Received: 16 January 2023 | Revised: 14 June 2023 | Accepted: 7 July 2023

DOI: 10.1002/prot.26558

RESEARCH ARTICLE

PROTEINS WILEY

Conserved functions of prion candidates suggest a primeval role of protein self-templating

Tomasz Zajkowski[1,2,3,4] | Michael D. Lee[5,6] | Siddhant Sharma[5,7] | Alec Vallota-Eastman[8] | Mikołaj Kuska[2,9] | Małgorzata Malczewska[2,9] | Lynn J. Rothschild[10]

[1]Universities Space Research Association at NASA Ames Research Center, Mountain View, California, USA

[2]Polish Astrobiology Society, Warsaw, Poland

[3]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California, USA

[4]Centre of New Technologies, University of Warsaw, Warsaw, Poland

[5]Blue Marble Space Institute of Science, Seattle, Washington, USA

[6]KBR, NASA Ames Research Center, Mountain View, California, USA

[7]School of Chemistry, University of New South Wales, Sydney, Australia

[8]Department of Earth Science, University of California, Santa Barbara, California, USA

[9]Department of Biophysics, Faculty of Physics, University of Warsaw, Warsaw, Poland

[10]Space Science and Astrobiology Division, NASA Ames Research Center, Mountain View, California, USA

Correspondence
Tomasz Zajkowski, Universities Space Research Association, NASA Ames Research Center 239-20, Moffett Field, Mountain View, CA 94035, USA.
Email: tzajkowski@usra.edu

Funding information
National Aeronautics and Space Administration's Planetary Science Division Research Program and Ames Research Innovation Award (ARIA); National Science Foundation, Grant/Award Number: INTERN NSF DCL 21-013

Abstract

Amyloid-based prions have simple structures, a wide phylogenetic distribution, and a plethora of functions in contemporary organisms, suggesting they may be an ancient phenomenon. However, this hypothesis has yet to be addressed with a systematic, computational, and experimental approach. Here we present a framework to help guide future experimental verification of candidate prions with conserved functions to understand their role in the early stages of evolution and potentially in the origins of life. We identified candidate prions in all high-quality proteomes available in UniProt computationally, assessed their phylogenomic distributions, and analyzed candidate-prion functional annotations. Of the 27 980 560 proteins scanned, 228 561 were identified as candidate prions (~0.82%). Among these candidates, there were 84 Gene Ontology (GO) terms conserved across the three domains of life. We found that candidate prions with a possible role in adaptation were particularly well-represented within this group. We discuss unifying features of candidate prions to elucidate the primeval roles of prions and their associated functions. Candidate prions annotated as transcription factors, DNA binding, and kinases are particularly well suited to generating diverse responses to changes in their environment and could allow for adaptation and population expansion into more diverse environments. We hypothesized that a relationship between these functions and candidate prions could be evolutionarily ancient, even if individual prion domains themselves are not evolutionarily conserved. Candidate prions annotated with these universally occurring functions potentially represent the oldest extant prions on Earth and are therefore excellent experimental targets.

KEYWORDS
adaptation, aggregation, bioinformatics, evolution, heredity, LUCA, phylogenetics, PLAAC, prion, protein inheritance, stress response, the origin of life, tree of life

1 | INTRODUCTION

1.1 | Amyloids and prions

Many proteins can form fibrillar aggregates called amyloids. These were originally discovered as factors involved in the development of neurodegenerative diseases such as Alzheimer, Parkinson, and Huntington diseases tree but have since been shown to have a role in many nonpathogenic and, indeed, beneficial physiological functions as well.[1-3] Three lines of argumentation suggest that amyloids might have been present on Earth even before the first forms of what we would consider life evolved. First, prebiotically plausible peptides can

form amyloid fibrils.[4,5] Second, the amyloid fold (the structure when proteins form amyloid fibrils) is considered the lowest point on the energy landscape for protein folding and aggregation. Therefore amyloidogenesis is considered as an intrinsic property of polypeptides.[6,7] Third, amyloids bear a phenomenological similarity to crystals and could have occurred abiotically.[8] Given these arguments, it is expected that amyloids should be old and very common in nature. Supporting this point of view, amyloids are known to perform multiple physiological functions in diverse organisms, from microorganisms to humans. Organisms capitalize both on the physical rigidity of preformed fibrils as well as on the process of aggregation itself. For example, the durability and strength of amyloid fibers are conveyed in the properties of silkworm and spider silk.[9,10] In humans, functional amyloids build scaffolds on which melanin (skin pigment) is deposited.[11] In bacteria and fungi, highly-conserved amyloid-forming proteins are components of the extracellular matrix and contribute to cell adhesion and biofilm formation.[12,13]

