

A novel predicted ADP-ribosyltransferase family conserved in eukaryotic evolution

Zbigniew Wyżewski^{2*}, *Marcin Gradowski*^{1*}, *Marianna Krysińska*¹, *Małgorzata Dudkiewicz*^{1#}
and *Krzysztof Pawłowski*^{1#}

¹ Warsaw University of Life Sciences SGGW, Warszawa, Poland

² Cardinal Wyszyński University, Warszawa, Poland

* these authors contributed equally to this work

corresponding authors:

malgorzata_dudkiewicz@sggw.edu.pl

krzysztof_pawlowski@sggw.edu.pl

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 involved in innate immune surveillance.

The novel family appears as likely novel ADP-ribosylation “writers”, previously unnoticed new players in cell signaling by this emerging post-translational modification. 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 signalling.

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 target molecules such as proteins, nucleic acids and small molecules. The 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 of functions performed by ARTs [1-4].

Prokaryotic ADP-ribosyltransferases often play a role of bacterial toxins or effectors. Several studies show that ADP-ribosylation of the host target molecules by ARTs from taxonomically diverse bacteria may be correlated with the development of infection as its essential step [5-9]. At the intracellular scale, it may change apoptotic potential of infected cell [10] and/or disturb organization of cellular membranes and actin cytoskeleton [11-13] whereas at systemic level, it can disturb cell-mediated immune response [6, 10], or increase the permeability of barriers limiting bacterial spread (epithelium and endothelium) [8, 10] and consequently contribute to serious structural and functional disorders of the host tissues and organs [14, 15]. Some ARTs act on small molecules, e.g. the bacterial Arr enzymes that ADP-ribosylate the antibiotics rifamycins [16].

Eukaryotic ADP-ribosyltransferases play important roles in both physiological and pathophysiological processes, contributing to change 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 [17, 18]. ADP-ribose moieties may function as signals that direct modified acceptors to ubiquitin-dependent proteolysis [1, 3]. Influencing the half-life of the target molecules, ADP-ribosyltransferases determine their intracellular level and thus affect their activity [3]. 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 [19]. Poly-ADP-ribosylation of DNA-binding proteins (e.g. core histones [20, 21]), chromatin remodeling enzymes (e.g. histone demethylase KDM5B [22], DEK protein [23] and DNA repair factors [24]) influences genome organization [21] and expression [22, 23] as well as effective DNA repair [20, 25]. Recently, growing evidence shows that PARPs also directly ADP-ribosylate mRNA [26]. Other examples of nucleic acid ADP-ribosylation are provided by the bacterial toxin DarT that acts on ssDNA [27]

and the toxin scabin, acting on mononucleosides, nucleotides, and both single-stranded and dsDNA [28].

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 [29-32]. PARP1 and PARP2 transferase activity is known to be necessary for repair of DNA damage such as deamination, hydroxylation and methylation of nitrogenous bases as well as for repair of single strand breaks. PARP1 and PARP also play a role in maintaining telomere stability[33, 34].

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 [33-36]. On the other hand, ADP-ribosyltransferases may also be involved in development of pathological states such as neurodegenerative disorders, diabetes, atherosclerosis, cataract and cancer [37-42]. 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) [43].

According to the Pfam database, the ART clan (superfamily) comprises 19 families of domains, eight of which, ART, DUF952, Enterotoxin_a, PARP, Pertussis_S1, PTS_2-RNA, RolB_RolC and TNT, include eukaryotic members [44]. 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). [1, 3]. 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 [45] 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 [1]. Three families of the ART clan, PARP, PTS_2-RNA and ART, are present in humans, and contain 16, 1 and 4 representatives, respectively. The NEURL-4 ART-like domain

family [46], 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 [5, 47, 48], novel ART/macro pairs in bacterial toxin/antitoxin systems [49] and novel human macro domains [50]. Other examples of the novel ADP-ribosylation players are provided by the viral macro domains, present in many dangerous viruses, including the SARS and SARS-CoV-2 coronaviruses [51], and by the recently characterized novel ART-like domain (DUF3715) in the human TASOR protein involved in gene silencing [52, 53].

In this paper, building up on experience in identification of novel enzyme families [50, 54-57], 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. 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 [58], HHpred [59] and Phyre2 [60] 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 [61] 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 [44]. 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 [62] was used to visualize results as sequence logos.

The CLANS [63] 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 four sequence sets representing the following ART-like domain families not included in Pfam database: ESPJ, SidE, 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 [64].

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 [65] was utilized to construct phylogenetic tree. A multiple sequence alignment was generated using the MUSCLE program [66]. Next, advanced mode of the platform was utilized to build phylogenetic tree using the maximum likelihood PhyML program [67] 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 [68].

Protein three-dimensional models for hypothetical structures of the human LRRC9-ART domain were built using homology modelling method Modeller9v21 [69] and were based on human PARP10 structure template (PDB ID: 3HKV).

Putative ligand binding site in the modelled structure was identified using the COFACTOR algorithm [70], an option delivered by the I-Tasser (Iterative Threading ASSEmbly Refinement) server for protein structure and function prediction [71] 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. Structures were visualized and analyzed using UCSF Chimera tools [72].

The three-domain model for whole human LRRC9 sequence was built using AIDA (Ab Initio Domain Assembly Server) [45] based on structures modelled using Modeller9v21 separately for three identified domains: N-terminal LRR region with helical fragment (PDB code of modelling template: 3OJA), ART-like domain (template: 3HKV), and C-terminal LRR region domain (template: 4LSX).

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 [73] and were automatically calculated as the number of conserved physicochemical properties for each column of the alignment [74].

The LRR repeats were identified using the LRRFinder tool [75]. 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 [76]. The domain organization of LRRC9 proteins was visualised using the DOG tool [77].

For analysis of protein and mRNA expression, the Proteomics-db and Protein Atlas databases were used [78, 79]. The BioGRID Interaction Database was used to obtain information about possible place of LRRC9 in human protein-protein interaction networks [80]. For prediction of the subcellular localization of LRRC9, the DeepLoc server was used [81].

Results and discussion

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

A sequence 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 (Fig. 1). 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 (Table 1). Such remote sequence similarity points to possible enzymatic function of the putative ART domain.

The relationship of the novel LRRC9 ART-like family, found in many eukaryotic lineages (Fig. 2), to the known ART families can be visualized using the sequence-based CLANS clustering approach (Fig. 3). CLANS analysis was performed at three different E-value levels for BLAST hits for all known families containing ADP-ribosyltransferase domain. 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 1, $1e^{-2}$ and even at $1e^{-4}$. This, together with sequence conservation analysis allowed us to hypothesize that LRRC9-ART belongs to the H-Y-[EDQ] clade.

