A novel conserved family of Macro-like domains—putative new players in ADP-ribosylation signaling

Małgorzata Dudkiewicz[1] and Krzysztof Pawłowski[1,2]

[1] Department of Experimental Design and Bioinformatics, Faculty of Agriculture and Biology, Warsaw University of Life Sciences, Warszawa, Poland
[2] Department of Translational Mecicine, Clinical Sciences, Lund University, Lund, Sweden

ABSTRACT

The presence of many completely uncharacterized proteins, even in well-studied organisms such as humans, seriously hampers a full understanding of the functioning of living cells. One such example is the human protein C12ORF4, which belongs to the DUF2362 family, present in many eukaryotic lineages and conserved in metazoans. The only functional information available on C12ORF4 (Chromosome 12 Open Reading Frame 4) is its involvement in mast cell degranulation and its being a genetic cause of autosomal intellectual disability.

Bioinformatics analysis of the DUF2362 family provides strong evidence that it is a novel member of the Macro clan/superfamily. Sequence similarity analysis versus other representatives of the Macro superfamily of ADP-ribose-binding proteins and mapping sequence conservation on predicted three-dimensional structure provides hypotheses regarding the molecular function for members of the DUF2362 family. For example, the available functional data suggest a possible role for C12ORF4 in ADP-ribosylation signaling in asthma and related inflammatory diseases.

This novel family appears to be a likely novel ADP-ribosylation "reader" and "eraser," a previously unnoticed putative new player in cell signaling by this emerging post-translational modification.

INTRODUCTION

In 2019, there are still 3918 protein families and three protein clans in the Pfam database (http://pfam.xfam.org) that are described simply as "DUFs"–domains of unknown function. These contain together over 1.3 million amino acid sequences, and as many as 512 (13%) Pfam "DUF" families contain human protein sequences. Proteins containing such domains are usually disregarded by biologists (*Pawlowski, 2008*) although often they turn out to be important signaling enzymes (e.g. *Dudkiewicz et al., 2012*; *Sreelatha et al., 2018*).

Recent findings of two independent groups from Saudi Arabia and Europe analyzing multiplex consanguineous families with members suffering from intellectual disability (ID) using whole-exome sequencing implicate the mysterious C12ORF4 as a causative

gene for autosomal ID, an important health problem in society (*Alazami et al., 2015*; *Philips et al., 2017*). Genetic causes of ID still remain poorly explained because of its vast heterogeneity (*Anazi et al., 2017*). The involvement of C12ORF4 in ID was confirmed by observation of the presence of specific mutations in this protein-coding gene in all ID affected members of the analyzed families, including a missense variant (L328P), a deletion (Q267fs) and a frameshift. Another study shows that, in rats, the gene orthologous to C12ORF4 encodes a cytoplasmic protein, involved in mast cell degranulation (*Mazuc et al., 2014*).

In this study, we present evidence that C12ORF4 is likely involved in ADP-ribosylation of proteins, an important post-translational modification that plays a key role in signal transduction in many important cellular processes such as DNA repair, regulation of transcription, maintenance of chromatin stability, cell differentiation and proliferation, necrosis and apoptosis (*Han, Li & Fu, 2011*; *Palazzo, Mikoč & Ahel, 2017*; *Cohen & Chang, 2018*; *Lüscher et al., 2018*). NAD+ is an essential cofactor and donor of the ADP-ribose group in a reaction which starts with breaking the bond between nicotinamide and ribose. As a result of the ADP-ribosylation reaction, nicotinamide and ADP-ribosylated proteins are released. Errors in the ADP-ribosylation of proteins can be associated with cancerogenesis processes (*Barkauskaite et al., 2013*). ADP-ribosylation of DNA is relatively less common and until recently has only been described for a small number of bacterial toxins (*Jankevicius et al., 2016*). However, there are new reports showing DNA ADP-ribosyltransferase activities of some mammalian PARPs (*Munnur & Ahel, 2017*; *Zarkovic et al., 2018*).

One of the important players in ADP-ribosylation, the Macro domain, is defined as a module of about 180 amino acids which can bind the important NAD metabolite–ADP-ribose–or related ligands (*Rack, Perina & Ahel, 2016*; *Rosenthal et al., 2013*). The Macro domain was originally named A1pp, as it was described in association with ADP-ribose 1-phosphate processing activity in yeast (*Martzen et al., 1999*), but later it was renamed Macro in reference to the fact that it makes up the C-terminal domain of mammalian core histone macro-H2A (*Allen et al., 2003*; *Karras et al., 2005*). Macro domain proteins are widespread in eukaryotes, but also present in pathogenic bacteria, archaea and in ssRNA viruses, such as coronaviruses, Rubella and Hepatitis E viruses (*Feijs et al., 2013*; *Rack, Perina & Ahel, 2016*).

The Macro domain is a universal "tool" for recognition of the ADP-ribose nucleotide and its polymers found in a number of otherwise unrelated proteins, which belongs to a cellular toolkit used for ADP-ribosylation of the proteome and its reversion. The typical three-dimensional structure of the Macro domain can be described as a mixed alpha/beta/alpha fold. A six-strand beta sheet is sandwiched between five α-helices with the ligand binding pocket lying between two loops (loop 1 and loop 2) (*Slade et al., 2011*; *Rack, Perina & Ahel, 2016*). Some Macro domains are shorter and lack either the first strand or the C-terminal helix 5. Well-conserved residues build a hydrophobic cleft and cluster around the ADP-ribose binding site. Most important for ligand binding by Macro domains are conserved aromatic residues in binding pocket (π–π stacking interactions with the adenosine ring) and the aspartate residue coordinating the N6 atom of adenosine.

The pyrophosphate and distal ribose are stabilized by two binding loops, the first of which can contain catalytic residues in cases when the Macro domain exhibits hydrolase activity.

Phylogenetic analysis of the Macro clan/superfamily distinguishes six classes of Macro domains: MacroD-type and MacroH2A-like (Macro family in the Pfam database, PF01661), Macro2-type (Macro_2, PF14519), ALC1-like, PARG-like (PARG_cat, PF05028) and SUD-M like (SUD-M, PF11633) (*Rack, Perina & Ahel, 2016*). The Pfam database additionally lists the M17 aminopeptidase N-terminal domain (PF18295 and PF02789) and DUF2362–an uncharacterized domain conserved in bacteria (PF10021)–as members of the Macro clan.

Depending on their functions, Macro domains can be divided into two distinct groups–"readers" and "erasers"–where the latter completes the "writer-reader-eraser" ADP-ribosylation triad, whereas the "writers" are ADP-ribosyltransferases. In humans, both "readers" and "erasers" are found in the Macro family (PF01661) (*Chen et al., 2011*; *Barkauskaite et al., 2013*; *Jankevicius et al., 2013*), while eraser-type PARG-like domains belong to a separate PARG-cat family (PF05028, poly-ADP-ribosyl-hydrolases) (*Slade et al., 2011*; *Dunstan et al., 2012*). Reader domains are found in PARPs (multidomain enzymes which catalyze mono-and poly-ADP-ribosylation of proteins) (*Gibson & Kraus, 2012*; *Langelier et al., 2018*), in histones MacroH2A (1 and 2) (*Posavec, Timinszky & Buschbeck, 2013*), in ALC1 (*Lehmann et al., 2017*) and GDAP2 (*Neuvonen & Ahola, 2009*) proteins and have no catalytic function, they accompany other domains and serve for recognition of ADP-ribose moieties.

Erasers have ADP-ribosyl-hydrolase activity and are found in mono-(MacroD1/D2, TARG) and poly-ADP-ribosyl-hydrolases (PARG proteins).

The mechanisms of ligand binding in both readers and erasers are similar, although reader Macro domains coordinate ADPr in a rather relaxed conformation while some erasers (e.g., MacroD2) keep ligands in strained form due to the presence of a structural water molecule in the binding cleft. However, the main difference between these two classes lies in the presence of catalytic residues in loops 1 or 2 of the "erasers." The structure and reaction mechanisms of the human hydrolytic Macro domains represent three different types.

Type-1: MacroD1/2 (*Chen et al., 2011*; *Barkauskaite et al., 2013*; *Jankevicius et al., 2013*): (O-acetyl-ADP-ribose deacetylase)-ADPr is bound in a constrained conformation forced by the presence of a structural water molecule and highly conserved aromatic residues located in MacroD-type signature motifs Nx(6)GG[V/L/I]D and G[V/I/A][Y/F]G identified in loops 1 and 2 (Fig. 1), respectively. Catalytic residues for hydrolysis reaction, Asn and Asp, are anchored in loop 1 and interact electrostatically with the distal ribose.

Type-2: TARG1 (*Sharifi et al., 2013*): (Terminal ADP-ribose protein Glycohydrolase 1)-ADPr binding cleft more closely resembles the "reader" ligand binding site and stabilizes the substrate in a relaxed conformation. Catalytic residues are lysine and aspartate, located in loops 1 and 2, respectively.