Other physiological processes rely on the ability of a protein to exist in two different states: soluble (an individual protein capable of carrying out its typically understood function(s)) and aggregated (multiple proteins aggregated into the amyloid form, not capable of carrying out their typically understood individual function(s)). The amyloid formation has characteristics of autocatalysis and, after reaching a certain threshold of aggregation, often leads to the quick depletion of the soluble fraction of protein that becomes trapped in the form of fibrillar aggregates.[14] This ability to exist in two very different and self-excluding states (soluble or aggregated) allows some amyloid-forming proteins to function as molecular switches.[15-17] In some cases, the amyloid form can be inherited cytoplasmically during cellular replication, where it will continue the conversion of a soluble fraction of freshly synthesized protein into the non-functional aggregated fibrillar form. A cell harboring such self-templating and self-perpetuating aggregation of a protein phenotypically resembles a cell with a deletion of the gene that codes for this protein.[18] These heritable protein aggregates are called *prions*. While some proteins can be prions without being amyloid-based, like intrinsically disordered RNA-binding proteins in *Saccharomyces cerevisiae*,[19] here we focus on amyloid-based prions. Recently, Dennis and Garcia reviewed both amyloid-forming and non-amyloid-forming prions.[20]

Prions can be sustained for many generations.[21] An individual protein that contributes to the prion aggregate is called a "prion protein," and a distinct protein domain that is responsible for prion aggregation is called a prion domain (PrD). Just like amyloids, prions were first discovered in association with diseases but are now known to regulate a variety of beneficial physiological functions. For example, in plants, prions regulate flowering time.[22] In yeast, among numerous other functions, prions can drive large shifts in metabolism, regulating whether cells are metabolic specialists or generalists.[15] In bacteria, prion aggregation can regulate translation[23] and plasmid copy number.[24] Prion inheritance is often referred to as epigenetic; epi- (ἐπι- "over, outside of, around") implies features that are "on top of" the traditional genetic basis for inheritance. Some prions even regulate other mechanisms of epigenetic inheritance via chromatin modification.[25,26]

Examples of proteins capable of forming prions are widespread on the phylogenetic tree of life, showing that prion formation is common and that it may be an early-evolved cellular mechanism. It is not clear if this mechanism did indeed evolve early and has been conserved as many forms of life have diverged or if it has evolved multiple times independently via convergent evolution. We call the second hypothesis *the recurring domestication of prions*.

To pursue the question of the origin of prions, we searched ~5000 high-quality proteomes for proteins predicted to contain prion-like domains (which we herein refer to as candidate prion proteins or cPrPs) and analyzed the functional annotations associated with cPrPs and all proteins not identified as cPrPs. We then focused on functional annotations of cPrPs that were identified in all three domains of life. We suspected that these conserved functions associated with cPrPs might be evolutionarily linked to prion aggregation. Determining the functions of the most conserved prion proteins might help unravel the role of the prion phenomenon in biology today, its evolutionary age, and the role it might have played in the origin of life.

1.2 | Conservation and distribution of prion-associated functions on the tree of life

Different types of prion-forming domains exist. Some, but not all, prions contain specific regions called PrDs that can be necessary and sufficient to grant a protein prionogenic behavior.[27] Many PrDs are also predicted to contain intrinsically disordered regions (IDRs). These regions have been shown to play a regulatory role in forming biomolecular condensates[28] and modulating protein solubility and phase behavior[29]—behaviors that can be important in cell signaling.[30,31] But it is worth noting that the presence of these IDRs does not necessarily indicate the encompassing protein is a prion. The inheritance of phenotypic traits due to prion formation involves a structural conversion, which in a majority of cases is driven by a Q/N-rich PrD.[32] Most known PrDs are enriched in glutamine and asparagine (Q/N enriched) and depleted in charged amino acids. However, known prion proteins like the mammalian PrP,[33] Het-s,[34] and Mod5[35] do not have this bias but are still characterized by an intrinsically disordered region. In both cases, prion proteins can be conserved across organisms. The first prion-forming protein ever described, PrP—has no Q/N bias and is conserved up to early chordates,[36-38] which gave rise to early tetrapods about 380 million years ago.[39] The Q/N bias in the PrD of the translation terminator (Sup35) is found in both *Ascomycota* and *Basidiomycota*, which diverged more than 1 billion years ago.[40] While many Q/N-rich prion proteins have been identified as conserved between different types of yeast, no conservation was detected in Q/N-rich prion proteins between humans and yeast.[41] And so far, no conservation has been shown to reach across different domains of life.

