., Wilkinson, J., Faccin, M., Sivertsson, E. M., Pelizzola, A., Lowe, A. R., et al. (2015). Mapping the Topography of a Protein Energy Landscape. *J. Am. Chem. Soc.* 137 (46), 14610–14625. doi:10.1021/jacs.5b07370
- Iakhiaeva, E., Hinck, C. S., Hinck, A. P., and Zwieb, C. (2009). Characterization of the SRP68/72 Interface of Human Signal Recognition Particle by Systematic Site-Directed Mutagenesis. *Protein Sci.* 18 (10), 2183–2195. doi:10.1002/pro.232
- Inoue, J., Kerr, L. D., Rashid, D., Davis, N., Bose, H. R., and Verma, I. M. (1992). Direct Association of Pp40/I Kappa B Beta with rel/NF-Kappa B Transcription Factors: Role of Ankyrin Repeats in the Inhibition of DNA Binding Activity. *Proc. Natl. Acad. Sci.* 89 (10), 4333–4337. doi:10.1073/pnas.89.10.4333
- Jernigan, K. K., and Bordenstein, S. R. (2014). Ankyrin Domains across the Tree of Life. *PeerJ* 2, e264. doi:10.7717/peerj.264
- Jernigan, K. K., and Bordenstein, S. R. (2015). Tandem-repeat Protein Domains across the Tree of Life. *PeerJ* 3, e732. doi:10.7717/peerj.732
- Jorda, J., Xue, B., Uversky, V. N., and Kajava, A. V. (2010). Protein Tandem Repeats - the More Perfect, the Less Structured. *Febs J.* 277 (12), 2673–2682. doi:10.1111/j.1742-4658.2010.07684.x
- Kajander, T., Sachs, J. N., Goldman, A., and Regan, L. (2009). Electrostatic Interactions of Hsp-Organizing Protein Tetratricopeptide Domains with Hsp70 and Hsp90. *J. Biol. Chem.* 284 (37), 25364–25374. doi:10.1074/jbc.m109.033894
- Kajava, A. V. (2001). Review: Proteins with Repeated Sequence - Structural Prediction and Modeling. *J. Struct. Biol.* 134 (2-3), 132–144. doi:10.1006/jbsi.2000.4328
- Kajava, A. V. (2012). Tandem Repeats in Proteins: From Sequence to Structure. *J. Struct. Biol.* 179 (3), 279–288. doi:10.1016/j.jsb.2011.08.009
- Kang, H., Sayner, S. L., Gross, K. L., Russell, L. C., and Chinkers, M. (2001). Identification of Amino Acids in the Tetratricopeptide Repeat and C-Terminal Domains of Protein Phosphatase 5 Involved in Autoinhibition and Lipid Activation. *Biochemistry* 40 (35), 10485–10490. doi:10.1021/bi010999i
- Kang, S. S. (2012). “The Mutation of Transient Receptor Potential Vanilloid 4 (TRPV4) Cation Channel in Human Diseases,” in *Mutagenesis*. Editor R. Mishra (Rijeka: IntechOpen). doi:10.5772/50520
- Karanicolas, J., Corn, J. E., Chen, I., Joachimiak, L. A., Dym, O., Peck, S. H., et al. (2011). A De Novo Protein Binding Pair by Computational Design and Directed Evolution. *Mol. Cel* 42 (2), 250–260. doi:10.1016/j.molcel.2011.03.010
- Katibah, G. E., Qin, Y., Sidote, D. J., Yao, J., Lambowitz, A. M., and Collins, K. (2014). Broad and Adaptable RNA Structure Recognition by the Human Interferon-Induced Tetratricopeptide Repeat Protein IIFT5. *Proc. Natl. Acad. Sci.* 111 (33), 12025–12030. doi:10.1073/pnas.1412842111
- Knight, J. P., Daly, T. M., and Bergman, L. W. (2004). Regulation by Phosphorylation of Pho81p, a Cyclin-dependent Kinase Inhibitor in *Saccharomyces cerevisiae*. *Curr. Genet.* 46 (1), 10–19. doi:10.1007/s00294-004-0502-z
