A novel predicted ADP-ribosyltransferase-like family conserved in eukaryotic evolution

Zbigniew Wyżewski[1,*], Marcin Gradowski[2,*], Marianna Krysińska[2], Małgorzata Dudkiewicz[2] and Krzysztof Pawłowski[2,3,4]

[1] Institute of Biological Sciences, Cardinal Stefan Wyszynski University in Warsaw, Warszawa, Poland
[2] Department of Biochemistry and Microbiology, Warsaw University of Life Sciences - SGGW, Warszawa, Poland
[3] Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX, United States
[4] Department of Translational Medicine, Lund University, Lund, Sweden
[*] These authors contributed equally to this work.

ABSTRACT

The presence of many completely uncharacterized proteins, even in well-studied organisms such as humans, seriously hampers full understanding of the functioning of the living cells. ADP-ribosylation is a common post-translational modification of proteins; also nucleic acids and small molecules can be modified by the covalent attachment of ADP-ribose. This modification, important in cellular signalling and infection processes, is usually executed by enzymes from the large superfamily of ADP-ribosyltransferases (ARTs). Here, using bioinformatics approaches, we identify a novel putative ADP-ribosyltransferase family, conserved in eukaryotic evolution, with a divergent active site. The hallmark of these proteins is the ART domain nestled between flanking leucine-rich repeat (LRR) domains. LRRs are typically involved in innate immune surveillance. The novel family appears as putative novel ADP-ribosylation-related actors, most likely pseudoenzymes. Sequence divergence and lack of clearly detectable "classical" ART active site suggests the novel domains are pseudoARTs, yet atypical ART activity, or alternative enzymatic activity cannot be excluded. We propose that this family, including its human member LRRC9, may be involved in an ancient defense mechanism, with analogies to the innate immune system, and coupling pathogen detection to ADP-ribosyltransfer or other signalling mechanisms.

Subjects Biochemistry, Bioinformatics, Biophysics, Evolutionary Studies, Molecular Biology
Keywords ADP-ribosyltransferases, Evolution, Protein domains, Pseudoenzymes, Protein structure and function prediction

INTRODUCTION

ADP-ribosyltransferases (ARTs) are enzymes catalyzing the transfer of ADP-ribose from oxidized nicotinamide adenine dinucleotide (NAD^+) to different acceptor molecules. Thus, they are responsible for chemical modification of various targets such as proteins, nucleic acids and small molecules. These enzymes are highly conserved in evolution and widespread in nature. They are common to all three domains of life: the Archaea, the Bacteria and the Eukarya. Multiplicity of ADP-ribosylated substrates is reflected in a variety

Submitted 24 July 2020
Accepted 11 February 2021
Published 10 March 2021

Corresponding authors
Małgorzata Dudkiewicz, malgorzata.dudkiewicz@gmail.com
Krzysztof Pawłowski, Krzysztof_Pawlowski@sggw.edu.pl, Krzysztof.Pawlowski@utsw.edu

Academic editor
Joseph Gillespie

Additional Information and Declarations can be found on page 20

DOI **10.7717/peerj.11051**

of functions performed by ARTs (*Aravind et al., 2015*; *Cohen & Chang, 2018*; *Munnur & Ahel, 2017*; *Otto et al., 2005*; *Teloni & Altmeyer, 2016*).

Prokaryotic ADP-ribosyltransferases often play a role of bacterial toxins or effectors. Several studies show that ADP-ribosylation of the host target molecules by bacterial ARTs is correlated with the development of infection (*Aktories et al., 2011*; *Bhogaraju et al., 2016*; *Boyer et al., 2006*; *Kalayil et al., 2018*; *Simon, Aktories & Barbieri, 2014*). At the intracellular scale, ADP-ribosylation may change apoptotic potential of infected cell (*Klockgether & Tümmler, 2017*) and/or disturb organization of cellular membranes and actin cytoskeleton (*Kagan & Roy, 2002*; *Komander & Randow, 2017*; *Maresso et al., 2007*) whereas at systemic level, it can disturb cell-mediated immune response (*Bhogaraju et al., 2016*; *Klockgether & Tümmler, 2017*), or increase the permeability of barriers limiting bacterial spread (epithelium and endothelium) (*Boyer et al., 2006*; *Klockgether & Tümmler, 2017*) and consequently contribute to serious structural and functional disorders of the host tissues and organs (*Chiu et al., 2009*; *Munro et al., 2010*). Some ARTs act on small molecules, e.g., the bacterial Arr enzymes that ADP-ribosylate the antibiotic rifamycin (*Baysarowich et al., 2008*).

Eukaryotic ADP-ribosyltransferases play important roles in both physiological and pathophysiological processes, contributing to changes in chemical properties of the ADP-ribose acceptors. Proteins, post-translationally modified by ARTs, can also lose the capacity to interact with their ligands or acquire the ability to bind new ones. ADP-ribosylation of enzymes may cause substantial changes in their catalytic activity (*Jwa & Chang, 2012*; *Liu & Yu, 2015*). ADP-ribose moieties may function as signals that direct modified acceptors to ubiquitin-dependent proteolysis (*Aravind et al., 2015*; *Cohen & Chang, 2018*). Influencing the half-life of the target molecules, ADP-ribosyltransferases determine their intracellular level and thus affect their activity (*Cohen & Chang, 2018*). Moreover, ADP-ribose moieties may form molecular scaffolds with a negative charge. Such structures play a role in recruitment of positively charged proteins, favoring specific intermolecular interactions (*Leung et al., 2012*). Poly-ADP-ribosylation of DNA-binding proteins (e.g., core histones (*Javle & Curtin, 2011*; *Weaver & Yang, 2013*), chromatin remodeling enzymes (e.g., histone demethylase KDM5B (*Krishnakumar & Kraus, 2010*), DEK protein (*Gamble & Fisher, 2007*) and DNA repair factors (*Kim et al., 2015*) influences genome organization (*Weaver & Yang, 2013*) and expression (*Gamble & Fisher, 2007*; *Krishnakumar & Kraus, 2010*) as well as effective DNA repair (*Javle & Curtin, 2011*; *McCann, 2019*). Recently, growing evidence shows that PARPs also directly ADP-ribosylate mRNA (*Kim et al., 2020*). Other examples of nucleic acid ADP-ribosylation are provided by the bacterial toxin DarT that acts on ssDNA (*Lawaree et al., 2020*) and the toxin scabin, acting on mononucleosides, nucleotides, and both single-stranded and dsDNA (*Lyons et al., 2018*).

The members of the poly-ADP-ribosyltransferase (PARP) family, PARP1 and PARP2, are examples of ARTs that modify proteins interacting with DNA. PARP1 and PARP2 are localized in the nucleus where they are involved in many cellular processes (*Choi et al., 2016*; *Liang et al., 2013*; *Riccio, Cingolani & Pascal, 2016*; *Szántó et al., 2014*). PARP1 and PARP2 are engaged in DNA repair, that is elimination of DNA damage such as deamination, hydroxylation and methylation as well as for repair of single strand breaks.

PARP1 and other PARPs also play a role in maintaining telomere stability (*Bai, 2015*; *De Vos, Schreiber & Dantzer, 2012*). PARPs are known as EMA/FDA approved drug targets, PARP inhibitors are used in treatment of prostate, breast and ovarian cancers. (*Curtin, 2005*; *Curtin & Szabo, 2013*; *Kamel et al., 2018*; *Kummar et al., 2012*).

In general, ADP-ribosyltransferases are responsible for regulation of intracellular and extracellular signal transduction. Therefore, their activity determines viability and proliferation potential of cells, DNA stability, immune system reactivity and thus the proper functioning of eukaryotic organisms (*Bai, 2015*; *Corda & Di Girolamo, 2002*; *De Vos, Schreiber & Dantzer, 2012*; *Yamamoto et al., 2017*). On the other hand, ADP-ribosyltransferases may also be involved in development of pathological states such as neurodegenerative disorders, diabetes, atherosclerosis, cataract and cancer (*Barreiro & Gea, 2018*; *Gunderson & Moore, 2015*; *Kamboj et al., 2013*; *Malyuchenko et al., 2015*; *Mangerich & Bürkle, 2012*; *Rossi, Ghosh & Bohr, 2010*). Similarly to protein phosphorylation, ADP-ribosylation is a reversible post-translational modification, performed by a trio of "writers" (ARTs), "readers" (e.g., macro domain proteins) and "erasers" (e.g., some macro domains, ADP-ribose hydrolases, and NUDIX phosphodiesterases) (*O'Sullivan et al., 2019*).

According to the Pfam database, the ART clan (superfamily) comprises 23 families of domains, fourteen of which, ADPrib_exo_Tox, ART, DarT, DUF2441, DUF3990, DUF952, Enterotoxin_a, Exotox-A_cataly, PARP, Pertussis_S1, PTS_2-RNA, RES, RolB_RolC and TNT, include eukaryotic members (*El-Gebali et al., 2019*). However, some of these 14 families are predominantly bacterial with just a handful of eukaryotic members, e.g. ADPrib_exo_Tox and DarT. PARPs, the best studied family of ADP-ribosyltransferases, are responsible for modification of target structures by covalently adding polymeric chains of ADP-ribose moieties instead of transferring only one moiety. Despite low conservation of amino acid sequence, ART clan members are characterized by a common spatial structure comprising a split β-sheet and two helical regions surrounding it. The "split" separates the β-sheet into two halves, each composed of three β-strands (4-5-2 and 1-3-6, respectively) (*Aravind et al., 2015*; *Cohen & Chang, 2018*). Aravind and colleagues divide ARTs into three main clades: the H-H-h clade, the H-Y-[EDQ] clade and the R-S-E clade (*Aravind et al., 2015*; *Cohen & Chang, 2018*) that are characterized by the different configurations of active site amino acid residues. In the H-H-h clade domains, the catalytic centre comprises two histidines and one hydrophobic residue supplied by β-strands 1, 2 and 5, respectively. The H-Y-[EDQ] clade domains, including PARPs, are characterized by active site composed of histidine, tyrosine and glutamate/aspartate/glutamine whereas in the R-[ST]-E clade domains, active site triad comprises arginine, a polar residue (serine/threonine) and glutamate (*Aravind et al., 2015*). Three families of the ART clan, PARP, PTS_2-RNA and ART, are present in humans (Table 1), and contain 16, 1 and 4 representatives, respectively. The NEURL-4 ART-like domain family (*De Souza & Aravind, 2012*), not included in Pfam database, is the fourth human ART-like family. The full catalogue of ADP-ribosylation enzymes is likely far from completion, as one may expect from recent discoveries of novel ART domains in effectors from pathogenic bacteria that perform non-canonical ubiquitination (*Akturk et al., 2018*; *Kalayil et al., 2018*; *Kim et al., 2018*), novel ART/macro pairs in bacterial toxin/antitoxin systems (*Jankevicius et al., 2016*) and novel human macro

Table 1 List of known human ART-like domains. Full list of ART-like domains from human, including the novel DUF3715 domains in TASOR, TASOR2 and TEX15.