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] [1] within β -strand 1, is responsible for binding the 2-OH group of the adenosine ribose of NAD^+ and NH_2 group of the nicotinamide amide *via* hydrogen bonds [3]. Tyrosine (motif II), localized within β -strand 2 [1], 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 [3].

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) among LRRC9-ART domains. 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 [82]. Some authors [82] include this second Tyr in the group of conserved residues being the hallmark of PARP ART-domains [82], 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. According to our results, this second tyrosine was 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; there are some suboptimal alignment paths to be considered (Suppl. Fig 1). A PARP region flanked by two tyrosines belonging to catalytic tetrad HYYE is aligned to a poorly conserved region 525-538 of the LRRC9 ART-like domain, 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 alignment 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-x(5)-PEY-(2)-EFEY motif in LRRC9 (logo positions 609-623), although not aligned by FFAS to PARP active site, may be speculated to be functionally important.

Structural analysis of the homology model

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 [83] 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 [83]. In PARP10 and PARP12 with mono-ADP-ribosylation (MAR) activity, the glutamic acid is substituted by isoleucine and instead of a 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). Histidine and glycine responsible for nicotinamide anchoring and binding are replaced by tyrosine and valine (Figure 5). 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 residues (604), there are poorly conserved Glu and Lys. Figure 6 presents a structure model of LRRC9 ART-like domain putative ligand binding site with hypothetical ligand molecule location according to COFACTOR predictions. COFACTOR is a functional annotation tool based on threading of the protein structure query through the BioLiP function database of ligand-protein binding interactions by local and global structure matches to identify functional sites and homologies [84]. 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 structure 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 moiety, is smaller and probably fits better in the LRRC9-ART model 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 nicotinamide adenine dinucleotide (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.

Replacing the catalytic Glu residue of the HYYE tetrad by Ile is characteristic for PARP family members having mono-ADP-ribosylation (MAR) activity instead of poly-ADP-ribosylation (PAR) [82]. Among FFAS/HHpred hits for LRRC9-ART domain, the only structures that showed up were proteins with MAR activity and PARP13 which has no confirmed ADP-ribosylation activity. 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 HYYE tetrad is replaced by glutamine 525, which is noticeably less conserved. However, five residues further towards the N-terminus, a conserved motif [ED]-x-P-R-[IL] can be seen. At position 532 in LRRC9-ART domain logo which corresponds to catalytic PARP

Lys (substituted by Leu in MAR enzymes) there is a lysine, which could take over some catalytic function in LRRC9 protein. Another highly conserved motif of the PARP family, YxEY[VI][VI][FY], containing catalytic Glu of the HYEE 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 misalignment may be just an artifact.

Figure 6 shows the active site of a modelled LRRC9-ART structure, with AMP molecule docked in the predicted active site. Fig. 6A presents the same residues as Figure 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 classical HYEE tetrad residues are not highly conserved. Fig. 6B presents all strongly conserved residues in the neighborhood of the putative LRRC9 ART-like domain active site. It demonstrates that there are some other conserved amino acids within 5 Å of predicted ligand location, like Phe473 and Leu472 (near YV dyad aligned to PARP HG), Ile490, close to adenosine ring of the docked ligand, Phe495 (part of conserved ExGF motif flanking ligand position near N-ribose ring) and Trp605 (next to Lys604 that replaces the catalytic Glu). Ligand pose predicted by COFACTOR enables only two hydrogen bonds between AMP and residues aligned to PARP ART-like domain catalytic tetrad: between AMP phosphate and Val 476 and between N-ribose ring and Glu 525. This method of docking 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 with coloring by within-family conservation (Fig. 6C) shows the conserved residues are grouped in the vicinity of the putative ligand binding groove.

The ART-like domain described here is only one region of the whole LRRC9 protein molecule, which is quite large, with clear similarity to leucine-rich repeats containing domains in its N- and C-terminal parts and a very long helical “spine” in the center. We were tempted to construct the whole molecule model, based on homology to LRR-containing templates and using an already modelled ART-like domain structure. The results are shown in Figure 7AB. 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 [85]), *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 [86]. Altogether, this taxonomic spread suggests an ancient function for LRRC9, common to many diverse eukaryotes.

Likely biological processes involving LRRC9

Apart from the ART-like domain, LRRC9 proteins comprise a variable number of leucine-rich repeats (LRRs) flanking the ART domain (Fig. 4). 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) [87]. 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 [88]. LRR motifs determine functionality of LRR-containing proteins, enabling them to make protein-protein interactions [87] as well as to bind various target structures such as small molecule hormones, lipids and nucleic acids [89]. The non-globular shape of LRR motifs may facilitate the contact between LRR-containing proteins and the target structures, increasing the interaction area [87]. 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 [90].

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 [91]. 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 [83].

The LRR domains, which are structures strongly 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, i.e. viral, bacterial, fungal and parasite antigens [92, 93]. On the other hand, also ADP-ribosylation is known to be involved in regulation of the host immune response [94, 95]. It may direct intracellular signaling to parthanatos (type of programmed death) [96-98], whereas such scenario implies induction of inflammatory response to infection. Poly-ADP-ribosylation is known to influence the stability of transcripts encoding proinflammatory cytokines [99]. PARP-dependent chromatin modification may counteract progression of viral infection. ADP-ribosylation also influences other innate immune mechanisms such as NF- κ B expression, phagocytosis and macrophage polarization [100]. 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 functional annotation analysis. 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), cellular response to hormone stimulus (7 proteins) as well as phosphatidylinositol 3-kinase signaling (6 proteins).

Many human LRR-containing proteins undergo extracellular secretion (36 proteins), and many are integral components of plasma membrane (25 proteins).

The ART-like domain combined with LRR motifs has a precedent, since, interestingly, 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 [101, 102]. 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.

Human LRRC9 protein is expressed mainly in the heart, tonsil as well as in the liver (Proteomics-db database). However, no high quality protein expression data is available for LRRC9, hence it is annotated as “existence validated on transcript level” in the NextProt database. This may suggest that LRRC9 protein expression is limited to specific circumstances, and is not easily captured in typical tissue proteomics experiments. 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). According to the DeepLoc prediction, the subcellular localization of LRRC9 is cytoplasmic (likelihood 0.79). Immunohistochemical staining revealed strong cytoplasmic expression of human LRRC9 in the gastrointestinal tract of a few patients suffering from colorectal cancer. Moreover, medium expression of the protein was shown in subjects with liver cancer (expression within liver/gall bladder), stomach cancer, testis cancer, renal and urothelial cancer (kidney/urinary bladder) and melanoma (Protein Atlas database).