- Kobayashi, K., Kawabata, M., Hisano, K., Kazama, T., Matsuoka, K., Sugita, M., et al. (2012). Identification and Characterization of the RNA Binding Surface of the Pentatricopeptide Repeat Protein. *Nucleic Acids Res.* 40 (6), 2712–2723. doi:10.1093/nar/gkr1084
- Koo, J., Tang, T., Harvey, H., Tammam, S., Sampaleanu, L., Burrows, L. L., et al. (2013). Functional Mapping of PilF and PilQ in the *Pseudomonas aeruginosa* Type IV Pilus System. *Biochemistry* 52 (17), 2914–2923. doi:10.1021/bi3015345
- Kumar, A., and Balbach, J. (2021). Folding and Stability of Ankyrin Repeats Control Biological Protein Function. *Biomolecules* 11, 840. doi:10.3390/biom11060840
- Kummer, L., Hsu, C.-W., Dagliyan, O., MacNevin, C., Kaufholz, M., Zimmermann, B., et al. (2013). Knowledge-Based Design of a Biosensor to Quantify Localized ERK Activation in Living Cells. *Chem. Biol.* 20 (6), 847–856. doi:10.1016/j.chembiol.2013.04.016
- Li, Y., Meng, X., Xiang, Y., and Deng, J. (2010). Structure Function Studies of Vaccinia Virus Host Range Protein K1 Reveal a Novel Functional Surface for Ankyrin Repeat Proteins. *J. Virol.* 84 (7), 3331–3338. doi:10.1128/jvi.02332-09
- Loizeau, K., Qu, Y., Depp, S., Fiechter, V., Ruwe, H., Lefebvre-Legendre, L., et al. (2014). Small RNAs Reveal Two Target Sites of the RNA-Maturation Factor Mbb1 in the Chloroplast of *Chlamydomonas*. *Nucleic Acids Res.* 42 (5), 3286–3297. doi:10.1093/nar/gkt1272
- Lovell, S. C., and Robertson, D. L. (2010). An Integrated View of Molecular Coevolution in Protein-Protein Interactions. *Mol. Biol. Evol.* 27 (11), 2567–2575. doi:10.1093/molbev/msq144
- Lubman, O. Y., Korolev, S. V., and Kopan, R. (2004). Anchoring Notch Genetics and Biochemistry. *Mol. Cel* 13 (5), 619–626. doi:10.1016/s1097-2765(04)00120-0
- Magliery, T. J., and Regan, L. (2004). Beyond Consensus: Statistical Free Energies Reveal Hidden Interactions in the Design of a TPR Motif. *J. Mol. Biol.* 343 (3), 731–745. doi:10.1016/j.jmb.2004.08.026
- Main, E., Lowe, A., Mochrie, S., Jackson, S., and Regan, L. (2005). A Recurring Theme in Protein Engineering: the Design, Stability and Folding of Repeat Proteins. *Curr. Opin. Struct. Biol.* 15 (4), 464–471. doi:10.1016/j.sbi.2005.07.003
- Main, E. R., Jackson, S. E., and Regan, L. (2003). The Folding and Design of Repeat Proteins: Reaching a Consensus. *Curr. Opin. Struct. Biol.* 13 (4), 482–489. doi:10.1016/s0959-440x(03)00105-2
- Marchi, J., Galpern, E. A., Espada, R., Ferreiro, D. U., Walczak, A. M., and Mora, T. (2019). Size and Structure of the Sequence Space of Repeat Proteins. *Plos Comput. Biol.* 15 (8), e1007282. doi:10.1371/journal.pcbi.1007282
- Marcotte, E. M., Pellegrini, M., Yeates, T. O., and Eisenberg, D. (1999). A Census of Protein Repeats. *J. Mol. Biol.* 293 (1), 151–160. doi:10.1006/jmbi.1999.3136
- Marold, J. D., Kavran, J. M., Bowman, G. D., and Barrick, D. (2015). A Naturally Occurring Repeat Protein with High Internal Sequence Identity Defines a New Class of TPR-like Proteins. *Structure* 23 (11), 2055–2065. doi:10.1016/j.str.2015.07.022