Type-3: PARG (*Slade et al., 2011*; *Dunstan et al., 2012*): (Poly(ADP-Ribosyl) Glycohydrolase) has a three-domain structure: a Macro domain together with a TARG accessory domain form the canonical catalytic core, the third domain is a putative regulatory
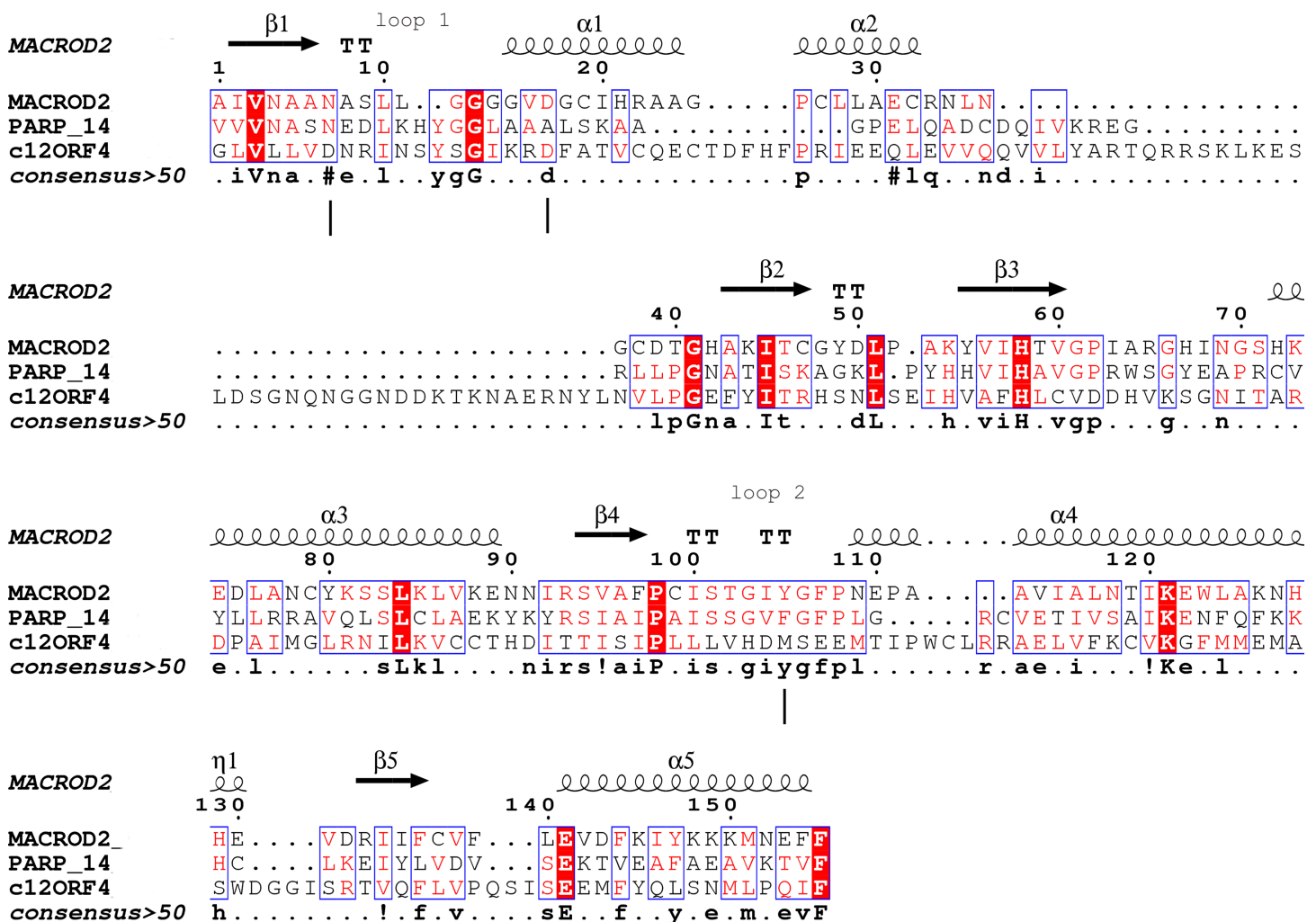
Figure 1 Sequence alignment for human amino acid sequences of C12ORF4, PARP14 and MacroD. Alignment constructed using Phyre[2] and HHpred algorithms. Secondary structure elements derived from human MacroD1 PDB structure are shown above, vertical lines indicate positions of catalytic and ligand binding residues Asn92, Asp102 and Tyr190 in human MacroD2 sequence and their possible counterparts in C12ORF4.

Full-size ⌸ DOI: 10.7717/peerj.6863/fig-1

region. The PARG Macro domain contains a specific catalytic motif within loop 1, and the main catalytic residues are glutamate in loop 1 and phenylalanine in loop 2. PARG is predominantly an exoglycohydrolase with only minor contributions of endocleavage activity.

Macro domains are a very divergent and evolutionarily old protein fold, their common feature is affinity to ADPr. The same function can be realized using different mechanisms and binding modes, but overall the Macro domain fold still remains recognizable.

In this paper, we present Macro domain-like structural predictions for the uncharacterized C12ORF4/DUF2362 family. We also discuss the relevance of the structural predictions for the likely molecular function in the DUF2362 family. Furthermore, we analyze the phylogenetic spread of the C12ORF4 genes and their relationship to members of the Macro superfamily/clan. We also summarize and analyze structural features of possible

ligand binding site and catalytic loop in the newly identified Macro domain using models constructed for its human C12ORF4 representative. Nevertheless, we stress that this in silico prediction report should be the starting point for experimental validation of the structural and functional hypotheses.

MATERIALS AND METHODS

Initial analysis of distant sequence similarities of human C12ORF4 and the DUF2362 family was performed using three state-of-the-art servers for protein structure and function prediction: Phyre2 (Protein Homology/analogY Recognition Engine V 2.0) (*Kelley et al., 2015*), FFAS03 (Fold and Function Assignment System) (*Jaroszewski et al., 2011*), and HHpred (Homology detection & structure Prediction by HMM–HMM comparison) (*Soding, Biegert & Lupas, 2005*), that implement Hidden Markov Models, profile–profile alignment methods, and related approaches for the detection of remote sequence similarity.

To assess phylogenetic spread of the DUF2362 family, the Pfam Representative Proteome rp75 sequence collection was chosen (*Chen et al., 2011*). Sequences were next classified and described using UniProt database Taxonomy tools, together with GeneTree (*Vilella et al., 2009*) and TreeFam (*Ruan et al., 2008*) gene orientated phylogenetic databases, providing curated phylogenetic trees for animal gene families. Thus, 381 sequences from 261 species were assigned to taxonomic groups and visualized using the iToL server on a dendrogram representing the Tree of Life (*Letunic & Bork, 2016*).

To assess similarities in the amino acid sequences of human C12ORF4 and PARP14/MacroD2 sequences, a multiple sequence alignment was constructed using the HHpred and Phyre2 algorithms, and manually adjusted. Secondary structure assignment from the PDB structure of MacroD2 (4IQY) was added to the final alignment. To graphically represent the resulting alignment, ESPript 3.0 software was used (*Robert & Gouet, 2014*).

A sequence logo for the rp75 set of DUF2362 sequences was created using the Pfam alignment. The alignment was edited to adjust it to the selected "master sequence" by removing columns containing gaps in human C12ORF4. Graphical representations of the sequence logo were prepared using the WebLogo3 server (*Schneider & Stephens, 1990*).

The structures of human MacroD2 (PDB code: 4IQY) and PARP14 (PDB code: 3Q6Z) were used to depict similarities between these proteins and human C12ORF4 according to BLOSUM62 matrix values (from 0 to 11) in pairwise alignments with human C12ORF4. High positive BLOSUM62 values signify residues which are identical or highly similar and frequently conserved by evolution.

In order to visualize sequence variation in the DUF2362 family, conservation of sequence positions was mapped onto known Macro domain structures as well as on the structure model of human C12ORF4. Conservation values derived from the Jalview alignment editor (*Waterhouse et al., 2009*) were automatically calculated as quantitative alignment annotation measuring the number of conserved physicochemical properties for each column of the alignment (*Livingstone & Barton, 1993*).

A protein three-dimensional structure model for human C12ORF4 was built using the homology modeling method in Modeller9v21 (*Webb & Sali, 2014*) based on a human

MacroD2 structure template (PDB ID: 4IQY). The loop (374–408) that had no counterpart in the template was refined using the I-Tasser (Iterative Threading ASSEmbly Refinement) server for protein structure and function prediction (*Yang et al., 2015*). The putative ligand binding site in the modeled structure was identified using the COFACTOR algorithm (*Zhang, Freddolino & Zhang, 2017*), which, based on threading of the query structural model through the BioLiP protein function database, annotates proteins and identifies potential functional sites by local and global structure matches. Structures were visualized and analyzed using UCSF Chimera tools (*Pettersen et al., 2004*).

For survey of similarities between the Macro domain clan and C12ORF4 homologs, the CLANS algorithm (*Frickey & Lupas, 2004*) was run. The set of sequences selected for analysis included all the Pfam rp15 sequences from the six families of the Macro clan (CL0223)–Macro (PF10021), Macro2 (PF14519), PARG-cat (PF05028), PeptidaseM17_N, (PF02789) SUD-M (PF11633), DUF2263 a.k.a. bacterial-type PARG (PF10021) (*Slade et al., 2011*)–and the novel DUF2362 family (PF10154). ALC1-like sequences were added manually, because they are not represented in Pfam. An ALC1 homolog collection was created based on BLAST search results for the human CHD1L (ALC1) Macro domain. Resulting sequences were clustered using CDHIT with sequence identity cutoff 0.9 and thus 3,432 representative sequences were obtained. For the CLANS run, relationships from a single iteration of BLAST were used, up to the *E*-value of 1, thus including also sub-significant similarities. The BLOSUM62 substitution matrix was used in BLAST. To find more potential members of the DUF2362 family, the Jackhmmer algorithm was applied (*Finn et al., 2015*). The interactors of C12ORF4 were elucidated using the STRING database (*Szklarczyk et al., 2015*) and the Ingenuity system (Qiagen, Hilden, Germany).