For an uncharacterised 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 [103]. ZFP36L2 is

an RNA-binding protein regulating cell cycle [104] that contributes to pathogenesis of pancreatic ductal adenocarcinoma (PDAC). ZFP36L2 is responsible for an increase in cancer cell aggressiveness [105]. 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.

Conclusions

Pseudoenzymes are characterized by the lack of enzymatic activity despite the common evolutionary origin and structural similarity to catalytically active homologues [106-108]. 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 [83]. It has been shown that substitution of catalytic His with Ala within the ART domain of PARP1 substantially decreases its catalytic activity [109]. 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 [110]. 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 [57, 111].

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- κ B-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 [21, 112]. 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. K.P. was supported by the Polish National Agency for Scientific Exchange scholarship PPN/BEK/2018/1/00431 and by the Polish National Science Centre grant 2019/33/B/NZ2/01409.

Table 1. Top structure predictions for ART-like domain of the human LRRC9 protein

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 [<i>Homo sapiens</i>]	E-value=6.2e-23	392-624	15%
Phyre2	c3HKVA , Human poly(ADP-ribose) polymerase 10, catalytic fragment	Confidence= 100%	395-624	16%

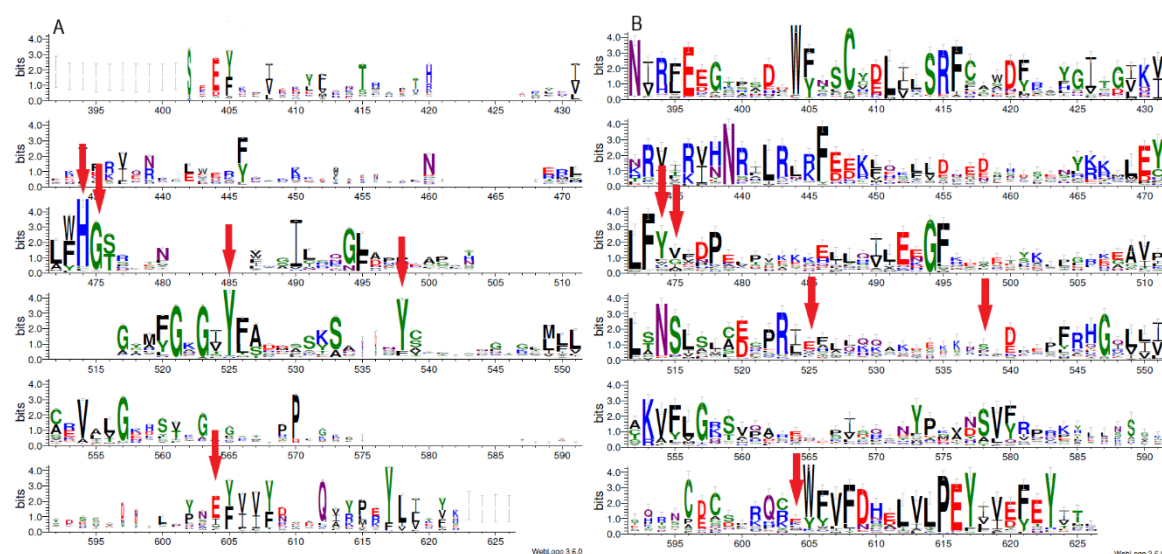


Fig. 1. Sequence logos for PARP homologues (A) and LRRC9 homologues (B), created from MAFFT alignment using WebLogo server. Sequence numbering according to human LRRC9 protein. The two logos are aligned using pairwise FFAS03 alignment. The catalytically important PARP residues His, Gly, Tyr, Tyr and Glu (aligned to positions 474, 475, 525, 538 and 604, respectively, of the human LRRC9 protein) marked with red arrows.