- Marold, J. D., Sforza, K., Geiger-Schuller, K., Aksel, T., Klein, S., Petersen, M., et al. (2021). A Collection of Programs for One-dimensional Ising Analysis of Linear Repeat Proteins with point Substitutions. *Protein Sci.* 30 (1), 168–186. doi:10.1002/pro.3977
- Marte, A., Russo, I., Rebosio, C., Valente, P., Belluzzi, E., Pischedda, F., et al. (2019). Leucine-rich Repeat Kinase 2 Phosphorylation on Synapsin I Regulates Glutamate Release at Pre-synaptic Sites. *J. Neurochem.* 150 (3), 264–281. doi:10.1111/jnc.14778
- Mello, C. C., Bradley, C. M., Tripp, K. W., and Barrick, D. (2005). Experimental Characterization of the Folding Kinetics of the Notch Ankyrin Domain (Vol 352, Pg 266, 2005). *J. Mol. Biol.* 353 (5), 1210. doi:10.1016/j.jmb.2005.07.026
- Michaely, P., and Bennett, V. (1993). The Membrane-Binding Domain of Ankyrin Contains Four Independently Folded Subdomains, Each Comprised of Six Ankyrin Repeats. *J. Biol. Chem.* 268 (30), 22703–22709. doi:10.1016/s0021-9258(18)41584-0
- Mohler, P. J., Davis, J. Q., Davis, L. H., Hoffman, J. A., Michaely, P., and Bennett, V. (2004). Inositol 1,4,5-trisphosphate Receptor Localization and Stability in Neonatal Cardiomyocytes Requires Interaction with Ankyrin-B. *J. Biol. Chem.* 279 (13), 12980–12987. doi:10.1074/jbc.m313979200
- Mosavi, L. K., Minor, D. L., and Peng, Z.-y. (2002). Consensus-derived Structural Determinants of the Ankyrin Repeat Motif. *Proc. Natl. Acad. Sci.* 99 (25), 16029–16034. doi:10.1073/pnas.252537899
- Mosavi, L. K., and Peng, Z.-y. (2003). Structure-based Substitutions for Increased Solubility of a Designed Protein. *Protein Eng. Des. Selection* 16 (10), 739–745. doi:10.1093/protein/gzg098
- Nekrep, N., Geyer, M., Jabrane-Ferrat, N., and Peterlin, B. M. (2001). Analysis of Ankyrin Repeats Reveals How a Single point Mutation in RFXANK Results in Bare Lymphocyte Syndrome. *Mol. Cel Biol.* 21 (16), 5566–5576. doi:10.1128/mcb.21.16.5566-5576.2001
- Paladin, L., Bevilacqua, M., Errigo, S., Piovesan, D., Mičetić, I., Necci, M., et al. (2021). RepeatsDB in 2021: Improved Data and Extended Classification for Protein Tandem Repeat Structures. *Nucleic Acids Res.* 49 (D1), D452–D457. doi:10.1093/nar/gkaa1097
- Paladin, L., Hirsh, L., Piovesan, D., Andrade-Navarro, M. A., Kajava, A. V., and Tosatto, S. C. E. (2017). RepeatsDB 2.0: Improved Annotation, Classification, Search and Visualization of Repeat Protein Structures. *Nucleic Acids Res.* 45 (6), 3613. doi:10.1093/nar/gkw1268
- Pallen, M. J., Francis, M. S., and Fütterer, K. (2003). Tetratricopeptide-like Repeats in Type-III-Secretion Chaperones and Regulators. *FEMS Microbiol. Lett.* 223 (1), 53–60. doi:10.1016/S0378-1097(03)00344-6
- Parra, R. G., Espada, R., Sánchez, I. E., Sippl, M. J., and Ferreiro, D. U. (2013). Detecting Repetitions and Periodicities in Proteins by Tiling the Structural Space. *J. Phys. Chem. B* 117 (42), 12887–12897. doi:10.1021/jp402105j