RESULTS AND DISCUSSION

Assignment of DUF2362 proteins to the Macro clan

Analysis of distant similarities between human proteins and ADP-ribosylation-related families performed using the FFAS server suggested that the uncharacterized family DUF2362 might be similar to Macro domains. Table 1 shows that both FFAS and two other popular structure prediction servers predicted a weak but suggestive similarity to Macro for the human C12ORF4 protein (see also Fig. 1). These methods use sequence profiles or Hidden Markov Models to detect distant sequence similarities compatible with similar structures. Phyre[2] and HHpred are remarkably concurring in their prediction of structural similarities for C12ORF4, identifying as best-scoring structure modeling templates the Macro domains of the human poly(ADP-ribose) polymerase 14 (PDB: 3Q6Z) and human protein-proximal ADP-ribosyl-hydrolase MacroD2 (PDB: 4IQY) or human MacroD1 (PDB: 2X47). Using the servers' results, it was difficult to distinguish between the predictions, as confidence values were very close. However, FFAS03 ran on DUF2362 multiple sequence alignment used as the query gave the highest score for human PARP14 (PDB: 3Q6Z) as structural prediction, but did not pick human MacroD1/2 structures as significant hits. Using HHpred and Phyre[2] local HMM

Table 1 Structure predictions for human C12ORF4 protein.

Template PDB id	Alignment coverage (% of C12ORF4 sequence)	Phyre2 confidence	FFAS03 score	HHpred confidence	Sequence identity %	HHpred E-value	Template info
3q6zA	39% (327–540)	98.1	−11.8	96.6	15	0.0088	Poly(ADP-ribose) polymerase 14
4iqyB	39% (327–540)	97.8	–	97.14	16	0.0011	O-acetyl-ADP-ribose deacetylase macrod2
2x47A	24% (410–540)	97.7	–	96.64	14	0.0047	Macro domain-containing protein 1

Note:
Phyre2, FFAS03 and HHpred results for human C12ORF4. Scores of the best PDB hits provided. Prediction confidence as %.

profile–profile alignment, an alignment for human C12ORF4 and the two best hits was constructed, considering secondary structure assignments from the human MacroD2 structure (Fig. 1). Manual adjustment was performed only in the loop 1 region. Investigating similarities between C12ORF4 and important regions of MacroD2, a similarity was noticed to the characteristic MacroD motif Nx(6)GG[V/L/I]D in the C12ORF4 region aligned with loop 1. However, there is a difference; whereas the first residue in the glycine dyad is substituted by serine: Nx(6)SG[V/L/I]D. MacroD2 residues N92 (matched with D333 in human C12ORF4) and D102 (aligned to D345 in human C12ORF4) are catalytic amino acids electrostatically interacting with the distal ribose ring of ADPr ligands. Conservation of these residues in C12ORF4 seems to support its similarity to the catalytic Macro domain class, albeit tyrosine 190, crucial for eraser function of MacroD, is substituted by a methionine (M478). The structural water molecule and aromatic ring of Y190 located in eraser loop 2 bind the ADPr moiety in constrained position enabling initiation of its hydrolysis. The reader-type Macro domain in PARP14 has conserved Asp residues in loop 1, but tyrosine within the G[IV]YG motif is replaced by a phenylalanine, also possessing an aromatic ring. In C12ORF4, there is a matched methionine at this position. This brings uncertainty regarding the catalytic activity of the C12ORF4 protein, taking into account the fact that Y->F substitution in this motif can convert a catalytically active Macro domain to a reader-only pseudoenzyme (*Rack, Perina & Ahel, 2016*).

Phylogenetic spread of the DUF2362 family and relationship to the Macro clan

The DUF2362 family in the Pfam database contains sequences belonging to 261 species (see Fig. 2), mainly Metazoa, including the early Metazoan *Trichoplax adherens* (Placozoa), and also the closest unicellular relative of metazoans, *Capsaspora owczarzaki* (Ichthyosporea). DUF2362 is also found in some fungi (albeit the fungal C12ORF4 homologs belong to specialized atypical groups of fungi, e.g., Mucoromycota, Chytridiomycota, Zoopagomycota and Cryptomycota), Amoebozoa, Apicomplexa, Choanoflagellida and Cryptophyta. C12ORF4 homologs have not been found in most fungi, or in plants, archaea or bacteria. Overall, the Jackhmmer search for new potential family members also picked up 410 C12ORF4 homologs in the Representative Proteomes and about 700 C12ORF4 homologs in UniProt. The broad phylogenetic spread of this protein family confirms its ancient origin
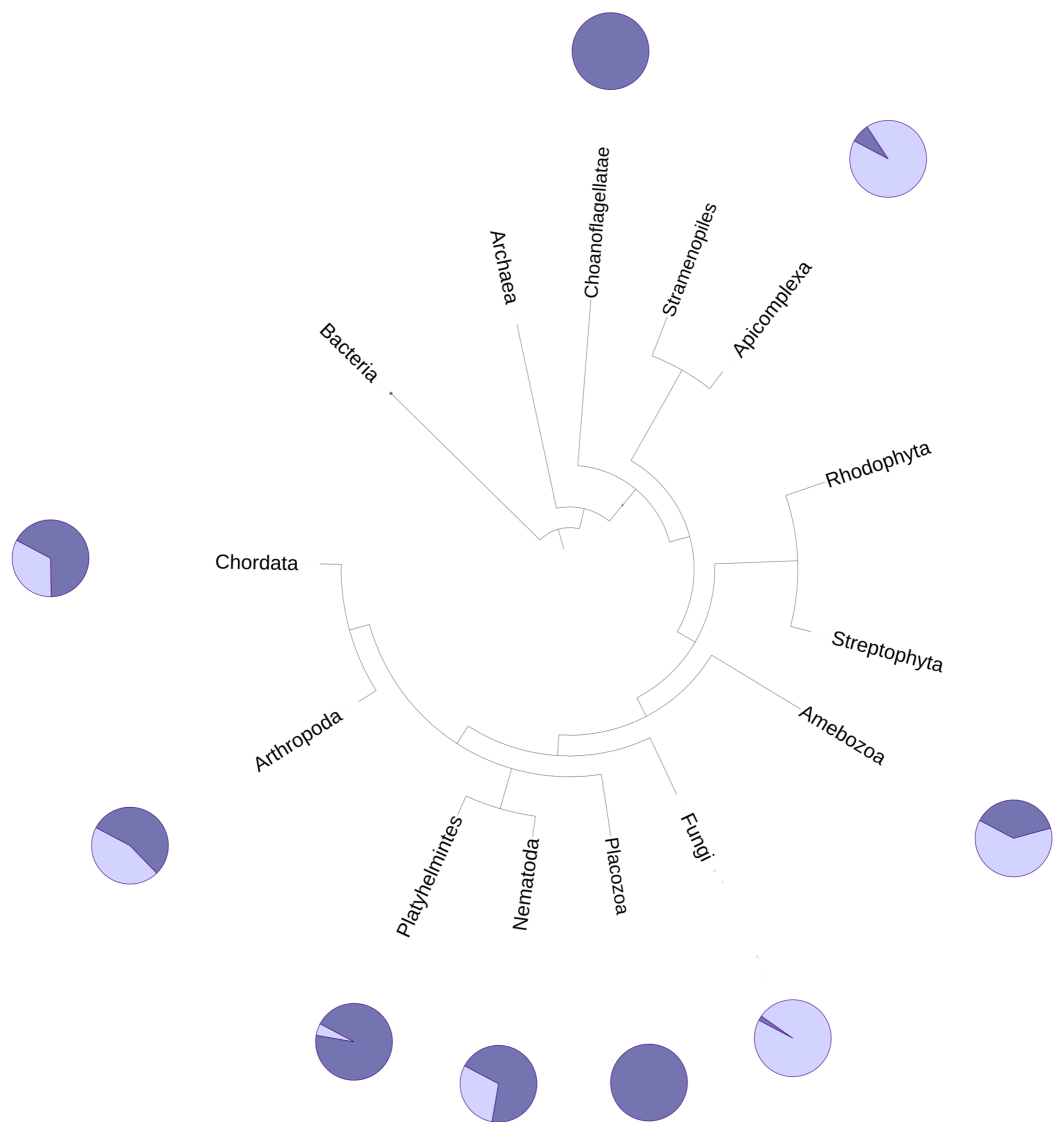
Figure 2 Phylogenetic spread of DUF2362 family. Selected branches of the universal Tree of Life. The dark blue sectors within the circles represent the percentage of representative genomes in the given taxonomic group possessing C12ORF4 homologs, as identified by Jackhmmer.

Full-size ◩ DOI: 10.7717/peerj.6863/fig-2

and potentially important cellular function. It is suggestive of a DUF2362 member being present in the last eukaryotic common ancestor and subsequent DUF2362 gene losses in some major eukaryotic lineages.