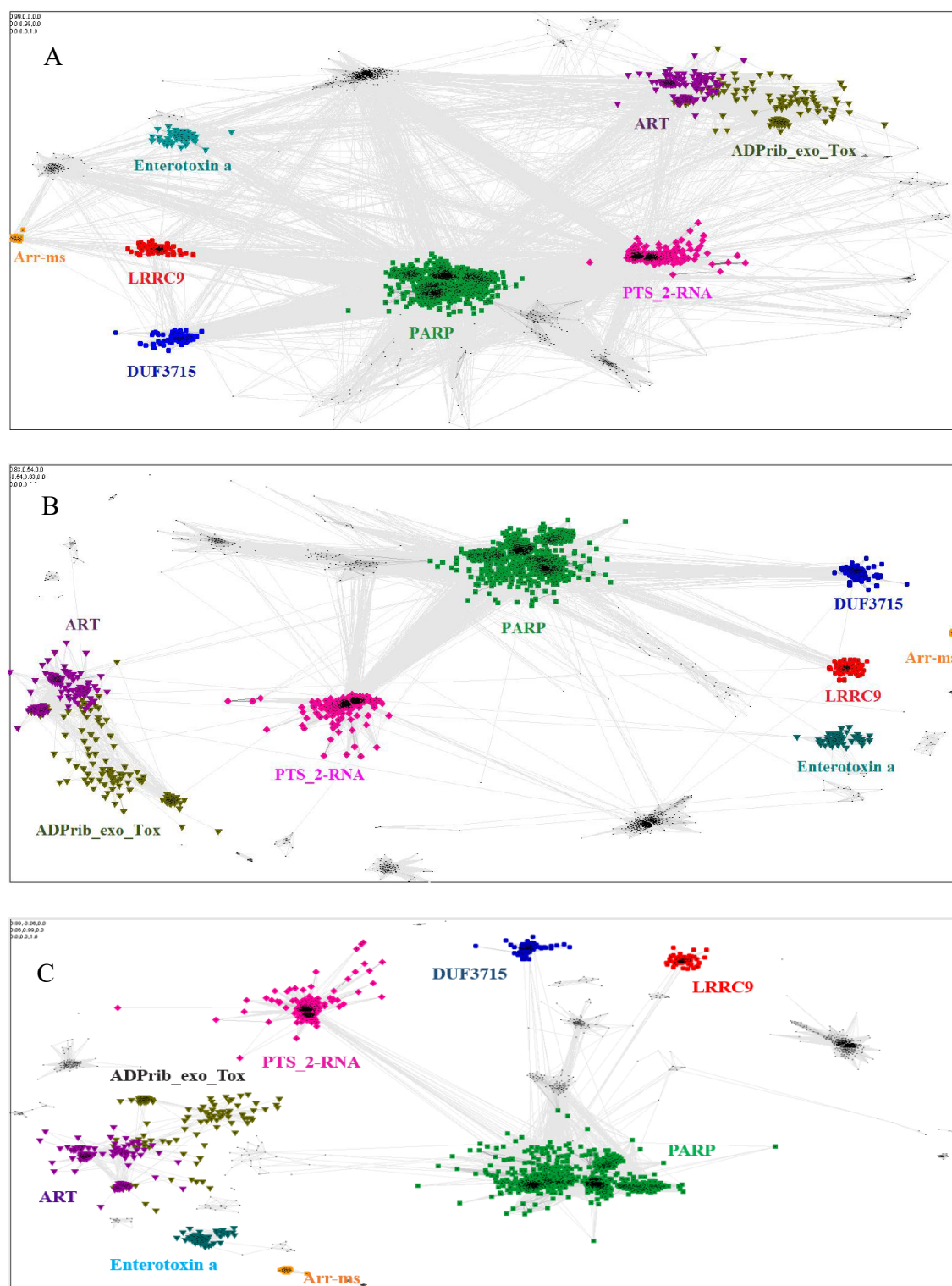


Fig. 3. Close and distant sequence similarities between the putative and the known ADP-ribosyltransferase domains, visualized using CLANS algorithm. A. sequence similarities up to BLAST E-value of A: 1, B: at E-value $1e-2$, C: at E-value $1e-4$

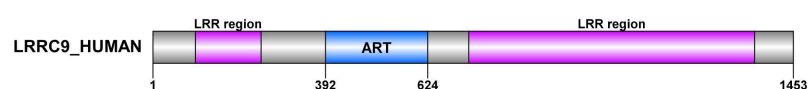


Fig. 4. Domain organization of LRRC9 proteins (created using the DOG tool).

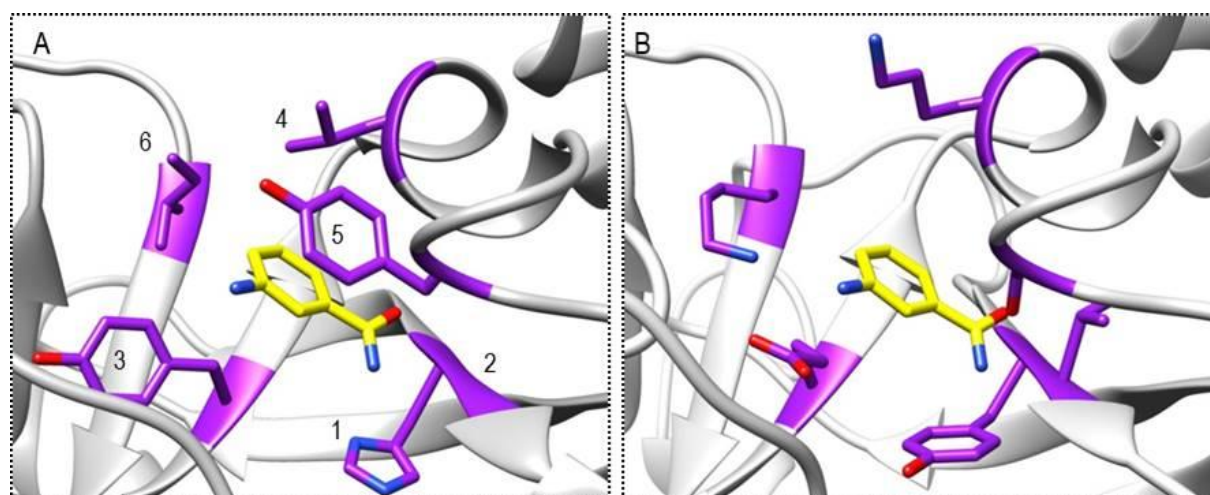
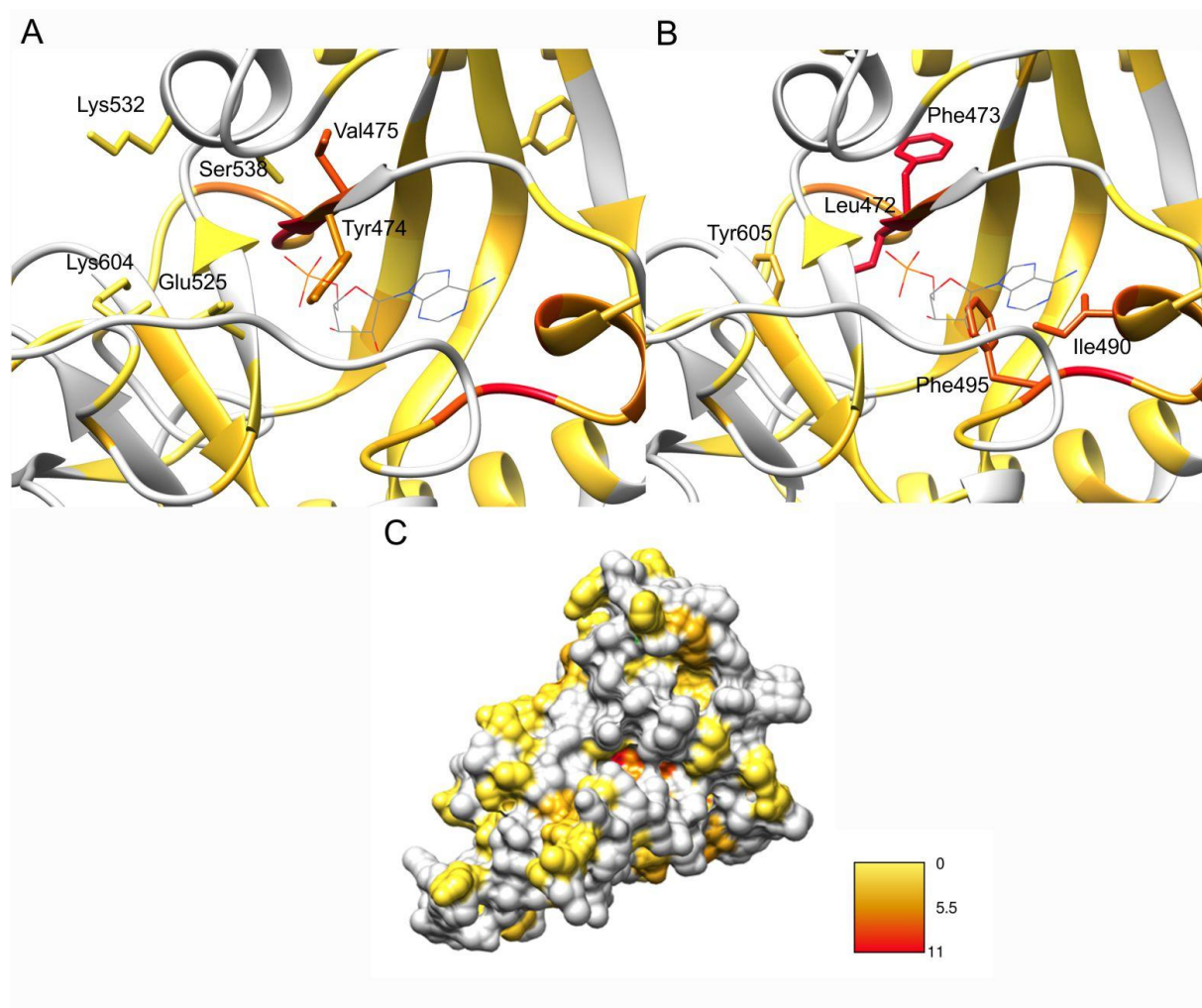


Fig. 5. Comparison of active site in the template structure (PARP10, 3HKV) (A) and modeled human LRRC9-ART domain. (B). Canonical PARP tetrad residues (listed in the table below) and their counterparts in the modelled structure are depicted in stick representation. The ligand is 3-aminobenzamide, a PARP inhibitor (yellow).