- Parra, R. G., Espada, R., Verstraete, N., and Ferreiro, D. U. (2015). Structural and Energetic Characterization of the Ankyrin Repeat Protein Family. *Plos Comput. Biol.* 11 (12), e1004659. doi:10.1371/journal.pcbi.1004659

- Pekkala, M., Hieta, R., Bergmann, U., Kivirikko, K. I., Wierenga, R. K., and Myllyharju, J. (2004). The Peptide-Substrate-Binding Domain of Collagen Prolyl 4-hydroxylases Is a Tetratricopeptide Repeat Domain with Functional Aromatic Residues. *J. Biol. Chem.* 279 (50), 52255–52261. doi:10.1074/jbc.m410007200
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera?A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 25 (13), 1605–1612. doi:10.1002/jcc.20084
- Plückthun, A. (2015). Designed Ankyrin Repeat Proteins (DARPs): Binding Proteins for Research, Diagnostics, and Therapy. *Annu. Rev. Pharmacol. Toxicol.* 55 (55), 489–511. doi:10.1146/annurev-pharmtox-010611-134654
- Preker, P. J., and Keller, W. (1998). The HAT helix, a Repetitive Motif Implicated in RNA Processing. *Trends Biochem. Sci.* 23 (1), 15–16. doi:10.1016/s0968-0004(97)01156-0
- Rafie, K., Raimi, O., Ferenbach, A. T., Borodkin, V. S., Kapuria, V., and van Aalten, D. M. F. (2017). Recognition of a Glycosylation Substrate by the O-GlcNAc Transferase TPR Repeats. *Open Biol.* 7 (6), 170078. doi:10.1098/rsob.170078
- Sampathkumar, P., Roach, C., Michels, P. A. M., and Hol, W. G. J. (2008). Structural Insights into the Recognition of Peroxisomal Targeting Signal 1 by Trypanosoma Brucei Peroxin 5. *J. Mol. Biol.* 381 (4), 867–880. doi:10.1016/j.jmb.2008.05.089
- Schaper, E., Gascuel, O., and Anisimova, M. (2014). Deep Conservation of Human Protein Tandem Repeats within the Eukaryotes. *Mol. Biol. Evol.* 31 (5), 1132–1148. doi:10.1093/molbev/msu062
- Schaper, E., Kajava, A. V., Hauser, A., and Anisimova, M. (2012). Repeat or Not Repeat?—Statistical Validation of Tandem Repeat Prediction in Genomic Sequences. *Nucleic Acids Res.* 40 (20), 10005–10017. doi:10.1093/nar/gks726
- Schüler, A., and Bornberg-Bauer, E. (2016). Evolution of Protein Domain Repeats in Metazoa. *Mol. Biol. Evol.* 33 (12), 3170–3182. doi:10.1093/molbev/msw194
- Schütz, M., Batyuk, A., Klenk, C., Kummer, L., de Picciotto, S., Gülbakan, B., et al. (2016). Generation of Fluorogen-Activating Designed Ankyrin Repeat Proteins (FADAs) as Versatile Sensor Tools. *J. Mol. Biol.* 428 (6), 1272–1289. doi:10.1016/j.jmb.2016.01.017
- Sedgwick, S. G., and Smerdon, S. J. (1999). The Ankyrin Repeat: a Diversity of Interactions on a Common Structural Framework. *Trends Biochem. Sci.* 24 (8), 311–316. doi:10.1016/s0968-0004(99)01426-7
- Seeger, M. A., Zbinden, R., Flütsch, A., Gutte, P. G. M., Engeler, S., Roschitzki-Voser, H., et al. (2013). Design, Construction, and Characterization of a Second-Generation DARPin Library with Reduced Hydrophobicity. *Protein Sci.* 22 (9), 1239–1257. doi:10.1002/pro.2312