The relationship of the DUF2362 family to the Macro clan can be visualized using a graph-based approach: the CLANS algorithm (*Frickey & Lupas, 2004*). The CLANS graph depicts proteins as nodes connected by edges representing BLAST-detected significant or sub-significant sequence similarities. The nodes are clustered by "attractive forces" representing BLAST *E*-values (see Fig. 3). DUF2362 appears to be connected not only to the main family of the clan–Macro (PF01661), but also Macro2, Peptidase_M17, PARG-cat and DUF2263. The biggest grouping of the superfamily, the Macro family,
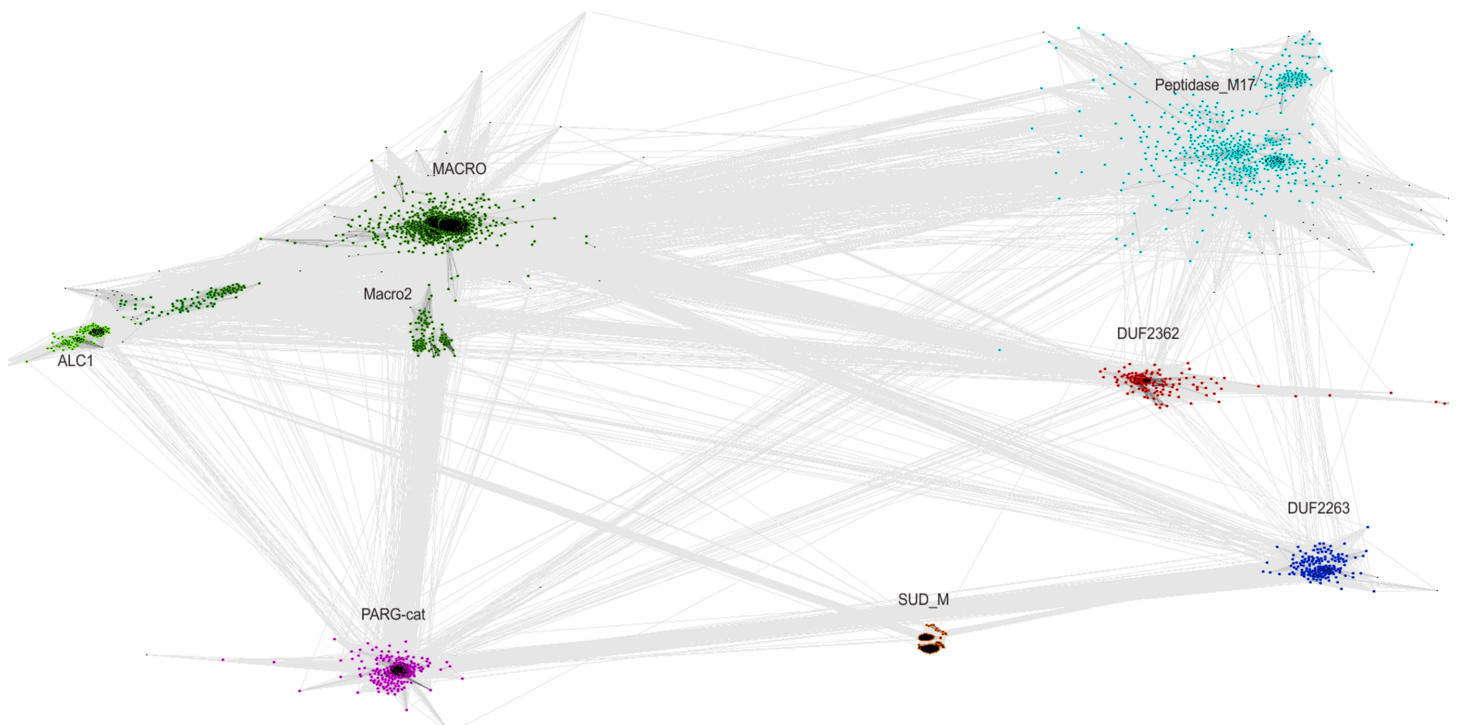
Figure 3 CLANS graph visualizing BLAST-detected sequence similarities among proteins of the Macro domain superfamily. CLANS clustering results for 3,232 representatives of the Macro domain superfamily (Pfam clan CL0223 plus ALC1 homologs and the DUF2362 family).

contains the majority of known human Macro domains, eraser and reader-like, except one that belongs to the PARG family, poly(ADP-ribose)-glycohydrolases.

Sequence variability in the DUF2362 family

While distant sequence similarity search tools strongly suggested the presence of a Macro-like domain in the C-terminal region of C12ORF4 (approx. residues 322–552), the presence of another domain in N-terminal part of the protein could not be established with confidence. HHpred reported a non-significant similarity to human all-helical nuclear pore complex protein Nup155, while FFAS03 reported a weak similarity to bacterial colicin E9.

Investigation of likely domain structure in uncharacterized proteins can be aided by the analysis of their amino acid sequence logo, a graphical representation of conservation of sequence positions in the protein family. A logo created using multiple alignment of the DUF2362 family shows strong sequence conservation in the C-terminal region, but also some strongly conserved motifs between positions 85 and 110 and in the region 230–260 (see Fig. 4 and Fig. S1 for full-length DUF2362 logo). Two Asp residues (D333 and D345) identified in profile–profile alignments (Fig. 1) as likely counterparts of MacroD2 catalytic residues are well-conserved in the family, although consensus does not reach 100% (Fig. 5). The long region following the putative catalytic motifs is the least conserved part of the DUF2362 alignment (Fig. 3). This region–378–408 in human

Figure 4 Sequence logo for Macro-like domain of the DUF2362 protein family. Multiple sequence alignment from the Pfam database (representative proteomes rp75 set), the logo was created for regions corresponding to positions 323–545 in human C12ORF4 amino acid sequences with Macro domain signal identified using remote protein homology detection tools.

Full-size ◩ DOI: 10.7717/peerj.6863/fig-4

C12ORF4–has no counterparts in known Macro domain structures, it is also very poorly conserved in the DUF2362 family (Fig. S1). Figure 5 presents a comparison of conservation within two catalytic loops in two Macro domain families–eraser-type (MacroD2) and reader-type (PARP14)–and in the corresponding regions of DUF2362. In C12ORF4, two conserved Asp residues, D333 and D345, are putative counterparts of MacroD2 active site N92 and D102. While similarities within loop 1 in C12ORF4 and two known Macro families are quite noticeable, the loop 2 motifs are more different. The presence of highly conserved methionine and histidine is noteworthy. These residues may be involved in binding/positioning of the distal ribose of ADPr, similarly to Y190 of MacroD2 and F926 of PARP14.

Apart from the predicted ADPr binding and catalytic site, the functions of conserved residues are less obvious. Several strikingly well-conserved cysteine residues could serve the building of disulfide bridges. The very well-conserved histidine 431 corresponds
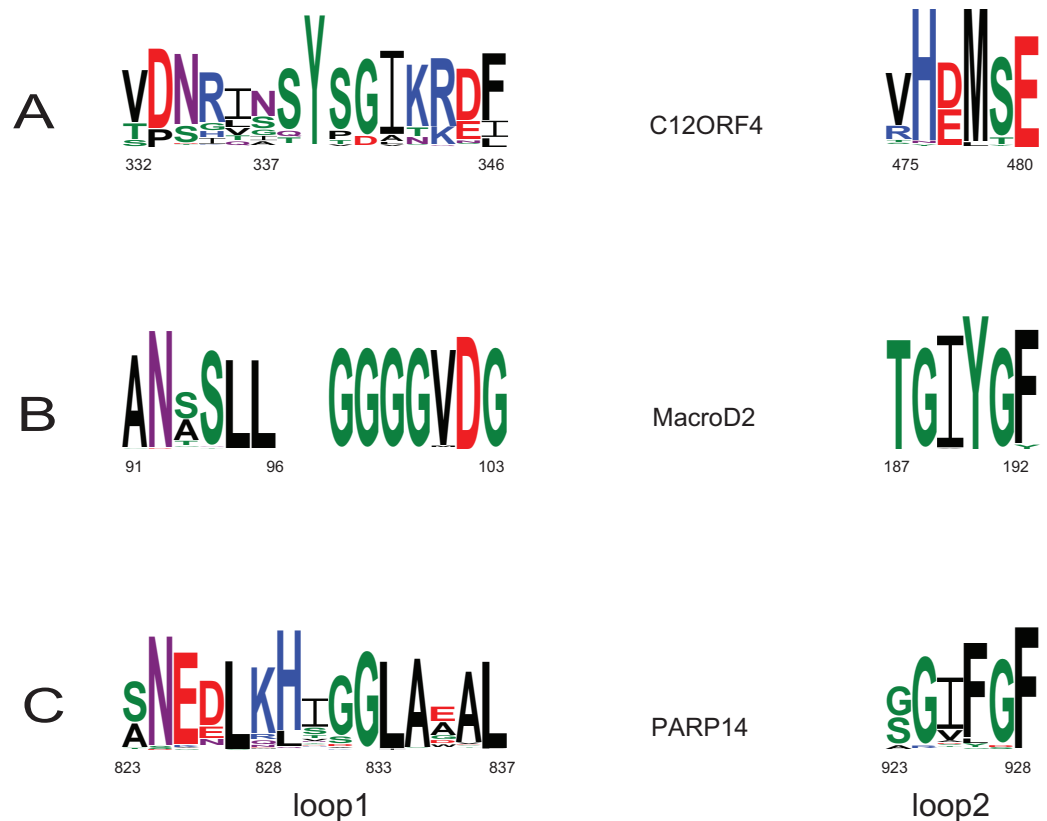
Figure 5 Sequence logos for active sites in loop 1 and loop 2 in two Macro families: MacroD2 and PARP14 compared to DUF2362. Conservation of catalytic residues in multiple sequence alignment constructed for MacroD2 and PARP14 homologs and corresponding conserved residues in aligned fragments of DUF2362. Amino acid position numbering according to human representatives for all three sequence groups. Full-size 🖼 DOI: 10.7717/peerj.6863/fig-5

to a histidine conserved in a VIH[TA] motif in Macro domains of PARP14 and MacroD2 and may be involved in the positioning of the active site Asp. The most conserved region (475–520) of the DUF2362/C12ORF4 family corresponds to helices 4 and 5 in known Macro domains which are distal to the ADP-ribose site and whose functions are not clear.