		H(1)	G(2)	Y(3)	K(L) (4)	Y (5)	E(I)(6)
A	PARP10	His 887	Gly 888	Tyr 919	Leu 926	Tyr 932	Ile 987
B	LRRC9	Tyr 474	Val 475	Glu 525	Lys 532	Ser 538	Lys 604



PARP family catalytic residues	PARP10 residue	Function in PARP10	LRRC9 residue aligned to PARP10 active site (A)	Conserved LRRC9 residues in vicinity of A (B)
H	His 888	Nicotinamide binding	Tyr 474	<i>Leu 472, Phe 473</i>
G	Gly 889	Nicotinamide anchoring	Val 475	-
Y	Tyr 919	N-ribose stacking	Glu 525	<i>Ile 490, Phe 495</i>
[KL]	Leu 926	L926 is replacing catalytic lysine of PARP1 and PARP2	Lys 532	-
Y	Tyr 932	Nicotinamide stacking	Ser 538	-
[EI]	Ile 987	Catalytic residue (in enzymes with MAR activity)	Lys 604	<i>Trp 605</i>

Fig. 6. Human LRRC9 homology model in ribbon representation colored according to amino acid conservation (red - highest) in multiple sequence alignment of LRRC9 homologues. An

AMP molecule docked in location and pose indicated by COFACTOR algorithm. A. LRRC9 residues aligned with PARP10 active site shown as sticks. B. LRRC9 conserved residues located in the vicinity of the active site shown as sticks. C. Surface representation of LRRC9 structure model with visible conserved residues in the putative binding groove.

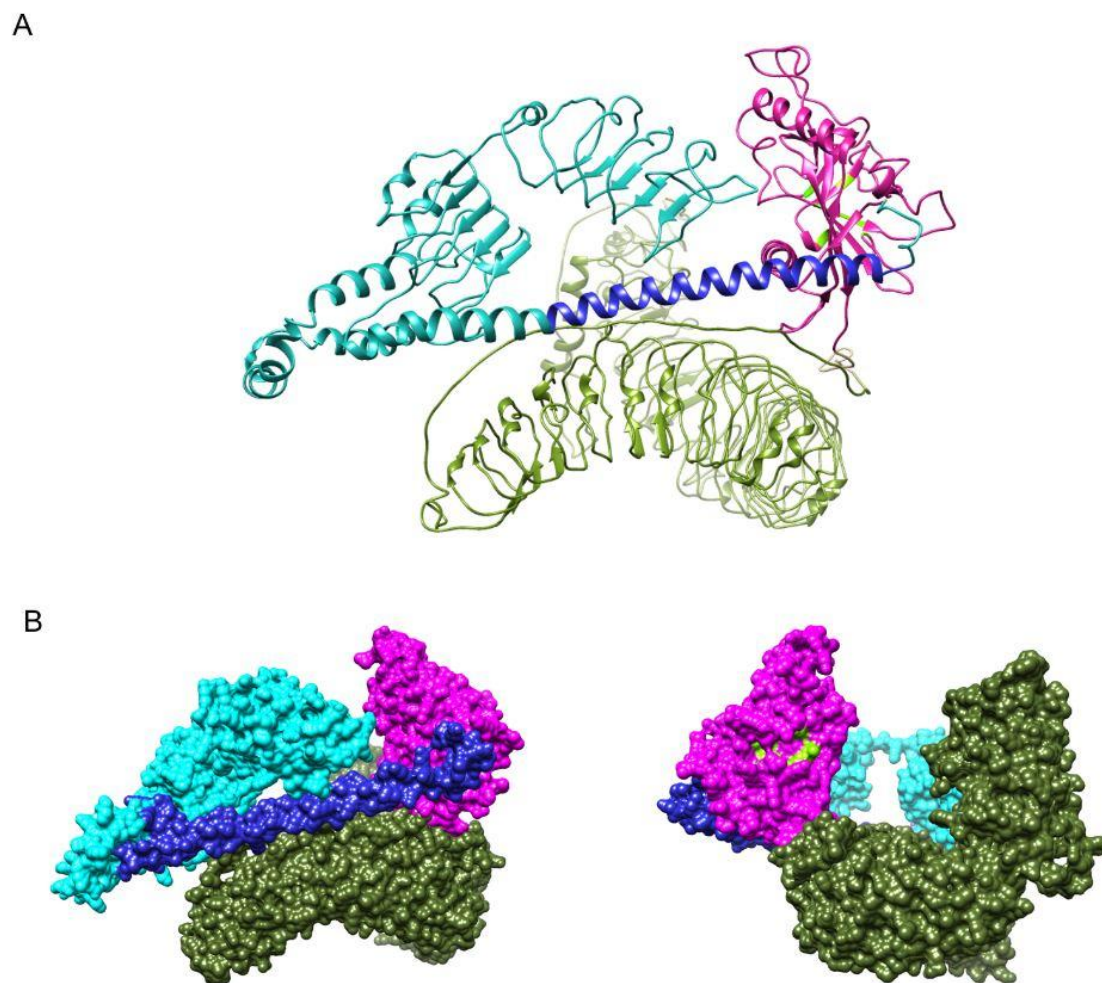


Fig. 7. A. Human LRRC9 structure model for the full-length protein in ribbon representation, colored by domain. ART-like domain (magenta) is stacked between two large leucine-rich repeat-containing domains (N-terminal LRR domain with long helical fragment is depicted in cyan and blue, C-terminal LRR domain - in green). B. The same model in surface representation (viewed from two directions).