Name	Gene names*	family classification for ART Domain (Pfam)		Triad motif	R/H-G-T/S motif in β-strand 1	S-X-S/Y-X-X motif in β-strand 2	X-X-E motif at front edge of β-strand 5	ADP-ribosylationactivity: mono (M), oligo (O), poly (P)	Cellular localization	ADP-ribosylation target: protein (P), DNA (D), RNA (R), and auto-ADP-ribosylation (A)	Total length (ART-like domain length)
TRPT1	TRPT1	PTS_2-RNA	PF01885	H-H-V	HGT	HLA	NGV	M	mitochondrion, nucleus, endoplasmic reticulum, cytosol	R	253 (181)
ART1	ART1	ART	PF01129	R-S-E	RGV	SAS	EEE	M	sarcoplasmic reticulum membrane, plasma membrane, ER membrane and lumen, glycosylphosphatidylinositol (GPI)-anchored ectoenzymes	P	327 (222)
ART3	ART3 TMART	ART	PF01129	K-L-V	RTS	SAK	ERI	M	cell membrane, extracellular region, extracellular exosome	P	389 (222)
ART4	ART4 DO DOK1	ART	PF01129	G-S-E	YRT	STS	KKE	M	cell membrane, extracellular region, extracellular exosome, glycosylphosphatidylinositol (GPI)-anchored proteins	P	314 (222)
ART5	ART5 UNQ575/PRO1137	ART	PF01129	R-S-E	RGV	SSS	ERE	M	secreted	P	291 (226)
PARP1	**PARP1 ADPRT PPOL**	PARP	PF00644	H-Y-E	HGS	YFA	YNE	P	nucleus	D, P	1014 (227)
PARP2	**PARP2 ADPRT2 ADPRTL2**	PARP	PF00644	H-Y-E	HGS	YFA	YNE	P	nucleus	D, P, A	583 (228)
PARP3	**PARP3 ADPRT3 ADPRTL3**	PARP	PF00644	H-Y-E	HGT	YFA	QSE	M	nucleus	D, P, A	533 (221)
PARP4	**PARP4 ADPRTL1 KIAA0177 PARPL**	PARP	PF00644	H-Y-E	HGS	YFS	DDE	M	nucleus, exosomes, cell membrane, spindle	P	1724 (205)
PARP5A (tankyrase-1)	TNKS PARP5A PARPL TIN1 TINF1 TNKS1	PARP	PF00644	H-Y-E	HGS	YFA	YAE	O, P	nucleus, telomeres, Golgi apparatus, cytoplasm	P, A	1327 (206)
PARP5B (tankyrase-2)	TNKS2 PARP5B TANK2 TNKL	PARP	PF00644	H-Y-E	HGS	YFA	LAE	O, P	nucleus, telomeres, Golgi apparatus, cytoplasm	P, A	1166 (206)
PARP6	PARP6	PARP	PF00644	H-Y-I	HGS	YLS	GEI	M	cytoplasm	P, A	630 (227)
PARP8	PARP8	PARP	PF00644	H-Y-I	HGS	YLS	GNI	M	nucleus, endoplasmic reticulum	P, A	854 (228)
PARP16	PARP16 ARTD15 C15orf30	PARP	PF00644	H-Y-Y	HGS	YLT	PKY	M	cell membrane, endoplasmic reticulum	P, A	322 (186)

(*continued on next page*)

Table 1 (*continued*)

Name	Gene names*	family classification for ART Domain (Pfam)		Triad motif	R/H-G-T/S motif in β-strand 1	S-X-S/Y-X-X motif in β-strand 2	X-X-E motif at front edge of β-strand 5	ADP-ribosylationactivity: mono (M), oligo (O), poly (P)	Cellular localization	ADP-ribosylation target: protein (P), DNA (D), RNA (R), and auto-ADP-ribosylation (A)	Total length (ART-like domain length)
PARP15	*PARP15 BAL3*	PARP	PF00644	H-Y-L	HGT	YFA	PKL	M	cytoplasm	R, P, A	678 (197)
PARP14	*PARP14 BAL2 KIAA1268*	PARP	PF00644	H-Y-L	HGT	YFA	PSL	M	cytoplasm, nucleus	P, A	1801 (197)
PARP10	*PARP10*	PARP	PF00644	H-Y-I	HGT	YFA	PSI	M	nucleus, cytoplasm	R, P, A	1025 (220)
PARP13	*ZC3HAV1 ZC3HDC2 PRO1677*	PARP	PF00644	Y-Y-V	YAT	YFA	PSV	inactive	cytoplasm	–	902 (187)
PARP7	*TIPARP PARP7*	PARP	PF00644	H-Y-I	HGT	YFA	PQI	M	nucleus, cytoplasm	P, A	657 (209)
PARP12	*PARP12 ZC3HDC1*	PARP	PF00644	H-Y-I	HGT	YFA	PSI	M	cytoplasm	P, A	701 (215)
PARP11	*PARP11 C12orf6*	PARP	PF00644	H-Y-I	HGT	YFA	PKI	M	nucleus (in mice)	R, P, A	338 (216)
PARP9	*PARP9 BAL, BAL1*	Undetected **/insignificant		Q-Y-T	QQV	YFT	PET	M	nucleus, cell membrane, cytoplasm, mitochondrion	ubiquitin	854 (223)
TASOR	TASOR	DUF3715 (partial alignment)	PF12509	L-Y-Q	LMV	YLS	LTQ	inactive	nucleus, chromosome	–	1670 (221)
TASOR2	TASOR2	DUF3715 (partial alignment)	PF12509	*V-T-E*	*VVA*	*TLD*	*LLE*	not determined	cytoplasm, nucleoplasm	not determined	2430 (199)
TEX15	TEX15	DUF3715 (partial alignment)	PF12509	*L-Y-S*	*LAL*	*YMF*	*VLS*	not determined	nucleus, cytoplasm	not determined	2789 (197)
NEURL4	NEURL4	Undetected**		*H-L-E*	*HGS*	*LLS*	*ELE*	not determined	cytoplasm, cytoskeleton	not determined	1562 (121)
LRRC9	LRRC9	Undetected**		*Y-V-K*	*YVF*	*EFL (SIS)*	*QCK*	not determined	not determined	not determined	1453 (233)

Notes.

*PARP family members are sorted by subgroup (font format) on the basis of similarities in amino acid sequence, intron positions and associated protein domains (*Otto et al., 2005*).

**For these domains, Pfam family assignments cannot be made using standard Pfam HMM tool.

Italicized motifs were identified based on HHpred and FFAS03 alignments to canonical PARPs, in some cases the two methods were not in agreement.DUF3715 domains correspond to a part of TASOR-like ART domains.

domains (*Dudkiewicz & Pawłowski, 2019*). Other examples of novel ADP-ribosylation players are provided by the viral macro domains, present in many dangerous viruses, including the SARS and SARS-CoV-2 coronaviruses (*Saikatendu et al., 2005*), and by the recently characterized novel ART-like domain (DUF3715) in the human TASOR protein involved in gene silencing (*Douse et al., 2020*; *Tchasovnikarova et al., 2015*).

In this paper, building up on experience in identification of novel enzyme families (*Dudkiewicz, Lenart & Pawłowski, 2013*; *Dudkiewicz & Pawłowski, 2019*; *Dudkiewicz et al., 2012*; *Pawłowski et al., 2006*; *Sreelatha et al., 2018*), we identified structural similarity between ART-like catalytic domains and a family of eukaryotic uncharacterized protein domains present in homologs of the leucine-rich repeat containing protein 9 (LRRC9). For clarity, we will call this domain LRRC9-ART. Human LRRC9 is annotated in the UniProt database as protein of unknown function with experimental evidence at transcript level. LRRC9 locus is located on chromosome 14 (q23.1). Gene and protein names refer to the twenty-two leucine-rich repeats located between positions 97 and 246 and 717-1365. According to The Human Protein Atlas (*Uhlen et al., 2015*), expression of LRRC9 mRNA is enhanced in brain, pituitary gland and testis.

We investigated sequence similarities between the novel and the known ADP-ribosyltransferases. We reconstructed phylogenetic relationships between sequences within LRRC9 domain family and other ART clan families. We also explored sequence variability in the novel ART-like domain in the context of predicted three-dimensional structures and proposed their likely biological functions.

MATERIALS AND METHODS

The FFAS03 (*Xu et al., 2014*), HHpred (*Zimmermann et al., 2017*) and Phyre2 (*Kelley et al., 2015*) servers were used to determine distant sequence similarities of LRRC9 central domain to proteins of known structures from public databases. Standard parameters and significance thresholds were selected.

The representative set of LRRC9 ART-like domain sequences was collected by submitting the ART-like domain of the human LRRC9 protein (UniProtKB: Q6ZRR7.2, positions 389-628) to two iterations of JackHMMER (also with standard parameters) ran on the Reference Proteomes database.

The MAFFT program (*Katoh, Rozewicki & Yamada, 2019*) was used to build the multiple sequence alignments of the novel domains and PARP catalytic domains obtained from the rp75 set from the Pfam database (*El-Gebali et al., 2019*). In-house scripts were used to merge family-wise multiple sequence alignments according to a FFAS pairwise alignment of representatives of the two families. The WebLogo server (*Crooks et al., 2004*) was used to visualize results as sequence logos.

The CLANS (*Frickey & Lupas, 2004*) algorithm was used to visualize close and distant similarities between the putative and the known ADP-ribosyltransferase domains. Sequence similarities up to BLAST *E*-value of 1, $1e^{-2}$ or $1e^{-4}$ and the BLOSUM45 substitution matrix were used. In order to acquire collection of the known ADP-ribosyltransferase domains, the following sets of sequences were obtained from Pfam database: ADPrib_exo_Tox

rp15, ADPRTs_Tse2 rp15, Anthrax-tox_M rp75, Arr-ms rp15, ART rp15, ART-PolyVal rp15, AvrPphF-ORF-2 rp15, Diphtheria_C rp75, Dot_icm_IcmQ rp35, DUF2441 rp15, DUF952 rp15, Enterotoxin_a rp15, Exotox-A_cataly rp75, NADase_NGA rp35, PARP rp15, Pertussis_S1 rp15, PTS_2-RNA rp15, RolB_RolC rp55 and TNT rp15. Next, collected data were supplemented with three sequence sets representing the following ART-like domain families not included in Pfam database: ESPJ, NEURL4 and DUF3715/TASOR. Sets of ESPJ and SidE domain sequences were obtained by submitting sequences from UniProtKB and Protein NCBI databases (UniProtKB: W0AL45, positions 52-214 and refseq ID: YP_094288.1, positions 686-912, respectively) to two iterations of JackHMMER with standard parameters, ran on UniProtKB database. The input sequence of NEURL4 (UniProtKB: F2TYZ7, positions 102-275) was subject to the same procedure, but extended to three iterations and ran on the Reference Proteomes database. Set of DUF3715 sequences was obtained by submitting the TASOR ART domain (UniProtKB: Q9UK61.3, positions 92-337) to two iterations of JackHMMER with standard parameters. Next, the whole collection of putative and known domain sequences were prepared for the CLANS procedure by clustering with CD-HIT and selecting representatives at a 70% sequence identity threshold (*Huang et al., 2010*).

In order to determine phylogenetic spread of the novel ART domains, a set of LRRC9 ART-like domain homologs was obtained using JackHMMER search seeded with input sequence UniProtKB: Q6ZRR7.2, positions 389-628. Then, Phylogeny.fr platform (*Dereeper et al., 2008*) was utilized to construct phylogenetic tree. A multiple sequence alignment was generated using the MUSCLE program (*Edgar, 2004*). Next, advanced mode of the platform was utilized to build phylogenetic tree using the maximum likelihood PhyML program (*Guindon et al., 2009*) with standard settings (WAG amino-acid substitution model, four substitution rate categories, aLRT test for bootstrap support). Phylogenetic relationships between homologs were visualized as a dendrogram using the iTOL server (*Letunic & Bork, 2016*).

Three three-dimensional structure models for the human LRRC9-ART domain were built using homology modelling method Modeller9v21 (*Webb & Sali, 2017*), I-Tasser server (*Yang et al., 2015*) and Robetta server (*Hiranuma et al., 2020*). Simple homology model (Modeller) was based on human PARP10 structure as single template (PDB ID: 3HKV, identity 16%). The I-Tasser multi template model (*Yang et al., 2015*) were based on best PDB hits identified automatically by the server, top three of them were human tankyrase 2 (PDB ID:3MHK, 3KR7 identity 17%), PARP10 (PDB ID 3HKV, identity 16%) and human poly(ADP-ribose) polymerase 14 (PDB ID 3SMJ, identity 16%). The Robetta model was constructed using a deep learning-based method, trRosetta (*Yang et al., 2020*).

Putative NAD binding pockets in all modelled structures and chosen native PARP structures were first found and described using DoGSiteScorer fully automatic algorithm for pocket and druggability prediction available at ProteinsPlus server (*Volkamer et al., 2012*) and independently identified using the COFACTOR algorithm (*Roy & Zhang, 2012*), an option delivered by the I-Tasser (Iterative Threading ASSEmbly Refinement) server for protein structure and function prediction (*Yang et al., 2015*) based on threading the query structural model through the BioLiP protein function database. COFACTOR identifies

potential functional sites by local and global structure matches and suggests annotations. Putative poses of NAD ligand in the LRRC9-ART trRosetta model were predicted by docking using AutodockVina module in UCSF Chimera (*Trott & Olson, 2010*). Structures were visualized and analyzed using UCSF Chimera tools (*Huang et al., 2014*).