Profile–profile alignment methods offer automatic pipelines enabling modeling of the putative three-dimensional structure of the query protein. In the case of C12ORF4, automated methods would be quite risky, mainly due to the presence of long insertions (likely loops) within the hypothetical Macro domain. Residues from 374 to 408 had to be modeled without a structural template. Secondary structure prediction run for amino acids in this region of C12ORF4 indicate a mainly unstructured character of this fragment, so we decided to build a preliminary model for the C12ORF4 structure in the form of a homology model based on a human MacroD2 template, using our own manually adjusted alignment, with the building and optimization of missing loops using the I-TASSER server. An "eraser" template (PDB:4IQY) was chosen because of the fact that only one Macro-like domain was detected in our query. This is similar to the "eraser" MacroD1/2 proteins that have only one catalytically active Macro domain. In contrast,
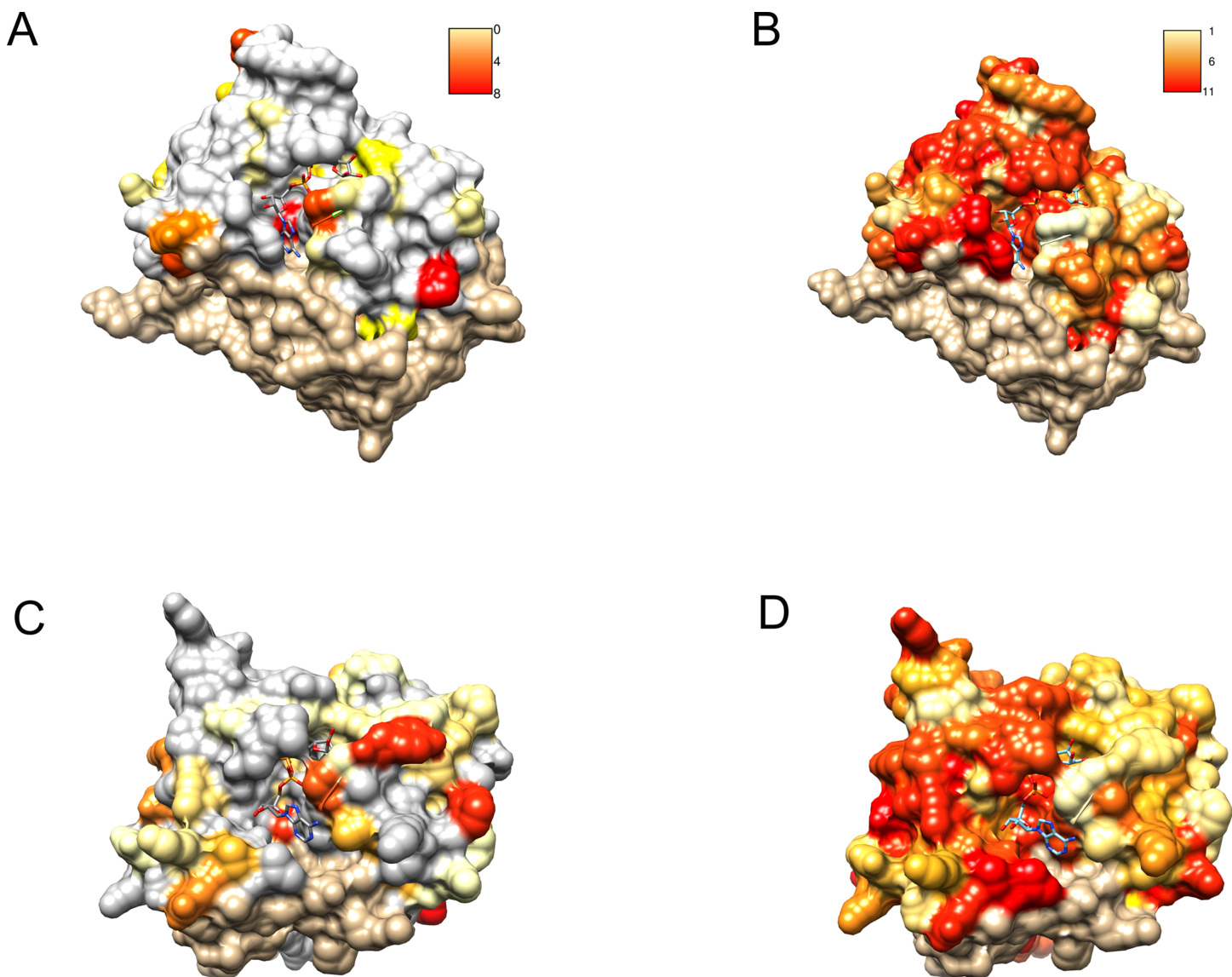
Figure 6 Similarity of DUF2362 to known Macros and sequence variation within the family mapped on three-dimensional structures. Bound ADPr molecules shown. (A) Conservation in human MacroD2/C12ORF4 pairwise sequence alignment mapped onto MacroD2 structure (PDB 4IQY). Coloring of the molecular surface according to BLOSUM62 values for alignment positions (red–high similarity, yellow–low similarity, white–no similarity (BLOSUM62 values below zero), beige–unaligned regions). (B) Conservation in the DUF2362 family sequence alignment mapped onto MacroD2 structure (PDB 4IQY). Coloring of the molecular surface: Jalview-derived alignment conservation value (red–high conservation, yellow–low conservation, white–no similarity, beige–unaligned regions. (C) Conservation in human PARP14/C12ORF4 pairwise sequence alignment mapped onto PARP14 structure (PDB 3Q6Z). Coloring as in (A). (D) Conservation in the DUF2362 family sequence alignment mapped onto PARP14 structure (PDB 3Q6Z). Coloring as in (B).

Full-size 🖼 DOI: 10.7717/peerj.6863/fig-6

the "reader" PARP14 contains three repeated Macro domains. Also, there are two likely catalytic Asp residues present in the predicted loop 1 of C12ORF4. Before building the structure model, we analyzed possible templates and conservation of important residues in C12ORF4 and the whole DUF2362 family, using sequence alignments and logo. As can be noticed in Fig. 6, conservation of DUF2362 alignment positions matched to the ligand binding cleft of MacroD2 and PARP14 is high (Figs. 6B and 6D). Also, considering
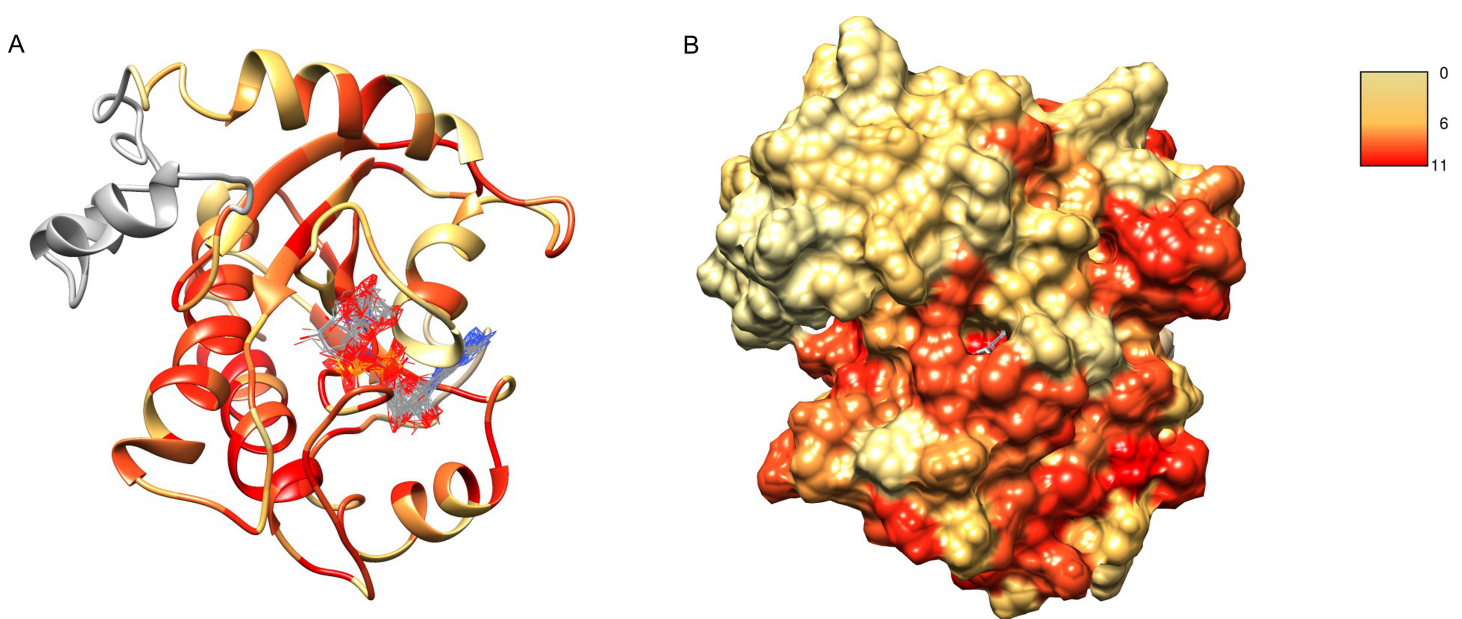
A

B

Figure 7 Structure model of C12ORF4 with COFACTOR positions of high scoring potential ligands–ADP-ribose molecules. Structure coloring: Jalview-derived sequence alignment conservation value for the DUF2362 family (red–high conservation, yellow–low conservation. (A) Ribbon representation (the region 378–408 modeled by I-Tasser server without template depicted in gray). (B) Molecular surface map.

Full-size 🖼 DOI: 10.7717/peerj.6863/fig-7

pairwise alignment mapping (Figs. 6A and 6C), one can see that apart from conserved (Asp/Asp) or almost conserved (Asn/Asp) catalytic positions there are also some residues conserved in C12ORF4 in the neighborhood of the putative ADPr binding site–proline, alanine and glycine in the case of both templates and tyrosine in the case of PARP14.