References

1. Aravind, L., et al., *The natural history of ADP-ribosyltransferases and the ADP-ribosylation system*. Curr Top Microbiol Immunol, 2015. **384**: p. 3-32.
2. Teloni, F. and M. Altmeyer, *Readers of poly(ADP-ribose): designed to be fit for purpose*. Nucleic Acids Res, 2016. **44**(3): p. 993-1006.
3. Cohen, M.S. and P. Chang, *Insights into the biogenesis, function, and regulation of ADP-ribosylation*. Nat Chem Biol, 2018. **14**(3): p. 236-243.
4. Munnur, D. and I. Ahel, *Reversible mono-ADP-ribosylation of DNA breaks*. FEBS J, 2017. **284**(23): p. 4002-4016.
5. Kalayil, S., et al., *Insights into catalysis and function of phosphoribosyl-linked serine ubiquitination*. Nature, 2018.
6. Bhogaraju, S., et al., *Phosphoribosylation of Ubiquitin Promotes Serine Ubiquitination and Impairs Conventional Ubiquitination*. Cell, 2016. **167**(6): p. 1636-1649 e13.
7. Simon, N.C., K. Aktories, and J.T. Barbieri, *Novel bacterial ADP-ribosylating toxins: structure and function*. Nat Rev Microbiol, 2014. **12**(9): p. 599-611.
8. Boyer, L., et al., *Induction of transient macroapertures in endothelial cells through RhoA inhibition by Staphylococcus aureus factors*. J Cell Biol, 2006. **173**(5): p. 809-19.
9. Aktories, K., et al., *Actin as target for modification by bacterial protein toxins*. FEBS J, 2011. **278**(23): p. 4526-43.
10. Klockgether, J. and B. Tümmler, *Recent advances in understanding Pseudomonas aeruginosa as a pathogen*. F1000Res, 2017. **6**: p. 1261.
11. Kagan, J.C. and C.R. Roy, *Legionella phagosomes intercept vesicular traffic from endoplasmic reticulum exit sites*. Nat Cell Biol, 2002. **4**(12): p. 945-54.
12. Komander, D. and F. Randow, *Strange New World: Bacteria Catalyze Ubiquitylation via ADP Ribosylation*. Cell Host Microbe, 2017. **21**(2): p. 127-129.
13. Maresso, A.W., et al., *Pseudomonas aeruginosa ExoS ADP-ribosyltransferase inhibits ERM phosphorylation*. Cell Microbiol, 2007. **9**(1): p. 97-105.
14. Chiu, C.C., et al., *Pseudomonas aeruginosa exotoxin A-induced hepatotoxicity: an animal model in rats*. J Vet Med Sci, 2009. **71**(1): p. 1-8.
15. Munro, P., et al., *The Staphylococcus aureus epidermal cell differentiation inhibitor toxin promotes formation of infection foci in a mouse model of bacteremia*. Infect Immun, 2010. **78**(8): p. 3404-11.
16. Baysarowich, J., et al., *Rifamycin antibiotic resistance by ADP-ribosylation: Structure and diversity of Arr*. Proc Natl Acad Sci U S A, 2008. **105**(12): p. 4886-91.
17. Jwa, M. and P. Chang, *PARP16 is a tail-anchored endoplasmic reticulum protein required for the PERK- and IRE1 α -mediated unfolded protein response*. Nat Cell Biol, 2012. **14**(11): p. 1223-30.
18. Liu, C. and X. Yu, *ADP-ribosyltransferases and poly ADP-ribosylation*. Curr Protein Pept Sci, 2015. **16**(6): p. 491-501.
19. Leung, A., et al., *Poly(ADP-ribose) regulates post-transcriptional gene regulation in the cytoplasm*. RNA Biol, 2012. **9**(5): p. 542-8.
20. Javle, M. and N.J. Curtin, *The role of PARP in DNA repair and its therapeutic exploitation*. Br J Cancer, 2011. **105**(8): p. 1114-22.
21. Weaver, A.N. and E.S. Yang, *Beyond DNA Repair: Additional Functions of PARP-1 in Cancer*. Front Oncol, 2013. **3**: p. 290.
22. Krishnakumar, R. and W.L. Kraus, *PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway*. Mol Cell, 2010. **39**(5): p. 736-49.
23. Gamble, M.J. and R.P. Fisher, *SET and PARP1 remove DEK from chromatin to permit access by the transcription machinery*. Nat Struct Mol Biol, 2007. **14**(6): p. 548-55.

24. Kim, I.K., et al., *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*. J Biol Chem, 2015. **290**(6): p. 3775-83.
25. McCann, K.E., *Advances in the use of PARP inhibitors for BRCA1/2-associated breast cancer: talazoparib*. Future Oncol, 2019. **15**(15): p. 1707-1715.
26. Kim, D.S., et al., *PARPs and ADP-ribosylation in RNA biology: from RNA expression and processing to protein translation and proteostasis*. Genes Dev, 2020. **34**(5-6): p. 302-320.
27. Lawaree, E., et al., *DNA ADP-Ribosylation Stalls Replication and Is Reversed by RecF-Mediated Homologous Recombination and Nucleotide Excision Repair*. Cell Rep, 2020. **30**(5): p. 1373-1384 e4.
28. Lyons, B., et al., *Characterization of the catalytic signature of Scabin toxin, a DNA-targeting ADP-ribosyltransferase*. Biochem J, 2018. **475**(1): p. 225-245.
29. Choi, J.R., et al., *PARP1 regulates the protein stability and proapoptotic function of HIPK2*. Cell Death Dis, 2016. **7**(10): p. e2438.
30. Liang, Y.C., et al., *PARP-2 regulates cell cycle-related genes through histone deacetylation and methylation independently of poly(ADP-ribosylation)*. Biochem Biophys Res Commun, 2013. **431**(1): p. 58-64.
31. Riccio, A.A., G. Cingolani, and J.M. Pascal, *PARP-2 domain requirements for DNA damage-dependent activation and localization to sites of DNA damage*. Nucleic Acids Res, 2016. **44**(4): p. 1691-702.
32. Szántó, M., et al., *Deletion of PARP-2 induces hepatic cholesterol accumulation and decrease in HDL levels*. Biochim Biophys Acta, 2014. **1842**(4): p. 594-602.
33. Bai, P., *Biology of Poly(ADP-Ribose) Polymerases: The Factotums of Cell Maintenance*. Mol Cell, 2015. **58**(6): p. 947-58.
34. De Vos, M., V. Schreiber, and F. Dantzer, *The diverse roles and clinical relevance of PARPs in DNA damage repair: current state of the art*. Biochem Pharmacol, 2012. **84**(2): p. 137-46.
35. Corda, D. and M. Di Girolamo, *Mono-ADP-ribosylation: a tool for modulating immune response and cell signaling*. Sci STKE, 2002. **2002**(163): p. pe53.
36. Yamamoto, M., et al., *Poly (ADP-ribose) polymerase-1 inhibition decreases proliferation through G2/M arrest in esophageal squamous cell carcinoma*. Oncol Lett, 2017. **14**(2): p. 1581-1587.
37. Barreiro, E. and J. Gea, *PARP-1 and PARP-2 activity in cancer-induced cachexia: potential therapeutic implications*. Biol Chem, 2018. **399**(2): p. 179-186.
38. Gunderson, C.C. and K.N. Moore, *Olaparib: an oral PARP-1 and PARP-2 inhibitor with promising activity in ovarian cancer*. Future Oncol, 2015. **11**(5): p. 747-57.
39. Kamboj, A., et al., *Poly(ADP-ribose) polymerase 2 contributes to neuroinflammation and neurological dysfunction in mouse experimental autoimmune encephalomyelitis*. J Neuroinflammation, 2013. **10**: p. 49.
40. Malyuchenko, N.V., et al., *PARP1 Inhibitors: antitumor drug design*. Acta Naturae, 2015. **7**(3): p. 27-37.
41. Mangerich, A. and A. Bürkle, *Pleiotropic cellular functions of PARP1 in longevity and aging: genome maintenance meets inflammation*. Oxid Med Cell Longev, 2012. **2012**: p. 321653.
42. Rossi, M.L., A.K. Ghosh, and V.A. Bohr, *Roles of Werner syndrome protein in protection of genome integrity*. DNA Repair (Amst), 2010. **9**(3): p. 331-44.
43. O'Sullivan, J., et al., *Emerging roles of eraser enzymes in the dynamic control of protein ADP-ribosylation*. Nat Commun, 2019. **10**(1): p. 1182.
44. El-Gebali, S., et al., *The Pfam protein families database in 2019*. Nucleic Acids Res, 2019. **47**(D1): p. D427-D432.
45. Xu, D., et al., *AIDA: ab initio domain assembly for automated multi-domain protein structure prediction and domain-domain interaction prediction*. Bioinformatics (Oxford, England), 2015. **31**(13): p. 2098-2105.