The three-domain model for full-length human LRRC9 protein was built using AIDA (Ab Initio Domain Assembly Server) (*Xu et al., 2015*) based on structures modelled using Modeller9v21 separately for three identified domains: N-terminal LRR region with helical fragment (PDB code of modelling template: 3OJA, 15% identity), ART-like domain (template: 3HKV), and C- terminal LRR region domain (template: 4LSX, 21% identity).

Conservation values of sequence positions in multiple sequence alignment created for the human LRRC9-ART domain homologues (for mapping onto the modelled structures), were derived from Jalview alignment editor (*Waterhouse et al., 2009*) and were automatically calculated as the number of conserved physicochemical properties for each column of the alignment (*Livingstone & Barton, 1993*).

The LRR repeats were identified using the LRRFinder tool (*Offord, Coffey & Werling, 2010*). Gene Ontology analysis of human genes encoding LRR-containing proteins (i.e., their cellular/extracellular localization, molecular function and association with biological processes) was performed with the use of Panther Classification System (*Mi et al., 2017*). The domain organization of LRRC9 proteins was visualised using the DOG2.0 tool (*Ren et al., 2009*).

For analysis of protein and mRNA expression, the Proteomics-db, PAXdb, neXtProt and Protein Atlas databases were used (*Samaras et al., 2020*; *Thul & Lindskog, 2018*; *Wang et al., 2015*; *Zahn-Zabal et al., 2020*). The BioGRID Interaction Database was used to obtain information about possible place of LRRC9 in human protein-protein interaction networks (*Oughtred et al., 2019*). For prediction of the subcellular localization of LRRC9, the DeepLoc server was used (*Almagro Armenteros et al., 2017*).

RESULTS

Assignment of the central domain of LRRC9 to the ADP-ribosyl clan

A sequence similarity screen of human proteins for remote homologs of ADP-ribosyltransferases (ARTs) using the FFAS03 method indicated that weak ART similarity might exist in the central region of the LRRC9 protein (Figs. 1, 2). In order to examine in detail this similarity, we used FFAS03, HHpred and Phyre2 servers. All of them are dedicated to protein structure prediction, allowing detection of non-obvious sequence similarity between very distant protein homologs. These three independent bioinformatics tools revealed statistically significant albeit distant sequence similarity of a region of human LRRC9 protein to representatives of the PARP catalytic domain family between positions 392 and 625 (Table 2). Such remote sequence similarity cannot be the basis for drawing a definite conclusion about an enzymatic function of the putative ART domain.

The relationship of the novel LRRC9 ART-like family, found in many eukaryotic lineages (Fig. 3), to the known ART families can be visualized using the sequence-based CLANS clustering approach (Fig. 4). CLANS analysis was performed at three different *E*-value

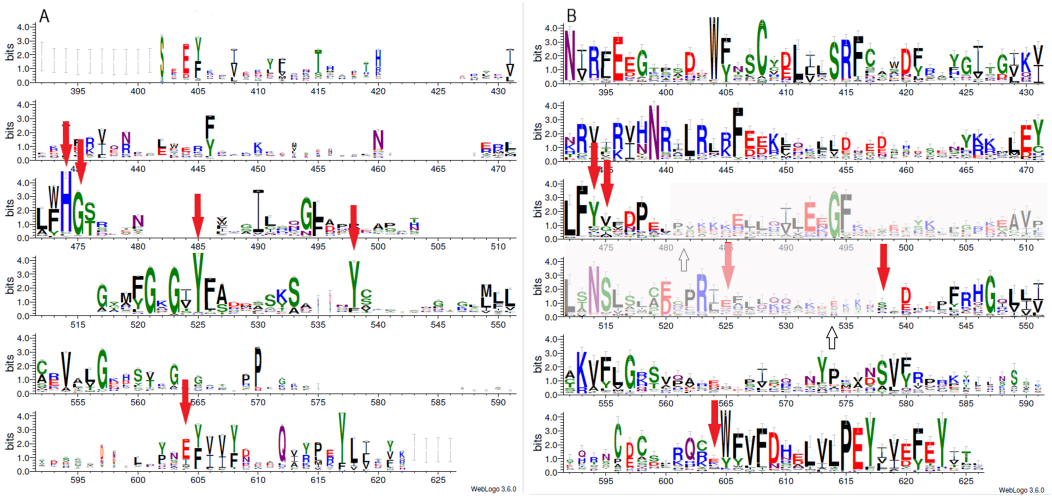

Figure 1 Sequence variability represented as sequence logos for PARP family (A) and LRRC9 ART-like domains (B). Logos were created from two separate MAFFT alignments for PARP family and LRRC9 sequences using WebLogo3 server. The two alignments were merged based on a FFAS03D alignment obtained for human PARP10 and LRRC9-ART domains. Both alignments were trimmed by removing positions with gaps in the human LRRC9 ART domain. Sequence numbering is according to human LRRC9. The catalytically important PARP residues His, Gly, Tyr, Tyr and Glu [matched to positions 474, 475, 525, 538 and 604, respectively, of the human LRRC9 protein] marked with red arrows. Region where FFAS and HHpred alignments between PARP and LRRC9 are inconsistent (between two white arrows) has been faded.

Full-size DOI: 10.7717/peerj.11051/fig-1

LRRC9_HUMAN

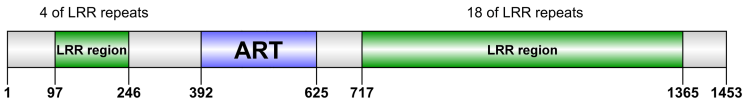

Figure 2 Domain organization of LRRC9 proteins. Sequence numbering as per human LRRC9.

Full-size DOI: 10.7717/peerj.11051/fig-2

Table 2 Structure predictions for the LRRC9 ART-like domain. Structure and distant sequence similarity predictions for human LRRC9 ART-like region (residues 389-628) obtained using different bioinformatics tools. For each method, only the first hit shown.

Bioinformatic tool for structure prediction	Top hit: PDB code, name	Statistical significance for top hit	Region of query aligned to the hit	Sequence identity
FFAS03	**4DVI** Tankyrase 1 with IWR2	Z-score =-12.500	427-625	20%
HHpred	**5LX6,** Human PARP10 (ARTD10), catalytic fragment in complex with PARP inhibitor Veliparib [*Homo sapiens*]	*E*-value =6.2e−23	392-624	15%
Phyre2	**3HKV**, Human poly(ADP-ribose) polymerase 10, catalytic fragment in complex with an inhibitor 3-aminobenzamide; chain A	Confidence= 100%	395-624	16%

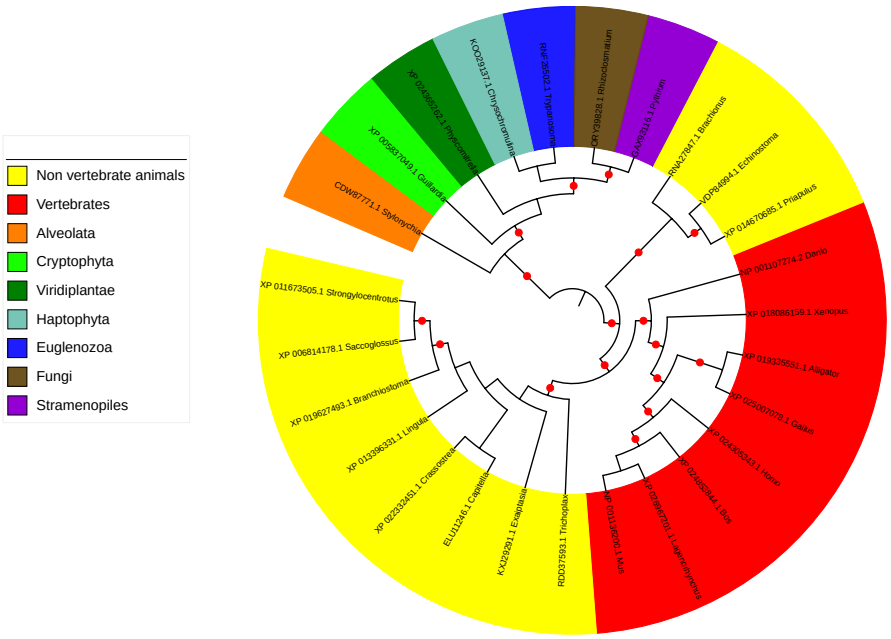

Figure 3 **Phylogenetic tree and taxonomic spread of LRRC9 ART-like domains.** A maximum likelihood phylogenetic tree for representative LRRC9 ART-like domains. NCBI protein identifiers and genus names shown. Colored by taxon. Red circles denote branches with bootstrap support of at least 75%.
Full-size DOI: 10.7717/peerj.11051/fig-3

levels for BLAST hits for all known families containing ADP-ribosyltransferase domains. In agreement with the FFAS03, HHpred and Phyre results, CLANS results suggest a closer relation between the novel ART family and the PARP domains, noticeable at *E*-value threshold 1, 0.01 and even at 0.0001. This, together with sequence conservation analysis allowed us to hypothesize that LRRC9-ART belongs to the H-Y-[EDQ] clade. The CLANS clustering analysis also suggests that LRRC9-ART should be regarded as a separate family within the ART-like clan/superfamily. The same applies to the DUF3715 family that was recently shown to possess an ART-like structure (*Douse et al., 2020*). Representatives of both families clearly cluster away from other ART-like clan members and the two novel families are also away from each other. Notably, detection of very distant sequence similarities depends on accurate domain boundary prediction. This may have hindered the recognition of the ART-like domain in TASOR whereas the domain of unknown function DUF3715 as defined in the Pfam database was missing approx. 50 residues at its N-terminus.

Putative active site of the novel ART-like family

Catalysis of poly-ADP-ribosylation performed by the members H-Y-[EDQ] clade (including PARP domain family) involves three non-consecutive conserved amino acid residues: His, Tyr and an acidic one (Glu or Asp) that can also be substituted by Gln. For consistency, these motifs will be called motifs I, II and III, respectively. Histidine, occurring in the conserved motif I Hx[ST] (*Aravind et al., 2015*) within β-strand 1, is responsible for binding the 2-OH group of the adenosine ribose of NAD^+ and NH_2

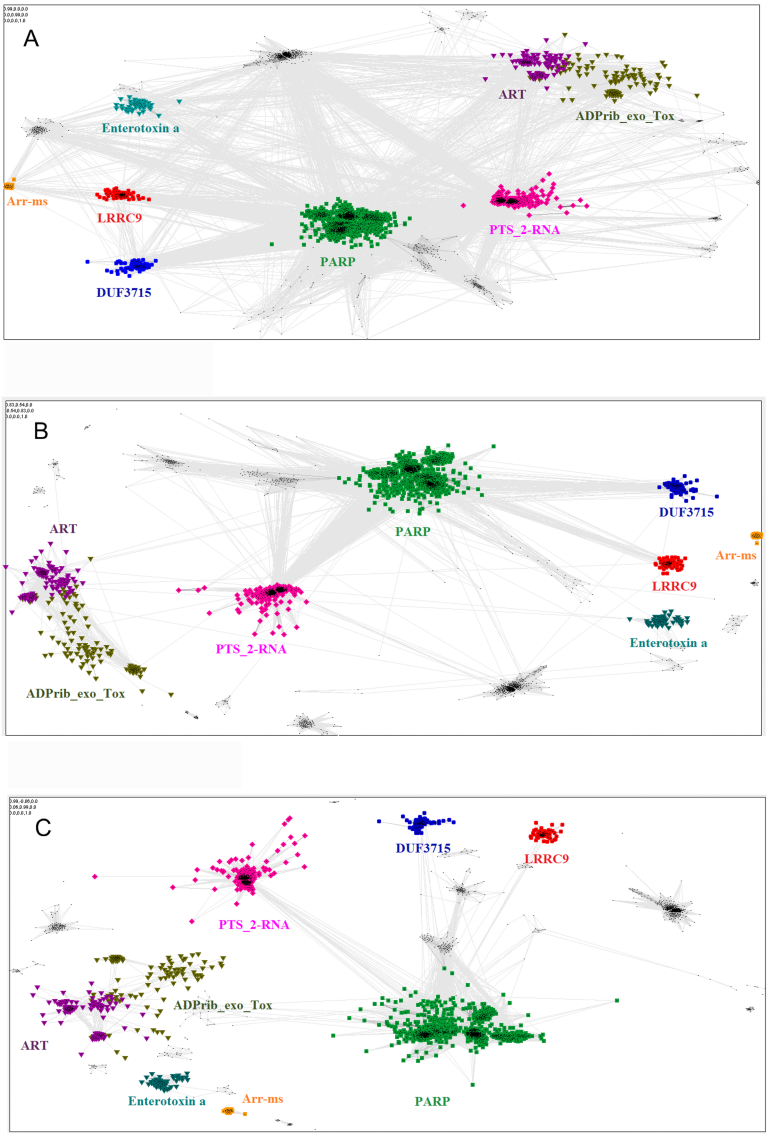

Figure 4 Close and distant sequence similarities between the putative and the known ADP-ribosyltransferase domains, visualized using CLANS algorithm. The graph groups sequences (graph nodes) according to BLAST-derived similarities (edges). (A) Sequence similarities up to BLAST *E*-value of 1, i.e., including distant, non-significant sequence similarities. (B) Up to *E*-value 1e-2. (C) up to *E*-value 1e-4.