After obtaining the structure model of human C12ORF4, with the ab initio constructed fragment, we used the COFACTOR algorithm to identify and place putative ligands within the obtained model. COFACTOR scans the PDB database for structures similar to the query and containing co-crystallized ligands to propose possible ligand binding site locations within the analyzed model. This represents an approximate result and one should not treat the obtained ligand binding poses as certain, but one can rank proposed ligands and identify clefts likely responsible for its binding. Results shown in Table S1 support ADPr binding by C12ORF4. The highest scoring COFACTOR ligands are ADP-riboses, which are frequently co-crystallized with Macro domain structures, which is not surprising considering the template used for homology modeling. The predicted ADPr binding cleft is localized between loop 3 (counterpart of MacroD loop 2) and helix 5 and flanked by the loop modeled by I-TASSER (see Figs. 7 and 8). Hydrogen bonds to ADPr detected in the model using UCSF Chimera tools (Pettersen et al., 2004) involve Ile 524, and His, Asp and Met of HDM (476–478) motif aligned to MacroD2 catalytic tyrosine. However, a detailed investigation of ligand binding based on the preliminary model and COFACTOR results would be too speculative.

Structure modeling for uncharacterized proteins is always a risky enterprise when no highly homologous template is available. In the case of C12ORF4, an additional
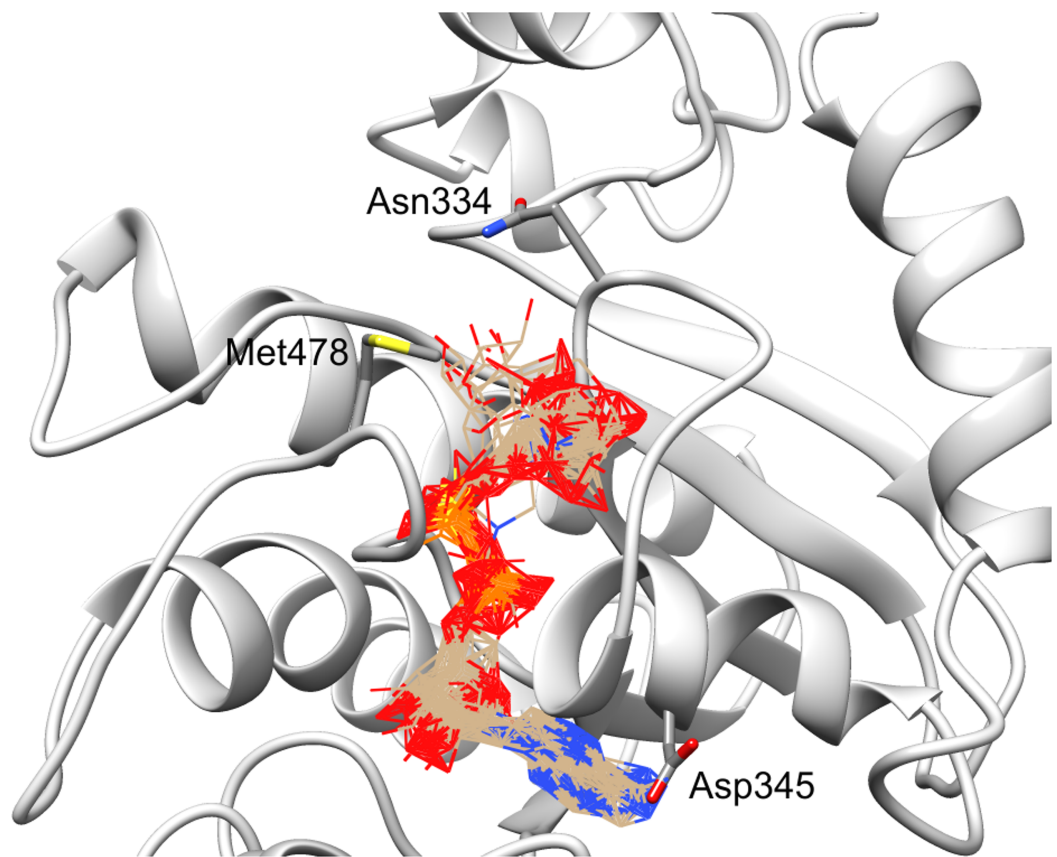
Figure 8 Ligand binding site in the structure model of C12ORF4 with hypothetical positions of predicted ligands–ADP-ribose molecules. Ribbon representation with Asn 334, Asp 345 in loop 1 and Met 478 in loop 2 (identified as counterparts of MacroD2 catalytic residues) depicted in the all-atom representation.   Full-size ⬚ DOI: 10.7717/peerj.6863/fig-8

complication is the presence of a putative N-terminal domain and a long insertion within the Macro-like region.

Biological function of C12ORF4

The only report available that investigated the molecular function of C12ORF4 is that of *Mazuc et al. (2014)* who used an intrabody approach to identify proteins involved in FcεRI-induced mast cell degranulation. They showed that silencing C12ORF4 influenced FcεRI-mediated signaling events; namely, it caused a decrease in FcεRI-mediated tyrosine phosphorylation of Src and Syk tyrosine kinases as well as the downstream kinases, MAPKs and Akt. This, in turn, had a clear effect on degranulation, by decreasing the release of TNF-α and β-hexaminidase.

The Mazuc study, together with our findings that link C12ORF4 to ADP-ribosylation, casts interesting light on the known function of PARP-1 in degranulation, and specifically in modulating asthma-associated production of cytokines (*Sethi, Dharwal & Naura, 2017*). Studies on asthma patient material as well as animal models have led to inhibitors of PARP1 and PARP14 being considered for asthma treatment and being investigated in animal models (*Zaffini, Gotte & Menegazzi, 2018*). For example, in an early study in guinea

pig, PARP inhibition prevented an induced asthma-like reaction (*Suzuki et al., 2004*). Allergen exposition induced a rapid increase in PARP activity, bronchiolar constriction, pulmonary air space inflation and mast cell degranulation in sensitized animals, while inhibition of PARP prevented the above morphological and biochemical changes to lung tissue (*Suzuki et al., 2004*). Our results allow us to propose a hypothesis that Macro-like C12ORF4, as a putative ADPr "eraser," regulates PARP-mediated ADP-ribosylation signaling, similarly to PARG, whose role in countering PARP action in airway disease is known (*Sethi, Dharwal & Naura, 2017*). The fact that C12ORF4 expression has been found to affect phosphorylation of several kinases (*Mazuc et al., 2014*) allows speculation that the effects of PARP and C12ORF4 on degranulation may be mediated by direct ADP-ribosylation of kinases. There are known cases of regulation of kinase signaling pathways by direct ADP-ribosylation of key kinases, by effectors from infectious bacteria (*Young et al., 2014*) and also by endogenous human ADP-ribosyltransferases (*Song et al., 2016*; *Li et al., 2018*).

An interesting connection between the degranulation-related phenotype attributed to C12ORF4 and its relationship to ID is the fact that mast cell degranulation is one of the processes involved in brain inflammation (*Dong et al., 2017*; *Skaper et al., 2018*). Brain inflammation, in turn, can lead to ID and related disorders (*Di Marco et al., 2016*).

Exploitation of in silico databases for additional clues regarding C12ORF4 functions provides rather few hints. According to databases such as Expression Atlas (*Papatheodorou et al., 2018*), Human Protein Atlas (*Uhlen et al., 2015*) and Human Proteome Map (*Kim et al., 2014*), the C12ORF4 protein-coding gene (on a protein and transcript level) is expressed in most normal tissues, with a high level of expression in the brain and nerves, and in the genitals. In disease-specific studies, high expression levels of C12ORF4 have been reported for samples collected from glioblastoma multiforme and glioma, lymphoma, pancreatic adenocarcinoma, and ovarian adenocarcinoma (*Cancer Genome Atlas Research Network et al., 2013*). The cBioPortal database (*Cerami et al., 2012*) gives evidence that the C12ORF4 protein-coding gene is altered (mutation, but not deletion) in 3% of 1,939 cases (59 patients) of endometrial cancer, and smaller fractions of patients with other malignancies, for example, breast cancer, bladder cancer, colorectal cancer and melanoma. The more frequently occurring alterations of the C12ORF4 protein-coding gene are presented in Table S2. The most frequent mutation, missense R335Q, has been observed in 14 cases, mainly in uterine endometrioid carcinoma samples, but also in patients with other malignancies (cBioportal). Arg335 is located within the putative Macro domain region of C12ORF4, in the conserved motif DNR, containing Asp333, suggested by us to be a likely catalytic residue (Fig. 5).

According to tools for elucidation of biological relationship networks, such as Ingenuity Pathway Analysis (https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis) and String (*Szklarczyk et al., 2015*), C12ORF4 is involved in a functional network involving a number of mostly cytoplasmic signaling proteins. One of these interactors, TBCK, described as a modulator of mammalian TOR signaling, is involved in the regulation of cell proliferation and growth and control of actin-cytoskeleton organization. Interestingly, mutations in TBCK lead to a syndrome of ID (*Bhoj et al., 2016*). Other

interactors include the interleukin receptor IL6R, that is a proposed asthma risk locus (*Ferreira et al., 2011*) and is functionally related to a number of degranulation-related signaling mediators (see Fig. S2)

CONCLUSIONS

In summary, we present strong in silico evidence that DUF2362 belongs to the Macro clan/superfamily. The similarity to known Macro domains is clear and suggestive. It remains unclear whether this new Macro domain has a catalytic activity or only a "reader" function. However, several lines of evidence support the catalytic, not pseudoenzyme function, including conservation of a putative active site.

Turning a structure prediction into a useful molecular function and biological process prediction is always a challenge. Here, we approached the DUF2362/C12ORF4 family from several angles, trying to combine structural prediction with available functional data. The final functional answers, nevertheless, will only come from experiments, both biochemical and biological, that should cast full light on C12ORF4, an intriguing putative novel addition to eukaryotic ADP-ribosylation signaling. Thus, until decisive biochemical and structural biology data provide unequivocal functional validation, DUF2362 has to be regarded as a tentative new member of the Macro clan/superfamily.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests
The authors declare that they have no competing interests.