46. de Souza, R.F. and L. Aravind, *Identification of novel components of NAD-utilizing metabolic pathways and prediction of their biochemical functions*. Mol Biosyst, 2012. **8**(6): p. 1661-77.
47. Akturk, A., et al., *Mechanism of phosphoribosyl-ubiquitination mediated by a single Legionella effector*. Nature, 2018.
48. Kim, L., et al., *Structural and Biochemical Study of the Mono-ADP-Ribosyltransferase Domain of SdeA, a Ubiquitylating/Deubiquitylating Enzyme from Legionella pneumophila*. J Mol Biol, 2018. **430**(17): p. 2843-2856.
49. Jankevicius, G., et al., *The Toxin-Antitoxin System DarTG Catalyzes Reversible ADP-Ribosylation of DNA*. Mol Cell, 2016. **64**(6): p. 1109-1116.
50. Dudkiewicz, M.P., K, *A novel conserved family of Macro-like domains—putative new players in ADP-ribosylation signalling*. PeerJ 2019. **7**: p. e6863.
51. Saikatendu, K.S., et al., *Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1''-phosphate dephosphorylation by a conserved domain of nsP3*. Structure, 2005. **13**(11): p. 1665-75.
52. Tchasovnikarova, I.A., et al., *GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells*. Science, 2015. **348**(6242): p. 1481-1485.
53. Douse, C.H., et al., *TASOR is a pseudo-PARP that directs HUSH complex assembly and epigenetic transposon control*. bioRxiv, 2020: p. 2020.03.09.974832.
54. Pawłowski, K., et al., *Novel conserved hydrolase domain in the CLCA family of alleged calcium-activated chloride channels*. Proteins-Structure Function and Bioinformatics, 2006. **63**(3): p. 424-439.
55. Dudkiewicz, M., A. Lenart, and K. Pawłowski, *A novel predicted calcium-regulated kinase family implicated in neurological disorders*. PLoS One, 2013. **8**(6): p. e66427.
56. Dudkiewicz, M., et al., *A novel protein kinase-like domain in a selenoprotein, widespread in the tree of life*. PLoS One, 2012. **7**(2): p. e32138.
57. Sreelatha, A., et al., *Protein AMPylation by an evolutionarily conserved pseudokinase*. Cell, 2018. **175**: p. 809-821.
58. Xu, D., et al., *FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking*. Bioinformatics, 2013.
59. Zimmermann, L., et al., *A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core*. J Mol Biol, 2017.
60. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis*. Nat Protoc, 2015. **10**(6): p. 845-58.
61. Katoh, K., J. Rozewicki, and K.D. Yamada, *MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization*. Briefings in bioinformatics, 2019. **20**(4): p. 1160-1166.
62. Crooks, G.E., et al., *WebLogo: a sequence logo generator*. Genome Res, 2004. **14**(6): p. 1188-90.
63. Frickey, T. and A. Lupas, *CLANS: a Java application for visualizing protein families based on pairwise similarity*. Bioinformatics, 2004. **20**(18): p. 3702-4.
64. Huang, Y., et al., *CD-HIT Suite: a web server for clustering and comparing biological sequences*. Bioinformatics, 2010. **26**(5): p. 680-2.
65. Dereeper, A., et al., *Phylogeny.fr: robust phylogenetic analysis for the non-specialist*. Nucleic Acids Res, 2008. **36**(Web Server issue): p. W465-9.
66. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space complexity*. BMC Bioinformatics, 2004. **5**: p. 113.
67. Guindon, S., et al., *Estimating maximum likelihood phylogenies with PhyML*. Methods Mol Biol, 2009. **537**: p. 113-37.
68. Letunic, I. and P. Bork, *Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees*. Nucleic Acids Res, 2016. **44**(W1): p. W242-5.
69. Webb, B. and A. Sali, *Protein Structure Modeling with MODELLER*. Methods Mol Biol, 2017. **1654**: p. 39-54.