Full-size DOI: 10.7717/peerj.11051/fig-4

group of the nicotinamide amide *via* hydrogen bonds (*Cohen & Chang, 2018*). Tyrosine (motif II), localized within β-strand 2 (*Aravind et al., 2015*), interacts with nicotinamide moiety whereas glutamate of β-strand 5 (first position in [QE]x[QED] motif according to Aravind 2018) seems to play a role in maintaining stability of the furanosyl oxocarbenium intermediate (*Cohen & Chang, 2018*).

Multiple sequence alignment comparison allowed us to identify few similarities and substantial differences between sequences of PARP catalytic domains and ART-like

domains, which we identified in human LRRC9 protein sequence in its central part, from position 392 to 625. The tyrosine residue (motif II), one of the catalytic triad elements in PARP, was found to be replaced by a weakly conserved Glu (E525) in human LRRC9-ART. Also, His and Ser residues (from the motif I, His-x-Ser, positions 474–476 in the logo in Fig. 1), highly conserved among PARP enzymatic domain members, were not conserved in LRRC9. The His residue was most often substituted by Tyr. Glutamate, the third element of the catalytic triad (motif III, position 604 in the sequence logo), was replaced by a poorly conserved Lys (Fig. 1). In PARP1, PARP2, PARP10, PARP12 and PARP15 sequences, there is also another important Tyr residue (Tyr932 in human PARP10) that is responsible for nicotinamide stacking (*Karlberg et al., 2015*). Some authors (*Karlberg et al., 2015*) include this second Tyr in the group of conserved residues being the hallmark of PARP ART-domains (*Karlberg et al., 2015*), hence we will use the term "catalytic tetrad HYYE", representing the non-contiguous catalytic motif H-x(n1)-Y-x(n2)-Y-x(n3)-E. This tyrosine is conserved in PARP (at the logo position 538, Fig. 1A) and not conserved in the LRRC9 ART-like domain (Fig. 1B). Lack of most of the catalytic amino acid residues, evolutionarily conserved in PARP catalytic domains, suggests that LRRC9 ART-like domains may be pseudoenzymes. Interestingly, the FFAS03 sequence alignment between human LRRC9 ART-like and PARP10 ART domains, especially in its central part, is not unequivocal. We obtained different alignments using FFAS03 and HHpred servers and even in FFAS results, there are some suboptimal alignment paths to be considered (Figs. S1 and S2). According to the FFAS alignment, PARP region flanked by two tyrosines belonging to the catalytic tetrad HYYE is aligned to a poorly conserved region 525–538 of the LRRC9 ART-like domain (Fig. 1A), but three positions upstream and downstream of this fragment there are two distinct motifs of LRRC9-ART domain: PR[IL] and [DE]xxxFRHG, respectively. This suggests that sequence-based alignments may be inaccurate in this region and one of both above-mentioned LRRC9 conserved motifs may in fact be involved in the active site. Also, the strongly conserved D-(5)-PEY-(2)-EFEY motif in LRRC9 (logo positions 609-623), although not aligned to PARP active site, may be speculated to be functionally important.

Structural analysis of the homology model

Because distant homology detection methods did not offer us consensus for template selection, we decided to create model based on three different approaches: classical one-template homology modeling (Modeller), reassembling structural fragments from threading templates using Monte Carlo simulations (I-Tasser approach, best model with C-score 0.5) and deep learning based methods which use covariance signals obtained from sequence alignments to predict inter-residue distances (trRosetta, best model confidence: 0.70). In the next step, we used DogSiteScorer to analyse the structure models to identify putative ligand binding pockets and compare them to the pockets in known PARP structures (Table S1, Fig. S3) To illustrate uncertainty of structural arrangements of putative ligand-binding residues identified by FFAS03 alignments, we used the homology model (PARP10/3HKV template-based) and the trRosetta (deep learning) model which according to ligand pocket analysis are more likely to accommodate a ligand. Comparison of the

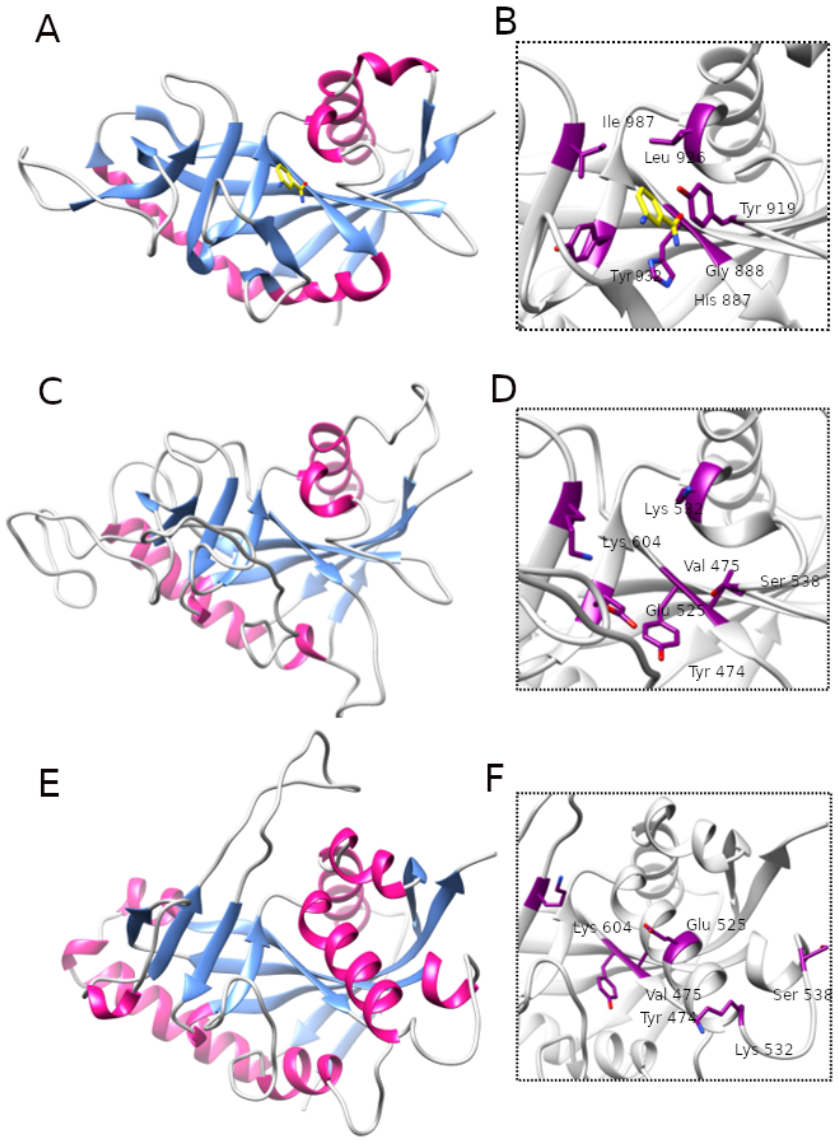

Figure 5 Comparison of ART-like domains and their active site regions. Overall structures of ART domains (A, C, E) and close-up active site regions (B, D, F) of: PARP10, PDB ID: 3HKV (A, B), human LRRC9-ART modelled on the PARP10 template (C, D), and human LRRC9-ART structure modelled using TrRosetta (E, F). Catalytic PARP residues and their counterparts in the modelled structures are depicted as sticks. The ligand shown in panels (A, B) is 3-aminobenzamide, a PARP inhibitor (yellow). Coloring by secondary structures.

Full-size DOI: 10.7717/peerj.11051/fig-5

template 3HKV structure and these two alternative models is presented in Fig. 5. To answer the question if the observed distant sequence similarity between LRRC9-ART and PARP families can translate into functional similarity, we focused first on the known structural determinants of ADP-ribosyltransferase activity (*Karlberg et al., 2015*) and relating these to LRRC9-ART features. In classical PARPs, like PARP1 and PARP2, possessing a poly-ADP-ribosylation (PAR) activity, besides the canonical H-x(n1)-Y-x(n2)-Y-x(n3)-[EI]

tetrad mentioned above, there are two other key positions associated with structural and functional roles of these proteins: glycine residue with nicotinamide anchoring function and an additional catalytic amino acid, lysine 903 (according to PARP1 numbering) (*Karlberg et al., 2015*). In PARP10 and PARP12 with mono-ADP-ribosylation (MAR) activity, instead of this catalytic lysine, there is a leucine or tyrosine residue. Investigating profile-profile alignments for human LRRC9 and PARP10 ART domains one can easily notice that only the first of the canonical tetrad residues is conserved, although substituted (H->Y). Thus, histidine and glycine responsible for nicotinamide anchoring and binding are replaced by tyrosine and valine (Figs. 5B, 5C). In PARPs, the glycine makes two conserved H-bonds to the nicotinamide via main chain carbonyl oxygen and amine hydrogen. These should not be disrupted by valine substitution as the side chain points outward from the active site. Interestingly, in LRRC9-ART at position 532 there is a weakly conserved Lys, as in PARP family members with poly-ADP-ribosylation (PAR) activity. However, instead of a second Tyr (logo position 538) and Glu/Ile residues (604), there are poorly conserved Ser and Lys.

To identify possible ligand pockets in modelled structures we used DoGSiteScorer. For comparison, we also analysed classic PARP structures where ligand binding sites are known and described. We selected PARP1 (ART domain with poly-ADP-ribosylation activity), PARP10 (mono-ADP-ribosylation activity) and two examples of inactive PARP/ART-like domains: PARP13 and TASOR. In case of native structures, DoGSiteScorer correctly identified pockets in the vicinity of PARP hallmark H-Y-Y-E tetrad or its counterpart as the best scoring ones. In case of models, pockets detected in the neighborhood of H-Y-Y-E aligned residues were the best scoring ones for I-Tasser and trRosetta models (Table S1). Overall shape, volume and surface of modelled pockets were very different, but the best score (close to the score obtained for native NAD binding pocket in PARP1) and parameters mostly resembling active PARPs were identified in trRosetta model of LRRC9-ART (Fig. S3). Thus, this model appears to best capture the putative ligand binding site in LRRC9.

Additionally, to investigate possible ligands for the LRRC9-ART domain we used COFACTOR. COFACTOR is a functional annotation tool based on threading of the protein structure query through a database of ligand-protein binding interactions by local and global structure matches to identify functional sites (*Zhang, Freddolino & Zhang, 2017*). COFACTOR suggests the best candidates fitting into the ligand binding site of a given structure. Adenosine monophosphate (AMP) turned out to be among the best scoring ligands for the LRRC9-ART homology model while NAD, a typical ligand for ART domains, was absent from the predicted ligand list. AMP molecule, which differs from NAD by the lack of nicotinamide mononucleotide moiety, is smaller and probably fits better in the modelled LRRC9-ART ligand binding groove. The docked AMP molecule pose was deduced from homology to crystal structure of the catalytic domain of *Pseudomonas aeruginosa* exotoxin A (PDB ID: 1DMA) which catalyzes the transfer of ADP ribose from NAD to elongation factor-2 in eukaryotic cells and hence inhibits protein synthesis. These docking results do not suggest that AMP is a physiological ligand of LRRC9-ART, rather, they provide a likely binding pose of the AMP moiety of NAD. For trRosetta model, COFACTOR found tankyrase-1 complexed with a PARP inhibitor, P4L, as the best ligand

binding site analogue. Here, NAD molecule was not present on the list of ligands complexed with identified binding site structural homologues.