Author Contributions
- Małgorzata Dudkiewicz conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Krzysztof Pawłowski conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Data Availability
The following information was supplied regarding data availability:

The raw data (full results of bioinformatics analyses) is available as Supplementary Files.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.6863#supplemental-information.

REFERENCES

Alazami AM, Patel N, Shamseldin HE, Anazi S, Al-Dosari MS, Alzahrani F, Hijazi H, Alshammari M, Aldahmesh MA, Salih MA, Faqeih E, Alhashem A, Bashiri FA, Al-Owain M, Kentab AY, Sogaty S, Tala SAl, Temsah MH, Tulbah M, Aljelaify RF, Alshahwan SA, Seidahmed MZ, Alhadid AA, Aldhalaan H, AlQallaf F, Kurdi W, Alfadhel M, Babay Z, Alsogheer M, Kaya N, Al-Hassnan ZN, Abdel-Salam GM, Al-Sannaa N, Mutairi FAl, Khashab HYEl, Bohlega S, Jia X, Nguyen HC, Hammami R, Adly N, Mohamed JY, Abdulwahab F, Ibrahim N, Naim EA, Al-Younes B, Meyer BF, Hashem M, Shaheen R, Xiong Y, Abouelhoda M, Aldeeri AA, Monies DM, Alkuraya FS. 2015. Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. *Cell Reports* **10(2)**:148–161 DOI 10.1016/j.celrep.2014.12.015.

Allen MD, Buckle AM, Cordell SC, Löwe J, Bycroft M. 2003. The crystal structure of AF1521 a protein from *Archaeoglobus fulgidus* with homology to the non-histone domain of macroH2A. *Journal of Molecular Biology* **330(3)**:503–511 DOI 10.1016/S0022-2836(03)00473-X.

Anazi S, Maddirevula S, Salpietro V, Asi YT, Alsahli S, Alhashem A, Shamseldin HE, AlZahrani F, Patel N, Ibrahim N, Abdulwahab FM, Hashem M, Alhashmi N, Murshedi FAl, Kindy AAl, Alshaer A, Rumayyan A, Tala SAl, Kurdi W, Alsaman A, Alasmari A, Banu S, Sultan T, Saleh MM, Alkuraya H, Salih MA, Aldhalaan H, Ben-Omran T, Musafri FAl, Ali R, Suleiman J, Tabarki B, El-Hattab AW, Bupp C, Alfadhel M, Tassan NAl, Monies D, Arold ST, Abouelhoda M, Lashley T, Houlden H, Faqeih E, Alkuraya FS. 2017. Expanding the genetic heterogeneity of intellectual disability. *Human Genetics* **136(11–12)**:1419–1429 DOI 10.1007/s00439-017-1843-2.

Barkauskaite E, Brassington A, Tan ES, Warwicker J, Dunstan MS, Banos B, Lafite P, Ahel M, Mitchison TJ, Ahel I, Leys D. 2013. Visualization of poly(ADP-ribose) bound to PARG reveals inherent balance between *exo-* and *endo-*glycohydrolase activities. *Nature Communications* **4(1)**:2164 DOI 10.1038/ncomms3164.

Barkauskaite E, Jankevicius G, Ladurner AG, Ahel I, Timinszky G. 2013. The recognition and removal of cellular poly(ADP-ribose) signals. *FEBS Journal* **280(15)**:3491–3507 DOI 10.1111/febs.12358.

Bhoj EJ, Li D, Harr M, Edvardson S, Elpeleg O, Chisholm E, Juusola J, Douglas G, Sacoto MJG, Siquier-Pernet K, Saadi A, Bole-Feysot C, Nitschke P, Narravula A, Walke M, Horner MB, Day-Salvatore DL, Jayakar P, Vergano SA, Tarnopolsky MA, Hegde M, Colleaux L, Crino P, Hakonarson H. 2016. Mutations in TBCK, encoding TBC1-domain-containing kinase, lead to a recognizable syndrome of intellectual disability and hypotonia. *American Journal of Human Genetics* **98(4)**:782–788 DOI 10.1016/j.ajhg.2016.03.016.

Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. 2013. The cancer genome atlas pan-cancer analysis project. *Nature Genetics* **45(10)**:1113–1120 DOI 10.1038/ng.2764.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* **2(5)**:401–404 DOI 10.1158/2159-8290.CD-12-0095.

Chen C, Natale DA, Finn RD, Huang H, Zhang J, Wu CH, Mazumder R. 2011. Representative proteomes: a stable, scalable and unbiased proteome set for sequence analysis and functional annotation. *PLOS ONE* **6(4)**:e18910 DOI 10.1371/journal.pone.0018910.

Chen D, Vollmar M, Rossi MN, Phillips C, Kraehenbuehl R, Slade D, Mehrotra PV, Von Delft F, Crosthwaite SK, Gileadi O, Denu JM, Ahel I. 2011. Identification of macrodomain proteins as novel *O*-acetyl-ADP-ribose deacetylases. *Journal of Biological Chemistry* **286(15)**:13261–13271 DOI 10.1074/jbc.M110.206771.

Cohen MS, Chang P. 2018. Insights into the biogenesis, function, and regulation of ADP-ribosylation. *Nature Chemical Biology* **14(3)**:236–243 DOI 10.1038/nchembio.2568.

Di Marco B, Bonaccorso CM, Aloisi E, D'Antoni S, Catania MV. 2016. Neuro-inflammatory mechanisms in developmental disorders associated with intellectual disability and autism spectrum disorder: a neuro-immune perspective. *CNS & Neurological Disorders - Drug Targets* **15(4)**:448–463 DOI 10.2174/1871527315666160321105039.

Dong H, Zhang X, Wang Y, Zhou X, Qian Y, Zhang S. 2017. Suppression of brain mast cells degranulation inhibits microglial activation and central nervous system inflammation. *Molecular Neurobiology* **54(2)**:997–1007 DOI 10.1007/s12035-016-9720-x.

Dudkiewicz M, Szczepinska T, Grynberg M, Pawlowski K. 2012. A novel protein kinase-like domain in a selenoprotein, widespread in the tree of life. *PLOS ONE* **7(2)**e32138 DOI 10.1371/journal.pone.0032138.

Dunstan MS, Barkauskaite E, Lafite P, Knezevic CE, Brassington A, Ahel M, Hergenrother PJ, Leys D, Ahel I. 2012. Structure and mechanism of a canonical poly(ADP-ribose) glycohydrolase. *Nature Communications* **3(1)**:878 DOI 10.1038/ncomms1889.

Feijs KL, Forst AH, Verheugd P, Luscher B. 2013. Macrodomain-containing proteins: regulating new intracellular functions of mono(ADP-ribosyl)ation. *Nature Reviews Molecular Cell Biology* **14(7)**:443–451 DOI 10.1038/nrm3601.

Ferreira MAR, Matheson MC, Duffy DL, Marks GB, Hui J, Souëf PL, Danoy P, Baltic S, Nyholt DR, Jenkins M, Hayden C, Willemsen G, Ang W, Kuokkanen M, Beilby J, Cheah F, De Geus EJC, Ramasamy A, Vedantam S, Salomaa V, Madden PA, Heath AC, Hopper JL, Visscher PM, Musk B, Leeder SR, Jarvelin M-R, Pennell C, Boomsma DI, Hirschhorn JN, Walters H, Martin NG, James A, Jones G, Abramson MJ, Robertson CF, Dharmage SC, Brown MA, Montgomery GW, Thompson PJ. 2011. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378(9795)**:1006–1014 DOI 10.1016/S0140-6736(11)60874-X.

Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015. HMMER web server: 2015 update. *Nucleic Acids Research* **43(W1)**:W30–W38 DOI 10.1093/nar/gkv397.

Frickey T, Lupas A. 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20(18)**:3702–3704 DOI 10.1093/bioinformatics/bth444.

Gibson BA, Kraus WL. 2012. New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nature Reviews Molecular Cell Biology* **13(7)**:411–424 DOI 10.1038/nrm3376.

Han W, Li X, Fu X. 2011. The macro domain protein family: structure, functions, and their potential therapeutic implications. *Mutation Research/Reviews in Mutation Research* **727(3)**:86–103 DOI 10.1016/j.mrrev.2011.03.001.

Jankevicius G, Ariza A, Ahel M, Ahel I. 2016. The toxin-antitoxin system DarTG catalyzes reversible ADP-ribosylation of DNA. *Molecular Cell* **64(6)**:1109–1116 DOI 10.1016/j.molcel.2016.11.014.

Jankevicius G, Hassler M, Golia B, Rybin V, Zacharias M, Timinszky G, Ladurner AG. 2013. A family of macrodomain proteins reverses cellular mono-ADP-ribosylation. *Nature Structural & Molecular Biology* **20(4)**:508–514 DOI 10.1038/nsmb.2523.

Jaroszewski L, Li Z, Cai XH, Weber C, Godzik A. 2011. FFAS server: novel features and applications. *Nucleic Acids Research* **39(suppl)**:W38–W44 DOI 10.1093/nar/gkr441.

Karras GI, Kustatscher G, Buhecha HR, Allen MD, Pugieux C, Sait F, Bycroft M, Ladurner AG. 2005. The macro domain is an ADP-ribose binding module. *EMBO Journal* **24(11)**:1911–1920 DOI 10.1038/sj.emboj.7600664.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10(6)**:845–858 DOI 10.1038/nprot.2015.053.

Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LD, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TS, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A. 2014. A draft map of the human proteome. *Nature* **509(7502)**:575–581 DOI 10.1038/nature13302.

Langelier MF, Eisemann T, Riccio AA, Pascal JM. 2018. PARP family enzymes: regulation and catalysis of the poly(ADP-ribose) posttranslational modification. *Current Opinion in Structural Biology* **53**:187–198 DOI 10.1016/j.sbi.2018.11.002.