- Sengupta, D., and Kundu, S. (2012). Role of Long- and Short-Range Hydrophobic, Hydrophilic and Charged Residues Contact Network in Protein's Structural Organization. *Bmc Bioinformatics* 13, 142. doi:10.1186/1471-2105-13-142
- Severi, E., Müller, A., Potts, J. R., Leech, A., Williamson, D., Wilson, K. S., et al. (2008). Sialic Acid Mutarotation Is Catalyzed by the *Escherichia coli*  $\beta$ -Propeller Protein YjhT. *J. Biol. Chem.* 283 (8), 4841–4849. doi:10.1074/jbc.m707822200
- Stanley, L. E., Ding, B., Sun, W., Mou, F., Hill, C., Chen, S., et al. (2020). A Tetratricopeptide Repeat Protein Regulates Carotenoid Biosynthesis and Chromoplast Development in Monkeyflowers (*Mimulus*). *Plant Cell* 32 (5), 1536–1555. doi:10.1105/tpc.19.00755
- Stumpp, M. T., Forrer, P., Binz, H. K., and Plückthun, A. (2003). Designing Repeat Proteins: Modular Leucine-Rich Repeat Protein Libraries Based on the Mammalian Ribonuclease Inhibitor Family. *J. Mol. Biol.* 332 (2), 471–487. doi:10.1016/s0022-2836(03)00897-0
- Tamura, K., Ohbayashi, N., Ishibashi, K., and Fukuda, M. (2011). Structure-Function Analysis of VPS9-Ankyrin-Repeat Protein (Varp) in the Trafficking of Tyrosinase-Related Protein 1 in Melanocytes. *J. Biol. Chem.* 286 (9), 7507–7521. doi:10.1074/jbc.m110.191205
- Tramontano, A., and Cozzetto, D. (2005). “The Relationship between Protein Sequence, Structure and Function,” in *Supramolecular Structure and Function*. Editor P.-M. G (Boston, MA: Springer).
- Travers, S. A. A., and Fares, M. A. (2007). Functional Coevolutionary Networks of the Hsp70-Hop-Hsp90 System Revealed through Computational Analyses. *Mol. Biol. Evol.* 24 (4), 1032–1044. doi:10.1093/molbev/msm022
- Tripp, K. W., and Barrick, D. (2004). The Tolerance of a Modular Protein to Duplication and Deletion of Internal Repeats. *J. Mol. Biol.* 344 (1), 169–178. doi:10.1016/j.jmb.2004.09.038
- Turjanski, P., Parra, R. G., Espada, R., Becher, V., and Ferreira, D. U. (2016). Protein Repeats from First Principles. *Scientific Rep.* 6, 23959. doi:10.1038/srep23959
- Urvoas, A., Guellouz, A., Valerio-Lepiniec, M., Graille, M., Durand, D., Desravines, D. C., et al. (2010). Design, Production and Molecular Structure of a New Family of Artificial Alpha-Helical Repeat Proteins ( $\alpha$ Rep) Based on Thermostable HEAT-like Repeats. *J. Mol. Biol.* 404 (2), 307–327. doi:10.1016/j.jmb.2010.09.048
- Vander Kooi, C. W., Ren, L., Xu, P., Ohi, M. D., Gould, K. L., and Chazin, W. J. (2010). The Prp19 WD40 Domain Contains a Conserved Protein Interaction Region Essential for its Function. *Structure* 18 (5), 584–593. doi:10.1016/j.str.2010.02.015
- Verardi, R., Kim, J.-S., Ghirlando, R., and Banerjee, A. (2017). Structural Basis for Substrate Recognition by the Ankyrin Repeat Domain of Human DHHC17 Palmitoyltransferase. *Structure* 25 (9), 1337–1347. doi:10.1016/j.str.2017.06.018
- Wall, M. A., Coleman, D. E., Lee, E., Iñiguez-Lluhi, J. A., Posner, B. A., Gilman, A. G., et al. (1995). The Structure of the G Protein Heterotrimer  $\text{G}\alpha 1\beta 1\gamma 2$ . *Cell* 83 (6), 1047–1058. doi:10.1016/0092-8674(95)90220-1