To examine the possibility of NAD binding, we attempted to dock the full NAD molecule into the LRRC9-ART hypothetical active site in the trRosetta (deep learning) structure model. The ligand was constructed in UCSF Chimera build tool and then it was docked using Autodock Vina procedure without specific constraints to the binding groove identified by DoGSiteScorer (Fig. S4B). The docking procedure located NAD in the best scoring pocket in few different poses (Fig. S4A), the best one with a good docking score of −6.8 (Fig. S4C). This suggests we cannot exclude the LRRC9-ART domain, at least as modelled using trRosetta approach, can bind NAD. However, docking a ligand to a structure model built using very distant templates has to be seen as speculative.

Replacing the catalytic Glu residue by Ile is characteristic for PARP family members having mono-ADP-ribosylation (MAR) activity (*Karlberg et al., 2015*). Among FFAS/HHpred hits for LRRC9-ART domain, proteins with MAR, as well as PAR activity and with no confirmed ADP-ribosylation activity appeared. In the sequence logos of ART and LRRC9-ART, there are no common conserved motifs between the two domain families. In LRRC9, the second tyrosine of the HYYE tetrad is replaced by glutamine 525, which is noticeably less conserved. Another highly conserved motif of the PARP family, YxEY[VI][VI][FY], containing catalytic Glu of the HYYE tetrad does not align to a relatively similar motif visible in the LRRC9-ART domain EY[VI][VI]E[FY]EY (logo positions 616-623). As mentioned earlier, these motifs may correspond to each other, and the apparent misalignment may be just an artifact.

Figure S4 shows the trRosetta model of LRRC9-ART structure, with NAD molecule docked in the predicted active site. Figure S4B focuses on the same residues as Fig. 5 but colored according to conservation in the LRRC9-ART family. Remarkably, only tyrosine 474 and valine 475 (out of the residues aligned to the catalytic tetrad) are quite strongly conserved in the whole family. The rest of LRRC9 amino acids aligned to the classical HYYE tetrad residues in the ligand-binding hydrophobic pocket are not highly conserved. This approach of docking a ligand to a structure model built using remote sequence similarity cannot be regarded as authoritative; it is only preliminary and can give a general idea of ligand placement, but no detailed information on its interactions. The model presented in surface representation (Fig. S4C) shows the conserved residues are grouped at the bottom of the putative ligand binding groove.

The ART-like domain described here is only one region of approx. 200 residues within the whole LRRC9 protein molecule, which is 1453 amino acids long, with clear similarity to leucine-rich repeat-containing domains in its N-and C-terminal parts, and a very long helical "spine" in the center. We were tempted to construct the full-length molecule model, based on homology to LRR-containing templates and using the already modelled ART-like domain structure. The results are shown in Fig. S5. In the full-length structure model proposed here, a large groove is surrounded by LRR "horseshoes" from three sides and an ART-like domain from the fourth. The size of the groove seems to suggest a possibility of binding a large molecule, e.g., a protein rather than a small cofactor. It is tempting to speculate that LRRC9, binding some "bait", rearranges its conformation and activates the

enzymatic ART-like domain, thereby triggering a cellular signal. However, the full-length model is only an illustration of possible domain arrangements.

Taxonomic distribution of the LRRC9-like proteins

The LRRC9 protein is widespread in several of the major eukaryotic lineages. It is found in *Alveolata* and *Stramenopiles* (from the TSAR supergroup *Burki et al., 2020*), *Cryptophyta*, *Haptophyta, Viridiplantae, Fungi and Metazoa (Opisthokonta)* and *Euglenozoa* while absent from several other main lineages, e.g., *Hemimastigophora*, CRuMS, *Metamonada*, *Malawimonadida*, *Ancyromonadida*, *Ancoracysta*, *Picozoa* (Fig. 3). Thus, this protein appears to be an ancestral eukaryotic feature, that was subject to losses in a number of lineages.

ART-like domain is present in all vertebrate classes as well as in 11 non-vertebrate animal phyla (*e.g., Cnidaria, Echinodermata, Brachiopoda, Mollusca, Annelida, Hemichordata, Chordata*). However, notable is its absence in several animal phyla, some of which include important model organisms (e.g., *Arthropoda*, *Nematoda*, *Porifera*; see Fig. 3). Also, strikingly, although LRRC9-ART domains are found in some *Viridiplantae*, e.g., mosses, green algae, liverworts, and club mosses, they are absent from flowering plants with the exception of the water lily *Nymphaea colorata*. Notably, water lilies belong to *Magnoliophyta*, a taxon thought to have diverged earliest from the lineage leading to most extant flowering plants (*Zhang et al., 2020*). Altogether, this taxonomic spread suggests an ancient function for LRRC9, common to many diverse eukaryotes.

DISCUSSION

Because the regions of LRRC9-ART domain corresponding to ADP-ribosyltransferase active sites are poorly conserved, the most straightforward conclusion appears to be that LRRC9-ART is a pseudoenzymatic domain. However, as exemplified by the recent cases of two pseudokinases identified by us, SelO and SidJ, apparent pseudoenzymes may in fact perform enzymatic functions somewhat similar to those performed by their "non-pseudo" cousins (*Black et al., 2019*; *Sreelatha et al., 2018*). Also, many inactive pseudoenzymes engage their non-pseudo cousins and modulate their activity, e.g., allosterically (*Adrain, 2020*). Thus, we discuss the significance of the LRR repeats flanking the ART-like domain while speculating that the central domain either performs by itself an ADP-ribosylation-related catalytic function, or modulates such function performed by yet another molecule.

Apart from the ART-like domain, LRRC9 proteins comprise a variable number of leucine-rich repeats (LRRs) flanking the ART domain (Fig. 2). In all the homologs analysed, the LRR-ART-LRR domain architecture is conserved in evolution. The N-terminal region possesses 1-6 LRR repeats while the C-terminal region possesses 11-20 such repeats. The N-terminal repeats and C-terminal repeats are most similar to the LRR repeats from proteins such as Leucine Rich Repeat Containing 23 (LRC23) and TLR4 Interactor with leucine-rich Repeats (TRIL), respectively, however, degree of similarity between LRR motifs from different proteins is generally low and makes it difficult to draw specific functional conclusions. The LRR motifs typically have a length of about 20-30 amino acid residues and are rich in leucines or other aliphatic residues (typically seven per motif,

localized at positions 2, 5, 7, 12, 16, 21 and 24 within the consensus LRR sequence) (*Kobe & Deisenhofer, 1994*). The LRR motifs may form various combinations of two or three secondary structures, usually involving two stretches of a β-strand, e.g., two β-strands and an α-helix. The LRR repeats typically fold into a horseshoe-like conformation composed of helices present on its convex face and parallel β-sheet localized on the concave one (*Enkhbayar et al., 2004*). LRR motifs determine functionality of LRR-containing proteins, enabling them to make protein-protein interactions (*Kobe & Deisenhofer, 1994*) as well as to bind various target structures such as small molecule hormones, lipids and nucleic acids (*Helft et al., 2011*). The non-globular shape of LRR motifs may facilitate the contact between LRR-containing proteins and the target structures, increasing the interaction area (*Kobe & Deisenhofer, 1994*). For example, ribonuclease inhibitor is able to bind pancreatic ribonucleases and inhibit their enzymatic activity. This interaction may affect RNA turnover in the context of angiogenesis (*Kobe & Deisenhofer, 1993*).

Employing the LRR motifs, LRR proteins perform various functions. They may play a role in signal transduction and cell development as well as they are able to participate in RNA splicing and effective response to DNA damage (*Kobe & Deisenhofer, 1995*). Many LRR-containing proteins are adhesive molecules that perform important functions in processes such as regulation of collagen-fibril formation, osteogenesis, myelination and platelet adhesion at a site of vascular injury. Many proteins with LRR motifs play a role in signal transduction as specific receptors. For example, in response to LPS stimulation, LRR-containing CD14 protein induces tyrosine phosphorylation of intracellular proteins to stimulate antibacterial activity of macrophages. Other receptors exhibit specificity to gonadotrophins such as luteinizing hormone, chorionic gonadotrophin and follicle-stimulating hormone (*Bluhm et al., 2004*).

The LRR domains, structurally conserved in evolution, occur widely within plant, invertebrate and vertebrate proteins responsible for innate immunity. Due to their ability to mediate protein-protein interactions, LRR repeats determine signaling function of many pattern recognition receptors (PRRs), such as Toll-like receptors (TLRs) and NOD-like receptors (NLRs), by conditioning their affinity to various ligands, e.g., viral, bacterial, fungal and parasite antigens (*Ng & Xavier, 2011*; *Ng et al., 2011*). On the other hand, ADP-ribosylation is known to be involved in regulation of the host immune response (*Fehr et al., 2020*; *Rosado et al., 2013*). It may direct intracellular signaling to parthanatos (type of programmed death) (*Fatokun, Dawson & Dawson, 2014*; *Greenwald & Pierce, 2019*; *Hoch & Polo, 2019*), whereas such scenario implies induction of inflammatory response to infection. Poly-ADP-ribosylation is known to influence the stability of transcripts encoding proinflammatory cytokines (*Ke et al., 2019*). PARP-dependent chromatin modification may counteract progression of viral infection. ADP-ribosylation also influences other innate immune mechanisms such as NF-kB expression, phagocytosis and macrophage polarization (*Kunze & Hottiger, 2019*). Here, co-occurrence of ART and LRR domains within the LRRC9 protein suggests that they may act together to support a version of host innate immunity. In this context, the LRR domains could perform signal sensors function, and the ART domain might play the effector role.

Overall, in the human proteome there are 234 proteins containing LRR repeats. The molecular functions significantly overrepresented among these proteins include peptide binding (11 proteins, FDR *p*-value 8E-6), ubiquitin-protein transferase activity (11 proteins), heparin binding (7 proteins) and G-protein-coupled peptide receptor activity (7 proteins). However, as many as 182 human LRR proteins are functionally uncharacterized, according to Panther-db. Human proteins containing LRR repeats are involved in processes such as neurogenesis (17 proteins, FDR 1E-5), ubiquitin-dependent protein catabolic process (11 proteins), positive regulation of innate immune response *via* toll-like receptor signaling pathway (10 proteins, FDR 8E-10).

The ART-like domain combined with LRR motifs has precedents. As many as 35 human LRR proteins possess enzymatic domains. Notable here are the NLRP immune sensors/effectors containing NACHT ATP-ase domains, totaling 17 human proteins (*Martinon & Tschopp, 2007*; *Pawlowski et al., 2001*). Other enzymes with LRR motifs include protein kinases (11 proteins), nucleases and peroxidases. Eight human LRR proteins also possess F-box motifs and are involved in ubiquitin ligase complexes.

A bioinformatics study analyzing non-LRR regions embedded within LRR proteins identified the LRRC9 family as containing such a "non-LRR island", noted a few conserved sequence motifs and homologs in a number of non-vertebrate eukaryotes but did not observe similarity to any characterized functional domains (*Matsushima et al., 2009*).

As no high-quality protein expression data is available for LRRC9, it is annotated as "existence validated on transcript level" in the neXtProt (*Zahn-Zabal et al., 2020*) database. According to PAXdb (*Wang et al., 2015*), LRRC9 protein expression is elevated in tonsil, esophagus and seminal vesicle tissues. This may suggest that LRRC9 protein expression is limited to specific circumstances, and is not easily captured in typical tissue proteomics experiments. However, recently, a study aimed at validating the "missing proteins" confirmed the existence of LRRC9 protein in human sperm by applying targeted mass spectrometry and antibody-based methods (*Carapito et al., 2017*). LRRC9 mRNA expression is highest within the brain as well as endocrine and male reproductive tissues, i.e., pituitary gland and testis, respectively (Protein Atlas database). The apparent discrepancy between mRNA and proteomics expression data for LRRC9 may be related to technical difficulties (e.g., too few unique peptides of sufficient length). According to the DeepLoc prediction, the subcellular localization of LRRC9 is cytoplasmic (likelihood 0.79).