70. Roy, A. and Y. Zhang, *Recognizing protein-ligand binding sites by global structural alignment and local geometry refinement*. Structure, 2012. **20**(6): p. 987-97.
71. Yang, J., et al., *The I-TASSER Suite: protein structure and function prediction*. Nat Methods, 2015. **12**(1): p. 7-8.
72. Huang, C.C., et al., *Enhancing UCSF Chimera through web services*. Nucleic Acids Res, 2014. **42**(Web Server issue): p. W478-84.
73. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench*. Bioinformatics, 2009. **25**(9): p. 1189-91.
74. Livingstone, C.D. and G.J. Barton, *Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation*. Comput Appl Biosci, 1993. **9**(6): p. 745-56.
75. Offord, V., T.J. Coffey, and D. Werling, *LRRfinder: a web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database*. Dev Comp Immunol, 2010. **34**(10): p. 1035-41.
76. Mi, H., et al., *PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements*. Nucleic Acids Res, 2017. **45**(D1): p. D183-d189.
77. Ren, J., et al., *DOG 1.0: illustrator of protein domain structures*. Cell Res, 2009. **19**(2): p. 271-3.
78. Samaras, P., et al., *ProteomicsDB: a multi-omics and multi-organism resource for life science research*. Nucleic Acids Res, 2020. **48**(D1): p. D1153-d1163.
79. Thul, P.J. and C. Lindskog, *The human protein atlas: A spatial map of the human proteome*. Protein Sci, 2018. **27**(1): p. 233-244.
80. Oughtred, R., et al., *The BioGRID interaction database: 2019 update*. Nucleic Acids Res, 2019. **47**(D1): p. D529-d541.
81. Almagro Armenteros, J.J., et al., *DeepLoc: prediction of protein subcellular localization using deep learning*. Bioinformatics, 2017. **33**(21): p. 3387-3395.
82. Karlberg, T., et al., *Structural basis for lack of ADP-ribosyltransferase activity in poly(ADP-ribose) polymerase-13/zinc finger antiviral protein*. J Biol Chem, 2015. **290**(12): p. 7336-44.
83. Karlberg, T., et al., *Structural basis for lack of ADP-ribosyltransferase activity in poly(ADP-ribose) polymerase-13/zinc finger antiviral protein*. J Biol Chem, 2015. **290**(12): p. 7336-44.
84. Zhang, C., P.L. Freddolino, and Y. Zhang, *COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information*. Nucleic Acids Res, 2017. **45**(W1): p. W291-W299.
85. Burki, F., et al., *The New Tree of Eukaryotes*. Trends Ecol Evol, 2020. **35**(1): p. 43-55.
86. Zhang, L., et al., *The water lily genome and the early evolution of flowering plants*. Nature, 2020. **577**(7788): p. 79-84.
87. Kobe, B. and J. Deisenhofer, *The leucine-rich repeat: a versatile binding motif*. Trends Biochem Sci, 1994. **19**(10): p. 415-21.
88. Enkhbayar, P., et al., *Structural principles of leucine-rich repeat (LRR) proteins*. Proteins, 2004. **54**(3): p. 394-403.
89. Helft, L., et al., *LRR conservation mapping to predict functional sites within protein leucine-rich repeat domains*. PLoS One, 2011. **6**(7): p. e21614.
90. Kobe, B. and J. Deisenhofer, *Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats*. Nature, 1993. **366**(6457): p. 751-6.
91. Kobe, B. and J. Deisenhofer, *Proteins with leucine-rich repeats*. Curr Opin Struct Biol, 1995. **5**(3): p. 409-16.
92. Ng, A. and R.J. Xavier, *Leucine-rich repeat (LRR) proteins: integrators of pattern recognition and signaling in immunity*. Autophagy, 2011. **7**(9): p. 1082-4.
93. Ng, A.C., et al., *Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity*. Proc Natl Acad Sci U S A, 2011. **108** Suppl 1(Suppl 1): p. 4631-8.
94. Fehr, A.R., et al., *The impact of PARPs and ADP-ribosylation on inflammation and host-pathogen interactions*. Genes Dev, 2020. **34**(5-6): p. 341-359.

95. Rosado, M.M., et al., *Beyond DNA repair, the immunological role of PARP-1 and its siblings*. Immunology, 2013. **139**(4): p. 428-37.
96. Hoch, N.C. and L.M. Polo, *ADP-ribosylation: from molecular mechanisms to human disease*. Genet Mol Biol, 2019. **43**(1 suppl 1): p. e20190075.
97. Greenwald, S.H. and E.A. Pierce, *Parthanatos as a Cell Death Pathway Underlying Retinal Disease*. Adv Exp Med Biol, 2019. **1185**: p. 323-327.
98. Fatokun, A.A., V.L. Dawson, and T.M. Dawson, *Parthanatos: mitochondrial-linked mechanisms and therapeutic opportunities*. Br J Pharmacol, 2014. **171**(8): p. 2000-16.
99. Ke, Y., et al., *The Role of PARPs in Inflammation-and Metabolic-Related Diseases: Molecular Mechanisms and Beyond*. Cells, 2019. **8**(9).
100. Kunze, F.A. and M.O. Hottiger, *Regulating Immunity via ADP-Ribosylation: Therapeutic Implications and Beyond*. Trends Immunol, 2019. **40**(2): p. 159-173.
101. Pawlowski, K., et al., *PAAD - a new protein domain associated with apoptosis, cancer and autoimmune diseases*. Trends Biochem Sci, 2001. **26**(2): p. 85-7.
102. Martinon, F. and J. Tschopp, *Inflammatory caspases and inflammasomes: master switches of inflammation*. Cell Death Differ, 2007. **14**(1): p. 10-22.
103. Noguchi, A., et al., *ZFP36L2 is a cell cycle-regulated CCCH protein necessary for DNA lesion-induced S-phase arrest*. Biol Open, 2018. **7**(3).
104. Galloway, A., et al., *RNA-binding proteins ZFP36L1 and ZFP36L2 promote cell quiescence*. Science, 2016. **352**(6284): p. 453-9.
105. Yonemori, K., et al., *ZFP36L2 promotes cancer cell aggressiveness and is regulated by antitumor microRNA-375 in pancreatic ductal adenocarcinoma*. Cancer Sci, 2017. **108**(1): p. 124-135.
106. Murphy, J.M., P.D. Mace, and P.A. Evers, *Live and let die: insights into pseudoenzyme mechanisms from structure*. Curr Opin Struct Biol, 2017. **47**: p. 95-104.
107. Ribeiro, A.J.M., et al., *Emerging concepts in pseudoenzyme classification, evolution, and signaling*. Sci Signal, 2019. **12**(594).
108. Jeffery, C.J., *Enzymes, pseudoenzymes, and moonlighting proteins: diversity of function in protein superfamilies*. Febs j, 2020.
109. Marsischky, G.T., B.A. Wilson, and R.J. Collier, *Role of glutamic acid 988 of human poly-ADP-ribose polymerase in polymer formation. Evidence for active site similarities to the ADP-ribosylating toxins*. J Biol Chem, 1995. **270**(7): p. 3247-54.
110. Blanke, S.R., et al., *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, 1994. **33**(17): p. 5155-61.
111. Black, M., et al., *Bacterial pseudokinase catalyzes protein polyglutamylation to inhibit the SidE all-in-one ubiquitin ligases*. Science, 2019. **364**: p. 787-792.
112. Hassa, P.O., et al., *Transcriptional coactivation of nuclear factor-kappaB-dependent gene expression by p300 is regulated by poly(ADP)-ribose polymerase-1*. J Biol Chem, 2003. **278**(46): p. 45145-53.