Lehmann LC, Hewitt G, Aibara S, Leitner A, Marklund E, Maslen SL, Maturi V, Chen Y, van der Spoel D, Skehel JM, Moustakas A, Boulton SJ, Deindl S. 2017. Mechanistic insights into autoinhibition of the oncogenic chromatin remodeler ALC1. *Molecular Cell* **68(5)**:847–859.e7 DOI 10.1016/j.molcel.2017.10.017.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* **44(W1)**:W242–W245 DOI 10.1093/nar/gkw290.

Li P, Huang P, Li X, Yin D, Ma Z, Wang H, Song H. 2018. Tankyrase mediates K63-linked ubiquitination of JNK to confer stress tolerance and influence lifespan in drosophila. *Cell Reports* **25(2)**:437–448 DOI 10.1016/j.celrep.2018.09.036.

Livingstone CD, Barton GJ. 1993. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Computer Applications in the Biosciences* **9(6)**:745–756.

Lüscher B, Bütepage M, Eckei L, Krieg S, Verheugd P, Shilton BH. 2018. ADP-ribosylation, a multifaceted posttranslational modification involved in the control of cell physiology in health and disease. *Chemical Reviews* **118(3)**:1092–1136 DOI 10.1021/acs.chemrev.7b00122.

Martzen MR, McCraith SM, Spinelli SL, Torres FM, Fields S, Grayhack EJ, Phizicky EM. 1999. A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286(5442)**:1153–1155 DOI 10.1126/science.286.5442.1153.

Mazuc E, Guglielmi L, Bec N, Parez V, Hahn CS, Mollevi C, Parrinello H, Desvignes JP, Larroque C, Jupp R, Dariavach P, Martineau P. 2014. In-cell intrabody selection from a diverse human library identifies C12orf4 protein as a new player in rodent mast cell degranulation. *PLOS ONE* **9(8)**e104998 DOI 10.1371/journal.pone.0104998.

Munnur D, Ahel I. 2017. Reversible mono-ADP-ribosylation of DNA breaks. *FEBS Journal* **284(23)**:4002–4016 DOI 10.1111/febs.14297.

Neuvonen M, Ahola T. 2009. Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites. *Journal of Molecular Biology* **385(1)**:212–225 DOI 10.1016/j.jmb.2008.10.045.

Palazzo L, Mikoč A, Ahel I. 2017. ADP-ribosylation: new facets of an ancient modification. *FEBS Journal* **284(18)**:2932–2946 DOI 10.1111/febs.14078.

Papatheodorou I, Fonseca NA, Keays M, Tang YA, Barrera E, Bazant W, Burke M, Fullgrabe A, Fuentes AM, George N, Huerta L, Koskinen S, Mohammed S, Geniza M, Preece J, Jaiswal P, Jarnuczak AF, Huber W, Stegle O, Vizcaino JA, Brazma A, Petryszak R. 2018. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Research* **46(D1)**:D246–D251 DOI 10.1093/nar/gkx1158.

Pawlowski K. 2008. Uncharacterized/hypothetical proteins in biomedical 'omics' experiments: is novelty being swept under the carpet? *Briefings in Functional Genomics and Proteomics* **7(4)**:283–290 DOI 10.1093/bfgp/eln033.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera? A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25(13)**:1605–1612 DOI 10.1002/jcc.20084.

Philips AK, Pinelli M, De Bie CI, Mustonen A, Määttä T, Arts HH, Wu K, Roepman R, Moilanen JS, Raza S, Varilo T, Scala G, Cocozza S, Gilissen C, van Gassen KL, Järvelä I. 2017. Identification of *C12orf4* as a gene for autosomal recessive intellectual disability. *Clinical Genetics* **91(1)**:100–105 DOI 10.1111/cge.12821.

Posavec M, Timinszky G, Buschbeck M. 2013. Macro domains as metabolite sensors on chromatin. *Cellular and Molecular Life Sciences* **70(9)**:1509–1524 DOI 10.1007/s00018-013-1294-4.

Rack JG, Perina D, Ahel I. 2016. Macrodomains: structure, function, evolution, and catalytic activities. *Annual Review of Biochemistry* **85(1)**:431–454 DOI 10.1146/annurev-biochem-060815-014935.

Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42(W1)**:W320–W324 DOI 10.1093/nar/gku316.

Rosenthal F, Feijs KL, Frugier E, Bonalli M, Forst AH, Imhof R, Winkler HC, Fischer D, Caflisch A, Hassa PO, Luscher B, Hottiger MO. 2013. Macrodomain-containing proteins are new mono-ADP-ribosylhydrolases. *Nature Structural & Molecular Biology* **20(4)**:502–507 DOI 10.1038/nsmb.2521.

Ruan J, Li H, Chen Z, Coghlan A, Coin LJ, Guo Y, Heriche JK, Hu Y, Kristiansen K, Li R, Liu T, Moses A, Qin J, Vang S, Vilella AJ, Ureta-Vidal A, Bolund L, Wang J, Durbin R. 2008. TreeFam: 2008 update. *Nucleic Acids Research* **36(Database)**:D735–D740.

Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research* **18(20)**:6097–6100 DOI 10.1093/nar/18.20.6097.

Sethi GS, Dharwal V, Naura AS. 2017. Poly(ADP-ribose) polymerase-1 in lung inflammatory disorders: a review. *Frontiers in Immunology* **8**:1172 DOI 10.3389/fimmu.2017.01172.

Sharifi R, Morra R, Appel CD, Tallis M, Chioza B, Jankevicius G, Simpson MA, Matic I, Ozkan E, Golia B, Schellenberg MJ, Weston R, Williams JG, Rossi MN, Galehdari H, Krahn J, Wan A, Trembath RC, Crosby AH, Ahel D, Hay R, Ladurner AG, Timinszky G, Williams RS, Ahel I. 2013. Deficiency of terminal ADP-ribose protein glycohydrolase TARG1/C6orf130 in neurodegenerative disease. *EMBO Journal* **32(9)**:1225–1237 DOI 10.1038/emboj.2013.51.

Skaper SD, Facci L, Zusso M, Giusti P. 2018. An inflammation-centric view of neurological disease: beyond the neuron. *Frontiers in Cellular Neuroscience* **12**:72 DOI 10.3389/fncel.2018.00072.

Slade D, Dunstan MS, Barkauskaite E, Weston R, Lafite P, Dixon N, Ahel M, Leys D, Ahel I. 2011. The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase. *Nature* **477(7366)**:616–620 DOI 10.1038/nature10404.

Soding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* **33(suppl_2)**:W244–W248 DOI 10.1093/nar/gki408.

Song GL, Jin CC, Zhao W, Tang Y, Wang YL, Li M, Xiao M, Li X, Li QS, Lin X, Chen WW, Kuang J. 2016. Regulation of the RhoA/ROCK/AKT/β-catenin pathway by arginine-specific ADP-ribosytransferases 1 promotes migration and epithelial-mesenchymal transition in colon carcinoma. *International Journal of Oncology* **49(2)**:646–656 DOI 10.3892/ijo.2016.3539.

Sreelatha A, Yee SS, Lopez VA, Park BC, Kinch LN, Pilch S, Servage KA, Zhang J, Jiou J, Karasiewicz-Urbanska M, Lobocka M, Grishin NV, Orth K, Kucharczyk R, Pawlowski K, Tomchick DR, Tagliabracci VS. 2018. Protein AMPylation by an evolutionarily conserved pseudokinase. *Cell* **175(3)**:809–821.e19 DOI 10.1016/j.cell.2018.08.046.

Suzuki Y, Masini E, Mazzocca C, Cuzzocrea S, Ciampa A, Suzuki H, Bani D. 2004. Inhibition of poly(ADP-ribose) polymerase prevents allergen-induced asthma-like reaction in sensitized Guinea pigs. *Journal of Pharmacology and Experimental Therapeutics* **311(3)**:1241–1248 DOI 10.1124/jpet.104.072546.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, Von Mering C. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43(Database)**:D447–D452 DOI 10.1093/nar/gku1003.

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, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F. 2015. Proteomics. Tissue-based map of the human proteome. *Science* **347(6220)**:1260419 DOI 10.1126/science.1260419.

Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. 2009. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research* **19(2)**:327–335 DOI 10.1101/gr.073585.107.

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25(9)**:1189–1191 DOI 10.1093/bioinformatics/btp033.

Webb B, Sali A. 2014. Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics* **47**:5.6.1–5.6.32 DOI 10.1002/0471250953.bi0506s47.

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(1)**:7–8 DOI 10.1038/nmeth.3213.

Young JC, Clements A, Lang AE, Garnett JA, Munera D, Arbeloa A, Pearson J, Hartland EL, Matthews SJ, Mousnier A, Barry DJ, Way M, Schlosser A, Aktories K, Frankel G. 2014. The *Escherichia coli* effector EspJ blocks Src kinase activity via amidation and ADP ribosylation. *Nature Communications* **5(1)**:5887 DOI 10.1038/ncomms6887.

Zaffini R, Gotte G, Menegazzi M. 2018. Asthma and poly(ADP-ribose) polymerase inhibition: a new therapeutic approach. *Drug Design, Development and Therapy* **12**:281–293 DOI 10.2147/DDDT.S150846.

Zarkovic G, Belousova EA, Talhaoui I, Saint-Pierre C, Kutuzov MM, Matkarimov BT, Biard D, Gasparutto D, Lavrik OI, Ishchenko AA. 2018. Characterization of DNA ADP-ribosyltransferase activities of PARP2 and PARP3: new insights into DNA ADP-ribosylation. *Nucleic Acids Research* **46(5)**:2417–2431 DOI 10.1093/nar/gkx1318.

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(W1)**:W291–W299 DOI 10.1093/nar/gkx366.