For an uncharacterized protein, its physical interactors may shed light on its function. However, according to the BioGRID database, affinity capture-mass spectrometry analysis provided evidence that LRRC9 physically interacts with a single protein, ZFP36L2 (*Noguchi et al., 2018*). ZFP36L2 is an RNA-binding protein regulating cell cycle (*Galloway et al., 2016*) that contributes to pathogenesis of pancreatic ductal adenocarcinoma (PDAC). ZFP36L2 is responsible for an increase in cancer cell aggressiveness (*Yonemori et al., 2017*). A Kaplan–Meier survival analysis (Protein Atlas) shows that LRRC9 mRNA expression is positively correlated with survival among patients with pancreatic cancer ($p = 0.0034$), although the gene is not classified as prognostic. This may suggest an antagonistic relation between LRRC9 and ZFP36L2 activities in the context of ZFP36L2-dependent promotion

of PDAC progression. However, the LRRC9-ZFP36L2 interaction was only reported in a high-throughput experiment, and further investigation is required here.

LRRC9-ART should be classified as a separate ART-like domain family, not only because of the co-occurrence with LRR domains, but primarily because the LRRC9-ART similarity to known ARTs is very remote, and the conservation patterns in LRRC9-ART family are distant from those typical for either PARP or TASOR families.

Pseudoenzymes are characterized by the lack of enzymatic activity despite the common evolutionary origin and structural similarity to catalytically active homologues (*Jeffery, 2020*; *Murphy, Mace & Eyers, 2017*; *Ribeiro et al., 2019*). Pseudoenzymes do occur among PARP family members, including PARP13 that is distinguished by amino acid substitutions and structure changes preventing catalysis. Similarly to LRRC9-ART, PARP13 catalytic domain is characterized by lack of conserved His and Glu residues (*Karlberg et al., 2015*). It has been shown that substitution of catalytic His with Ala within the ART domain of PARP1 substantially decreases its catalytic activity (*Marsischky, Wilson & Collier, 1995*). The results were consistent with previous studies on diphtheria toxin, another member of H-Y-[EDQ] clade. Replacement of catalytic His reduced the toxin's activity by at least 70-fold (*Blanke et al., 1994*). Thus, the LRRC9-ART family, lacking most of the ART active site, can be hypothesized to be pseudoenzymes. However, one cannot exclude the possibility that this ART-like domain has retained or regained the ART catalytic function, or acquired a novel enzymatic activity, in both cases employing an atypical and/or migrated active site. Such a scenario has been recently reported for apparent pseudokinases, SelO and SidJ, that were shown to be AMPylases and polyglutamylases, respectively (*Black et al., 2019*; *Sreelatha et al., 2018*). It can also be envisaged that LRRC9 may require another protein factor to complement its active site, similarly to ADP-ribosyltransferases PARP1 and PARP2 that form complexes with HPF1 which contributes a "missing residue" to the otherwise "incomplete" ART active site (*Suskiewicz et al., 2020*). Another precedent for a "third party" protein required for ART activity is provided by the PARP9/DTX3L heterodimer whereas the otherwise inactive PARP9 needs the interaction partner to become active (*Yang et al., 2017*).

On the other side, in eukaryotic cells, poly-ADP-ribosylating proteins may perform some of their functions without utilizing the enzymatic activity. For example, PARP1 participates in NF-kB-dependent gene transcription *via* two different mechanisms but only one requires poly-ADP-ribosylation whereas the second seems not to involve PARP1 enzymatic function and depends on the structure of the protein (*Hassa et al., 2003*; *Weaver & Yang, 2013*). Both a pseudo-enzymatic character of LRRC9-ART, and migration of the active site are possible for this intriguing novel family. Experimental studies are needed to confirm one of those hypotheses. The unique, conserved domain architecture of LRRC9, suggests that this mysterious protein family could be involved in a defense mechanism, with some analogies to the innate immune system, coupling within a single molecule the detection of foreign objects (LRRs) and downstream signalling (ART domain).

ACKNOWLEDGEMENTS

We thank Drs Vincent Tagliabracci, Miles Black, Anna Muszewska and Marcin Grynberg for critical reading of the manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Krzysztof Pawłowski was supported by the Polish National Agency for Scientific Exchange scholarship PPN/BEK/2018/1/00431 and by the Polish National Science Centre grants 2018/31/B/NZ2/00758 and 2019/33/B/NZ2/01409. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
The Polish National Agency for Scientific Exchange scholarship PPN/BEK/2018/1/00431.
The Polish National Science Centre grants: 2018/31/B/NZ2/00758, 2019/33/B/NZ2/01409.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Zbigniew Wyżewski performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Marcin Gradowski and Marianna Krysińska performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Małgorzata Dudkiewicz and Krzysztof Pawłowski conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Raw data (including protein sequence alignments, CLANS analysis files, subsignificant alignment plots and PDB file for structure model) are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.11051#supplemental-information.

REFERENCES

Adrain C. 2020. Pseudoenzymes: dead enzymes with a lively role in biology. *FEBS Journal* **287**:4102–4105 DOI 10.1111/febs.15535.

Aktories K, Lang AE, Schwan C, Mannherz HG. 2011. Actin as target for modification by bacterial protein toxins. *FEBS Journal* **278**:4526–4543 DOI 10.1111/j.1742-4658.2011.08113.x.

Akturk A, Wasilko DJ, Wu X, Liu Y, Zhang Y, Qiu J, Luo Z-Q, Reiter KH, Brzovic PS, Klevit RE, Mao Y. 2018. Mechanism of phosphoribosyl-ubiquitination mediated by a single Legionella effector. *Nature* **557**:729–733 DOI 10.1038/s41586-018-0147-6.

Aravind L, Zhang D, De Souza RF, Anand S, Iyer LM. 2015. The natural history of ADP-ribosyltransferases and the ADP-ribosylation system. *Current Topics in Microbiology and Immunology* **384**:3–32 DOI 10.1007/82_2014_414.

Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. 2017. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**:3387–3395 DOI 10.1093/bioinformatics/btx431.

Bai P. 2015. Biology of Poly(ADP-Ribose) polymerases: the factotums of cell maintenance. *Molecular Cell* **58**:947–958 DOI 10.1016/j.molcel.2015.01.034.

Barreiro E, Gea J. 2018. PARP-1 and PARP-2 activity in cancer-induced cachexia: potential therapeutic implications. *Biological Chemistry* **399**:179–186 DOI 10.1515/hsz-2017-0158.

Baysarowich J, Koteva K, Hughes DW, Ejim L, Griffiths E, Zhang K, Junop M, Wright GD. 2008. Rifamycin antibiotic resistance by ADP-ribosylation: structure and diversity of Arr. *Proceedings of the National Academy of Sciences of the United States of America* **105**:4886–4891 DOI 10.1073/pnas.0711939105.

Bhogaraju S, Kalayil S, Liu Y, Bonn F, Colby T, Matic I, Dikic I. 2016. Phosphoribosylation of ubiquitin promotes serine ubiquitination and impairs conventional ubiquitination. *Cell* **167**:1636–1649 DOI 10.1016/j.cell.2016.11.019.

Black M, Osinski A, Gradowski M, Servage K, Pawłowski K, Tomchick D, Tagliabracci V. 2019. Bacterial pseudokinase catalyzes protein polyglutamylation to inhibit the SidE all-in-one ubiquitin ligases. *Science* **364**:787–792 DOI 10.1126/science.aaw7446.

Blanke SR, Huang K, Wilson BA, Papini E, Covacci A, Collier RJ. 1994. Active-site mutations of the diphtheria toxin catalytic domain: role of histidine-21 in nicotinamide adenine dinucleotide binding and ADP-ribosylation of elongation factor 2. *Biochemistry* **33**:5155–5161 DOI 10.1021/bi00183a019.

Bluhm AP, Toledo RA, Mesquita FM, Pimenta MT, Fernandes FM, Ribela MT, Lazari MF. 2004. Molecular cloning, sequence analysis and expression of the snake follicle-stimulating hormone receptor. *General and Comparative Endocrinology* **137**:300–311 DOI 10.1016/j.ygcen.2004.03.014.

Boyer L, Doye A, Rolando M, Flatau G, Munro P, Gounon P, Clément R, Pulcini C, Popoff MR, Mettouchi A, Landraud L, Dussurget O, Lemichez E. 2006. Induction of transient macroapertures in endothelial cells through RhoA inhibition by Staphylococcus aureus factors. *Journal of Cell Biology* **173**:809–819 DOI 10.1083/jcb.200509009.

Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The new tree of eukaryotes. *Trends in Ecology & Evolution* **35**:43–55 DOI 10.1016/j.tree.2019.08.008.

Carapito C, Duek P, Macron C, Seffals M, Rondel K, Delalande F, Lindskog C, Freour T, Vandenbrouck Y, Lane L, Pineau C. 2017. Validating missing proteins in human sperm cells by targeted mass-spectrometry- and antibody-based methods. *Journal of Proteome Research* **16**:4340–4351 DOI 10.1021/acs.jproteome.7b00374.

Chiu CC, Chen HH, Chuang HL, Chung TC, Chen SD, Huang YT. 2009. Pseudomonas aeruginosa exotoxin A-induced hepatotoxicity: an animal model in rats. *Journal of Veterinary Medical Science* **71**:1–8 DOI 10.1292/jvms.71.1.

Choi JR, Shin KS, Choi CY, Kang SJ. 2016. PARP1 regulates the protein stability and proapoptotic function of HIPK2. *Cell Death & Disease* **7**:e2438 DOI 10.1038/cddis.2016.345.

Cohen MS, Chang xx. 2018. Insights into the biogenesis, function, and regulation of ADP-ribosylation. *Nature Chemical Biology* **14**:236–243 DOI 10.1038/nchembio.2568.

Corda D, Di Girolamo M. 2002. Mono-ADP-ribosylation: a tool for modulating immune response and cell signaling. *Science STKE [Signal Transduction Knowledge Environment]* **2002**:pe53 DOI 10.1126/stke.2002.163.pe53.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Research* **14**:1188–1190 DOI 10.1101/gr.849004.

Curtin NJ. 2005. PARP inhibitors for cancer therapy. *Expert Reviews in Molecular Medicine* **7**:1–20 DOI 10.1017/S146239940500904X.

Curtin NJ, Szabo C. 2013. Therapeutic applications of PARP inhibitors: anticancer therapy and beyond. *Molecular Aspects of Medicine* **34**:1217–1256 DOI 10.1016/j.mam.2013.01.006.

De Souza RF, Aravind L. 2012. Identification of novel components of NAD-utilizing metabolic pathways and prediction of their biochemical functions. *Molecular BioSystems* **8**:1661–1677 DOI 10.1039/c2mb05487.

De Vos M, Schreiber V, Dantzer F. 2012. The diverse roles and clinical relevance of PARPs in DNA damage repair: current state of the art. *Biochemical Pharmacology* **84**:137–146 DOI 10.1016/j.bcp.2012.03.018.

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* **36**:W465–W469 DOI 10.1093/nar/gkn180.

Douse CH, Tchasovnikarova IA, Timms RT, Protasio AV, Seczynska M, Prigozhin DM, Albecka A, Wagstaff J, Williamson JC, Freund SMV, Lehner PJ, Modis Y. 2020. TASOR is a pseudo-PARP that directs HUSH complex assembly and epigenetic transposon control. *bioRxiv*. 2020.2003.2009.974832 DOI 10.1101/2020.03.09.974832.

Dudkiewicz M, Lenart A, Pawłowski K. 2013. A novel predicted calcium-regulated kinase family implicated in neurological disorders. *PLOS ONE* **8**:e66427 DOI 10.1371/journal.pone.0066427.

Dudkiewicz M, Pawłowski K. 2019. A novel conserved family of Macro-like domains—-putative new players in ADP-ribosylation signalling. *PeerJ* **7**:e6863 DOI 10.7717/peerj.6863.

Dudkiewicz M, Szczepinska T, Grynberg M, Pawłowski K. 2012. A novel protein kinase-like domain in a selenoprotein, widespread in the tree of life. *PLOS ONE* **7**:e32138 DOI 10.1371/journal.pone.0032138.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**:113 DOI 10.1186/1471-2105-5-113.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Research* **47**:D427–D432 DOI 10.1093/nar/gky995.