- Wang, C., and Lambert, M. W. (2010). The Fanconi Anemia Protein, FANCG, Binds to the ERCC1-XPF Endonuclease via its Tetratricopeptide Repeats and the Central Domain of ERCC1. *Biochemistry* 49 (26), 5560–5569. doi:10.1021/bi100584c
- Ward, B. K., Allan, R. K., Mok, D., Temple, S. E., Taylor, P., Dornan, J., et al. (2002). A Structure-Based Mutational Analysis of Cyclophilin 40 Identifies Key Residues in the Core Tetratricopeptide Repeat Domain that Mediate Binding to Hsp90. *J. Biol. Chem.* 277 (43), 40799–40809. doi:10.1074/jbc.m207097200
- Wittwer, M., and Dames, S. A. (2015). Expression and Purification of the Natively Disordered and Redox Sensitive Metal Binding Regions of *Mycobacterium tuberculosis* Protein Kinase G. *Protein Expr. Purif.* 111, 68–74. doi:10.1016/j.pep.2015.03.015
- Wu, S.-J., Liu, F.-H., Hu, S.-M., and Wang, C. (2001). Different Combinations of the Heat-Shock Cognate Protein 70 (Hsc70) C-Terminal Functional Groups Are Utilized to Interact with Distinct Tetratricopeptide Repeat-Containing Proteins. *Biochem. J.* 359, 419–426. doi:10.1042/bj3590419
- Yang, Y. D., Gao, J. Z., Wang, J. H., Heffernan, R., Hanson, J., Paliwal, K., et al. (2018). Sixty-five Years of the Long March in Protein Secondary Structure Prediction: the Final Stretch? *Brief. Bioinform.* 19 (3), 482–494. doi:10.1093/bib/bbw129
- Yang, Z., Liang, H., Zhou, Q., Li, Y., Chen, H., Ye, W., et al. (2012). Crystal Structure of ISG54 Reveals a Novel RNA Binding Structure and Potential Functional Mechanisms. *Cell Res.* 22 (9), 1328–1338. doi:10.1038/cr.2012.111
- Yarbrough, W. G., Buckmire, R. A., Bessho, M., and Liu, E. T. (1999). Biologic and Biochemical Analyses of p16INK4a Mutations from Primary Tumors. *JNCI J. Natl. Cancer Inst.* 91 (18), 1569–1574. doi:10.1093/jnci/91.18.1569
- Yuzawa, S., Kamakura, S., Iwakiri, Y., Hayase, J., and Sumimoto, H. (2011). Structural Basis for Interaction between the Conserved Cell Polarity Proteins Inscuteable and Leu-Gly-Asn Repeat-Enriched Protein (LGN). *Proc. Natl. Acad. Sci.* 108 (48), 19210–19215. doi:10.1073/pnas.1110951108
- Zhang, X., Hoey, R. J., Lin, G., Koide, A., Leung, B., Ahn, K., et al. (2012). Identification of a Tetratricopeptide Repeat-like Domain in the Nicastrin Subunit of -secretase Using Synthetic Antibodies. *Proc. Natl. Acad. Sci.* 109 (22), 8534–8539. doi:10.1073/pnas.1202691109
- Zhao, G., Li, G., Schindelin, H., and Lennarz, W. J. (2009). An Armadillo Motif in Ufd3 Interacts with Cdc48 and Is Involved in Ubiquitin Homeostasis and Protein Degradation. *Proc. Natl. Acad. Sci.* 106 (38), 16197–16202. doi:10.1073/pnas.0908321106
- Zhu, H. B., Sepulveda, E., Hartmann, M. D., Kogenaru, M., Ursinus, A., Sulz, E., et al. (2016). Origin of a Folded Repeat Protein from an Intrinsically Disordered Ancestor. *Elife* 5, e16761. doi:10.7554/elife.16761

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Izert, Szybowska, Górna and Merski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.