Enkhbayar P, Kamiya M, Osaki M, Matsumoto T, Matsushima N. 2004. Structural principles of leucine-rich repeat (LRR) proteins. *Proteins* **54**:394–403 DOI 10.1002/prot.10605.

Fatokun AA, Dawson VL, Dawson TM. 2014. Parthanatos: mitochondrial-linked mechanisms and therapeutic opportunities. *British Journal of Pharmacology* **171**:2000–2016 DOI 10.1111/bph.12416.

Fehr AR, Singh SA, Kerr CM, Mukai S, Higashi H, Aikawa M. 2020. The impact of PARPs and ADP-ribosylation on inflammation and host-pathogen interactions. *Genes and Development* **34**:341–359 DOI 10.1101/gad.334425.119.

Frickey T, Lupas A. 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**:3702–3704 DOI 10.1093/bioinformatics/bth444.

Galloway A, Saveliev A, Łukasiak S, Hodson DJ, Bolland D, Balmanno K, Ahlfors H, Monzón-Casanova E, Mannurita SC, Bell LS, Andrews S, Díaz-Muñoz MD, Cook SJ, Corcoran A, Turner M. 2016. RNA-binding proteins ZFP36L1 and ZFP36L2 promote cell quiescence. *Science* **352**:453–459 DOI 10.1126/science.aad5978.

Gamble MJ, Fisher RP. 2007. SET and PARP1 remove DEK from chromatin to permit access by the transcription machinery. *Nature Structural & Molecular Biology* **14**:548–555 DOI 10.1038/nsmb1248.

Greenwald SH, Pierce EA. 2019. Parthanatos as a cell death pathway underlying retinal disease. *Advances in Experimental Medicine and Biology* **1185**:323–327 DOI 10.1007/978-3-030-27378-1_53.

Guindon S, Delsuc F, Dufayard JF, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. *Methods in Molecular Biology* **537**:113–137 DOI 10.1007/978-1-59745-251-9_6.

Gunderson CC, Moore KN. 2015. Olaparib: an oral PARP-1 and PARP-2 inhibitor with promising activity in ovarian cancer. *Future Oncology* **11**:747–757 DOI 10.2217/fon.14.313.

Hassa PO, Buerki C, Lombardi C, Imhof R, Hottiger MO. 2003. Transcriptional coactivation of nuclear factor-kappaB-dependent gene expression by p300 is regulated by poly(ADP)-ribose polymerase-1. *Journal of Biological Chemistry* **278**:45145–45153 DOI 10.1074/jbc.M307957200.

Helft L, Reddy V, Chen X, Koller T, Federici L, Fernández-Recio J, Gupta R, Bent A. 2011. LRR conservation mapping to predict functional sites within protein leucine-rich repeat domains. *PLOS ONE* **6**:e21614 DOI 10.1371/journal.pone.0021614.

Hiranuma N, Park H, Baek M, Anishchanka I, Dauparas J, Baker D. 2020. Improved protein structure refinement guided by deep learning based accuracy estimation. *bioRxiv*.

Hoch NC, Polo LM. 2019. ADP-ribosylation: from molecular mechanisms to human disease. *Genetics and Molecular Biology* **43**:e20190075 DOI 10.1590/1678-4685-gmb-2019-0075.

Huang CC, Meng EC, Morris JH, Pettersen EF, Ferrin TE. 2014. Enhancing UCSF Chimera through web services. *Nucleic Acids Research* **42**:W478–W484 DOI 10.1093/nar/gku377.

Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**:680–682 DOI 10.1093/bioinformatics/btq003.

Jankevicius G, Ariza A, Ahel M, Ahel I. 2016. The toxin-antitoxin system DarTG catalyzes reversible ADP-Ribosylation of DNA. *Molecular Cell* **64**:1109–1116 DOI 10.1016/j.molcel.2016.11.014.

Javle M, Curtin NJ. 2011. The role of PARP in DNA repair and its therapeutic exploitation. *British Journal of Cancer* **105**:1114–1122 DOI 10.1038/bjc.2011.382.

Jeffery CJ. 2020. Enzymes, pseudoenzymes, and moonlighting proteins: diversity of function in protein superfamilies. *FEBS Journal* DOI 10.1111/febs.15446.

Jwa M, Chang . 2012. PARP16 is a tail-anchored endoplasmic reticulum protein required for the PERK- and IRE1 α-mediated unfolded protein response. *Nature Cell Biology* **14**:1223–1230 DOI 10.1038/ncb2593.

Kagan JC, Roy CR. 2002. Legionella phagosomes intercept vesicular traffic from endoplasmic reticulum exit sites. *Nat Cell Biol* **4**:945–954 DOI 10.1038/ncb883.

Kalayil S, Bhogaraju S, Bonn F, Shin D, Liu Y, Gan N, Basquin J, Grumati P, Luo ZQ, Dikic I. 2018. Insights into catalysis and function of phosphoribosyl-linked serine ubiquitination. *Nature* **557**:734–738 DOI 10.1038/s41586-018-0145-8.

Kamboj A, Lu P, Cossoy MB, Stobart JL, Dolhun BA, Kauppinen TM, Murcia Gde, Anderson CM. 2013. Poly(ADP-ribose) polymerase 2 contributes to neuroinflammation and neurological dysfunction in mouse experimental autoimmune encephalomyelitis. *Journal of Neuroinflammation* **10**:821 DOI 10.1186/1742-2094-10-49.

Kamel D, Gray C, Walia JS, Kumar V. 2018. PARP inhibitor drugs in the treatment of breast, ovarian, prostate and pancreatic cancers: an update of clinical trials. *Current Drug Targets* **19**:21–37 DOI 10.2174/1389450118666170711151518.

Karlberg T, Klepsch M, Thorsell AG, Andersson CD, Linusson A, Schüler H. 2015. Structural basis for lack of ADP-ribosyltransferase activity in poly(ADP-ribose) polymerase-13/zinc finger antiviral protein. *Journal of Biological Chemistry* **290**:7336–7344 DOI 10.1074/jbc.M114.630160.

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**:1160–1166 DOI 10.1093/bib/bbx108.

Ke Y, Wang C, Zhang J, Zhong X, Wang R, Zeng X, Ba X. 2019. The role of PARPs in inflammation-and metabolic-related diseases: molecular mechanisms and beyond. *Cell* **8**:1047 DOI 10.3390/cells8091047.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**:845–858 DOI 10.1038/nprot.2015.053.

Kim DS, Challa S, Jones A, Kraus WL. 2020. PARPs and ADP-ribosylation in RNA biology: from RNA expression and processing to protein translation and proteostasis. *Genes and Development* **34**:302–320 DOI 10.1101/gad.334433.119.

Kim IK, Stegeman RA, Brosey CA, Ellenberger T. 2015. A quantitative assay reveals ligand specificity of the DNA scaffold repair protein XRCC1 and efficient disassembly of complexes of XRCC1 and the poly(ADP-ribose) polymerase 1 by poly(ADP-ribose) glycohydrolase. *Journal of Biological Chemistry* **290**:3775–3783 DOI 10.1074/jbc.M114.624718.

Kim L, Kwon DH, Kim BH, Kim J, Park MR, Park ZY, Song HK. 2018. Structural and biochemical study of the Mono-ADP-Ribosyltransferase Domain of SdeA, a Ubiquitylating/Deubiquitylating Enzyme from Legionella pneumophila. *Journal of Molecular Biology* **430**:2843–2856 DOI 10.1016/j.jmb.2018.05.043.

Klockgether J, Tümmler B. 2017. Recent advances in understanding Pseudomonas aeruginosa as a pathogen. *F1000 Research* **6**:1261 DOI 10.12688/f1000research.10506.1.

Kobe B, Deisenhofer J. 1993. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **366**:751–756 DOI 10.1038/366751a0.

Kobe B, Deisenhofer J. 1994. The leucine-rich repeat: a versatile binding motif. *Trends in Biochemical Sciences* **19**:415–421 DOI 10.1016/0968-0004(94)90090-6.

Kobe B, Deisenhofer J. 1995. Proteins with leucine-rich repeats. *Current Opinion in Structural Biology* **5**:409–416 DOI 10.1016/0959-440x(95)80105-7.

Komander D, Randow F. 2017. Strange new world: bacteria catalyze ubiquitylation via ADP ribosylation. *Cell Host & Microbe* **21**:127–129 DOI 10.1016/j.chom.2017.01.014.

Krishnakumar R, Kraus WL. 2010. PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway. *Molecular Cell* **39**:736–749 DOI 10.1016/j.molcel.2010.08.014.

Kummar S, Chen A, Parchment RE, Kinders RJ, Ji J, Tomaszewski JE, Doroshow JH. 2012. Advances in using PARP inhibitors to treat cancer. *BMC Medicine* **10**:25 DOI 10.1186/1741-7015-10-25.

Kunze FA, Hottiger MO. 2019. Regulating immunity via ADP-Ribosylation: therapeutic implications and beyond. *Trends in Immunology* **40**:159–173 DOI 10.1016/j.it.2018.12.006.

Lawaree E, Jankevicius G, Cooper C, Ahel I, Uphoff S, Tang CM. 2020. DNA ADP-ribosylation stalls replication and is reversed by RecF-mediated homologous recombination and nucleotide excision repair. *Cell Reports* **30**:1373–1384 DOI 10.1016/j.celrep.2020.01.014.

Letunic I, Bork . 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* **44**:W242–W245 DOI 10.1093/nar/gkw290.

Leung A, Todorova T, Ando Y, Chang . 2012. Poly(ADP-ribose) regulates post-transcriptional gene regulation in the cytoplasm. *RNA Biology* **9**:542–548 DOI 10.4161/rna.19899.

Liang YC, Hsu CY, Yao YL, Yang WM. 2013. PARP-2 regulates cell cycle-related genes through histone deacetylation and methylation independently of poly(ADP-ribosyl)ation. *Biochemical and Biophysical Research Communications* **431**:58–64 DOI 10.1016/j.bbrc.2012.12.092.

Liu C, Yu X. 2015. ADP-ribosyltransferases and poly ADP-ribosylation. *Current Protein & Peptide Science* **16**:491–501 DOI 10.2174/1389203716666150504122435.

Livingstone CD, Barton GJ. 1993. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Computer Applications in the Biosciences* **9**:745–756.

Lyons B, Lugo MR, Carlin S, Lidster T, Merrill AR. 2018. Characterization of the catalytic signature of Scabin toxin, a DNA-targeting ADP-ribosyltransferase. *Biochemical Journal* **475**:225–245 DOI 10.1042/bcj20170818.

Malyuchenko NV, Kotova EY, Kulaeva OI, Kirpichnikov MP, Studitskiy VM. 2015. PARP1 inhibitors: antitumor drug design. *Acta Naturae* **7**:27–37 DOI 10.32607/20758251-2015-7-3-27-37.

Mangerich A, Bürkle A. 2012. Pleiotropic cellular functions of PARP1 in longevity and aging: genome maintenance meets inflammation. *Oxidative Medicine and Cellular Longevity* **2012**:321653 DOI 10.1155/2012/321653.

Maresso AW, Deng Q, Pereckas MS, Wakim BT, Barbieri JT. 2007. Pseudomonas aeruginosa ExoS ADP-ribosyltransferase inhibits ERM phosphorylation. *Cellular Microbiology* **9**:97–105 DOI 10.1111/j.1462-5822.2006.00770.x.

Marsischky GT, Wilson BA, Collier RJ. 1995. Role of glutamic acid 988 of human poly-ADP-ribose polymerase in polymer formation, Evidence for active site similarities to the ADP-ribosylating toxins. *Journal of Biological Chemistry* **270**:3247–3254 DOI 10.1074/jbc.270.7.3247.

Martinon F, Tschopp J. 2007. Inflammatory caspases and inflammasomes: master switches of inflammation. *Cell Death and Differentiation* **14**:10–22 DOI 10.1038/sj.cdd.4402038.

Matsushima N, Mikami T, Tanaka T, Miyashita H, Yamada K, Kuroki Y. 2009. Analyses of non-leucine-rich repeat (non-LRR) regions intervening between LRRs in proteins. *Biochimica et Biophysica Acta/General Subjects* **1790**:1217–1237 DOI 10.1016/j.bbagen.2009.06.014.

McCann KE. 2019. Advances in the use of PARP inhibitors for BRCA1/2-associated breast cancer: talazoparib. *Future Oncology* **15**:1707–1715 DOI 10.2217/fon-2018-0751.

Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research* **45**:D183–D189 DOI 10.1093/nar/gkw1138.

Munnur D, Ahel I. 2017. Reversible mono-ADP-ribosylation of DNA breaks. *FEBS Journal* **284**:4002–4016 DOI 10.1111/febs.14297.

Munro P, Benchetrit M, Nahori MA, Stefani C, Clément R, Michiels JF, Landraud L, Dussurget O, Lemichez E. 2010. The Staphylococcus aureus epidermal cell differentiation inhibitor toxin promotes formation of infection foci in a mouse model of bacteremia. *Infection and Immunity* **78**:3404–3411 DOI 10.1128/iai.00319-10.

Murphy JM, Mace PD, Eyers PA. 2017. Live and let die: insights into pseudoenzyme mechanisms from structure. *Current Opinion in Structural Biology* **47**:95–104 DOI 10.1016/j.sbi.2017.07.004.

Ng AC, Eisenberg JM, Heath RJ, Huett A, Robinson CM, Nau GJ, Xavier RJ. 2011. Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proceedings of the National Academy of Sciences of the United States of America* **108(Suppl 1)**:4631–4638 DOI 10.1073/pnas.1000093107.

Ng A, Xavier RJ. 2011. Leucine-rich repeat (LRR) proteins: integrators of pattern recognition and signaling in immunity. *Autophagy* **7**:1082–1084 DOI 10.4161/auto.7.9.16464.

Noguchi A, Adachi S, Yokota N, Hatta T, Natsume T, Kawahara H. 2018. ZFP36L2 is a cell cycle-regulated CCCH protein necessary for DNA lesion-induced S-phase arrest. *Biology Open* **7**:bio031575 DOI 10.1242/bio.031575.

Offord V, Coffey TJ, Werling D. 2010. LRRfinder: a web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database. *Developmental and Comparative Immunology* **34**:1035–1041 DOI 10.1016/j.dci.2010.05.004.

O'Sullivan J, Ferreira MTedim, Gagne JP, Sharma AK, Hendzel MJ, Masson JY, Poirier GG. 2019. Emerging roles of eraser enzymes in the dynamic control of protein ADP-ribosylation. *Nature Communications* **10**:1182 DOI 10.1038/s41467-019-08859-x.

Otto H, Reche PA, Bazan F, Dittmar K, Haag F, Koch-Nolte F. 2005. In silico characterization of the family of PARP-like poly(ADP-ribosyl)transferases (pARTs). *BMC Genomics* **6**:139 DOI 10.1186/1471-2164-6-139.

Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R, Zhang F, Dolma S, Willems A, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M. 2019. The BioGRID interaction database: 2019 update. *Nucleic Acids Research* **47**:D529–D541 DOI 10.1093/nar/gky1079.

Pawłowski K, Lepisto M, Meinander N, Sivars U, Varga M, Wieslander E. 2006. Novel conserved hydrolase domain in the CLCA family of alleged calcium-activated chloride channels. *Proteins-Structure Function and Bioinformatics* **63**:424–439 DOI 10.1002/prot.20887.

Pawlowski K, Pio F, Chu Z, Reed JC, Godzik A. 2001. PAAD - a new protein domain associated with apoptosis, cancer and autoimmune diseases. *Trends in Biochemical Sciences* **26**:85–87 DOI 10.1016/S0968-0004(00)01729-1.

Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X. 2009. DOG 1.0: illustrator of protein domain structures. *Cell Research* **19**:271–273 DOI 10.1038/cr.2009.6.

Ribeiro AJM, Das S, Dawson N, Zaru R, Orchard S, Thornton JM, Orengo C, Zeqiraj E, Murphy JM, Eyers PA. 2019. Emerging concepts in pseudoenzyme classification, evolution, and signaling. *Science Signaling* **12**:eaat9797 DOI 10.1126/scisignal.aat9797.

Riccio AA, Cingolani G, Pascal JM. 2016. PARP-2 domain requirements for DNA damage-dependent activation and localization to sites of DNA damage. *Nucleic Acids Research* **44**:1691–1702 DOI 10.1093/nar/gkv1376.

Rosado MM, Bennici E, Novelli F, Pioli C. 2013. Beyond DNA repair, the immunological role of PARP-1 and its siblings. *Immunology* **139**:428–437 DOI 10.1111/imm.12099.

Rossi ML, Ghosh AK, Bohr VA. 2010. Roles of Werner syndrome protein in protection of genome integrity. *DNA Repair* **9**:331–344 DOI 10.1016/j.dnarep.2009.12.011.

Roy A, Zhang Y. 2012. Recognizing protein-ligand binding sites by global structural alignment and local geometry refinement. *Structure* **20**:987–997 DOI 10.1016/j.str.2012.03.009.

Saikatendu KS, Joseph JS, Subramanian V, Clayton T, Griffith M, Moy K, Velasquez J, Neuman BW, Buchmeier MJ, Stevens RC, Kuhn . 2005. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1-phosphate dephosphorylation by a conserved domain of nsP3. *Structure* **13**:1665–1675 DOI 10.1016/j.str.2005.07.022.

Samaras P, Schmidt T, Frejno M, Gessulat S, Reinecke M, Jarzab A, Zecha J, Mergner J, Giansanti P, Ehrlich HC, Aiche S, Rank J, Kienegger H, Krcmar H, Kuster B, Wilhelm M. 2020. ProteomicsDB: a multi-omics and multi-organism resource for life science research. *Nucleic Acids Research* **48**:D1153–D1163 DOI 10.1093/nar/gkz974.

Simon NC, Aktories K, Barbieri JT. 2014. Novel bacterial ADP-ribosylating toxins: structure and function. *Nature Reviews. Microbiology* **12**:599–611 DOI 10.1038/nrmicro3310.

Sreelatha A, Yee SS, Lopez V, Park B, Pilch S, Servage KA, Zhang J, Jiou J, Kinch LN, Karasiewicz M, Łobocka M, Grishin NV, Orth K, Kucharczyk R, Pawłowski K, Tomchick DR, Tagliabracci VS. 2018. Protein AMPylation by an evolutionarily conserved pseudokinase. *Cell* **175**:809–821 DOI 10.1016/j.cell.2018.08.046.

Suskiewicz MJ, Zobel F, Ogden TEH, Fontana P, Ariza A, Yang JC, Zhu K, Bracken L, Hawthorne WJ, Ahel D, Neuhaus D, Ahel I. 2020. HPF1 completes the PARP active site for DNA damage-induced ADP-ribosylation. *Nature* **579**:598–602 DOI 10.1038/s41586-020-2013-6.

Szántó M, Brunyánszki A, Márton J, Vámosi G, Nagy L, Fodor T, Kiss B, Virág L, Gergely P, Bai P. 2014. Deletion of PARP-2 induces hepatic cholesterol accumulation and decrease in HDL levels. *Biochimica et Biophysica Acta/General Subjects* **1842**:594–602 DOI 10.1016/j.bbadis.2013.12.006.

Tchasovnikarova IA, Timms RT, Matheson NJ, Wals K, Antrobus R, Göttgens B, Dougan G, Dawson MA, Lehner PJ. 2015. GENE SILENCING, Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* **348**:1481–1485 DOI 10.1126/science.aaa7227.

Teloni F, Altmeyer M. 2016. Readers of poly(ADP-ribose): designed to be fit for purpose. *Nucleic Acids Research* **44**:993–1006 DOI 10.1093/nar/gkv1383.

Thul PJ, Lindskog C. 2018. The human protein atlas: a spatial map of the human proteome. *Protein Science* **27**:233–244 DOI 10.1002/pro.3307.

Trott O, Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**:455–461 DOI 10.1002/jcc.21334.

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, Feilitzen Kvon, Forsberg M, Persson L, Johansson F, Zwahlen M, Heijne Gvon, Nielsen J, Ponten F. 2015. Proteomics. Tissue-based map of the human proteome. *Science* **347**:1260419 DOI 10.1126/science.1260419.

Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M. 2012. Combining global and local measures for structure-based druggability predictions. *Journal of Chemical Information and Modeling* **52**:360–372 DOI 10.1021/ci200454v.

Wang M, Herrmann CJ, Simonovic M, Szklarczyk D, Von Mering C. 2015. Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**:3163–3168 DOI 10.1002/pmic.201400441.

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:1189–1191 DOI 10.1093/bioinformatics/btp033.

Weaver AN, Yang ES. 2013. Beyond DNA repair: additional functions of PARP-1 in cancer. *Frontiers in Oncology* **3**:290 DOI 10.3389/fonc.2013.00290.

Webb B, Sali A. 2017. Protein structure modeling with MODELLER. *Methods in Molecular Biology* **1654**:39–54 DOI 10.1007/978-1-4939-7231-9_4.

Xu D, Jaroszewski L, Li Z, Godzik A. 2014. FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics* **30**:660–667.

Xu D, Jaroszewski L, Li Z, Godzik A. 2015. AIDA: ab initio domain assembly for automated multi-domain protein structure prediction and domain-domain interaction prediction. *Bioinformatics* **31**:2098–2105 DOI 10.1093/bioinformatics/btv092.

Yamamoto M, Yamasaki M, Tsukao Y, Tanaka K, Miyazaki Y, Makino T, Takahashi T, Kurokawa Y, Nakajima K, Takiguchi S, Mori M, Doki Y. 2017. Poly (ADP-ribose) polymerase-1 inhibition decreases proliferation through G2/M arrest in esophageal squamous cell carcinoma. *Oncology Letters* **14**:1581–1587 DOI 10.3892/ol.2017.6334.

Yang CS, Jividen K, Spencer A, Dworak N, Ni L, Oostdyk LT, Chatterjee M, Kuśmider B, Reon B, Parlak M, Gorbunova V, Abbas T, Jeffery E, Sherman NE, Paschal BM. 2017. Ubiquitin modification by the E3 Ligase/ADP-Ribosyltransferase Dtx3L/Parp9. *Molecular Cell* **66**:503–516 DOI 10.1016/j.molcel.2017.04.028.

Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. 2020. Improved protein structure prediction using predicted interresidue orientations. *Proceedings*

of the National Academy of Sciences of the United States of America **117**:1496–1503 DOI 10.1073/pnas.1914677117.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nature Methods* **12**:7–8 DOI 10.1038/nmeth.3213.

Yonemori K, Seki N, Kurahara H, Osako Y, Idichi T, Arai T, Koshizuka K, Kita Y, Maemura K, Natsugoe S. 2017. ZFP36L2 promotes cancer cell aggressiveness and is regulated by antitumor microRNA-375 in pancreatic ductal adenocarcinoma. *Cancer Science* **108**:124–135 DOI 10.1111/cas.13119.

Zahn-Zabal M, Michel PA, Gateau A, Nikitin F, Schaeffer M, Audot E, Gaudet P, Duek PD, Teixeira D, Rechde Laval V, Samarasinghe K, Bairoch A, Lane L. 2020. The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Research* **48**:D328–D334 DOI 10.1093/nar/gkz995.

Zhang C, Freddolino PL, Zhang Y. 2017. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Research* **45**:W291–W299 DOI 10.1093/nar/gkx366.

Zhang L, Chen F, Zhang X, Li Z, Zhao Y, Lohaus R, Chang X, Dong W, Ho SYW, Liu X, Song A, Chen J, Guo W, Wang Z, Zhuang Y, Wang H, Chen X, Hu J, Liu Y, Qin Y, Wang K, Dong S, Liu Y, Zhang S, Yu X, Wu Q, Wang L, Yan X, Jiao Y, Kong H, Zhou X, Yu C, Chen Y, Li F, Wang J, Chen W, Chen X, Jia Q, Zhang C, Jiang Y, Zhang W, Liu G, Fu J, Chen F, Ma H, Van de Peer Y, Tang H. 2020. The water lily genome and the early evolution of flowering plants. *Nature* **577**:79–84 DOI 10.1038/s41586-019-1852-5.

Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V. 2017. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *Journal of Molecular Biology* DOI 10.1016/j.jmb.2017.12.007.