Prion proteins can be highly similar in sequence or, more often, similar in the chemical properties of the sequence.[42] In *S. cerevisiae*, Ure2p, and Sup35p sustain their ability to form prions even when the amino-acid sequence of their PrDs is "scrambled" as long as they

maintain the same general chemical properties of the amino-acid composition.[43,44] The fact that prion aggregation depends on the composition of PrDs and is (somewhat) independent of exact primary sequence challenges conventional bioinformatics approaches when applied to prions. Nevertheless, Su and Harrison[42] were able to identify PrDs that are both conserved in the primary sequence and amino-acid composition in *S. cerevisiae*. In their dataset, the most conserved prion-like domain across the *Saccharomyces* genus was the protein NRP1, originally described as a prion by Alberti et al.[45] NRP1 is a putative RNA-binding protein that localizes to dense amorphous aggregations in the cytosol called stress granules.[46] Other examples of confirmed conserved prion include GLFG-motif nucleoporin NUP100, a confirmed prion-forming protein, which is part of the nuclear pore complex[47]; GLN3, a transcriptional activator of genes regulated by nitrogen catabolite repression; RBS1, a protein involved in the assembly of the RNA polymerase III (Pol III) complex[48]; and MED3/PGD1, a subunit of the RNA polymerase II mediator complex.[49] Having a role in RNA binding and gene regulation is a common theme for many prion candidates, not only in *Saccharomyces*,[50,51] and is often connected to phase transitions,[52] again, promoting the formation of stress granules.[53,54]

Our current information on confirmed prions is limited because it is difficult and time-consuming to verify them experimentally. To broaden our understanding of the ancient functions of prions beyond this small group of confirmed prions, we include functions of proteins that are predicted to harbor prion-like domains and, thus, may behave as true prions. The development of prion-prediction algorithms has enabled proteome-wide analysis of proteins revealing thousands of candidate prions.[45,55-63] These efforts were recently reviewed by Batlle and Gil-Garcia.[64,65] The most widely used prediction program is Prion-Like Amino Acid Composition (PLAAC).[22,59] PLAAC was designed to identify prion-like proteins, specifically proteins with similar amino-acid compositions to yeast prions, and is not a perfect prion-prediction tool. In fact, analysis of predictions from a first-generation version of the algorithm[45] showed that only 19% of predicted proteins were experimentally verified to form prions though later work scanning archaea with PLAAC yielded 6/16 (37.5%) experimentally verified prion-like elements.[66] Despite its limitations, has proven successful in facilitating the identification of prions in the domain archaea,[66] Bacteria,[23,24,67,68] and even in viruses and phages.[69-71] This observation alone speaks to the chemically conserved nature of at least a subset of candidate prion domains (cPrDs) across all domains of life.

Several studies have made a rudimentary attempt to describe a phylogenetic distribution of predicted prion candidates in all domains of life. In one such study, Angarica noticed that there are much fewer prion candidates in the domains Bacteria and Archaea than in Eukarya.[72] This may be because bacterial and archaeal proteomes appear to have fewer Q/N-rich regions in general as compared with eukaryotes.[60] Investigations into the phylogenetic relationships of prion candidates were further augmented with analysis of predicted function derived from Gene Ontology (GO) annotations.[73] Angarica et al. performed the first wide-ranging study of prion candidates in proteomes.[57] Using GO annotations, they observed that predicted prionogenic domains (or candidate prion domains [cPrDs]) co-exist with different functional domains of proteins that localize to different cellular compartments, depending on the taxon and organism group. In bacteria, cPrDs are significantly enriched in proteins with annotations involved with the cell wall. Accordingly, bacterial cPrPs appear to be involved in metabolic and catabolic processes resulting in the construction and disassembly of the cell wall.[57] Another study similarly showed that bacterial prion candidates are associated with peripheral rearrangement, macromolecular assembly, cell adaptability, and invasion.[74] This logical convergence of GO cellular localization and function encourages further use of this method as a potentially valuable tool for studying global trends in the roles of prions in biology.

In our previous work, we identified prion candidates in archaea using PLAAC and looked into their GO annotations. Similar to what Angarica et al. found in bacteria, many prion candidates were involved in the construction and adhesion of the cell wall. Archaeal prion candidates were also significantly functionally enriched in regulation through transcription, calcium-binding, and copper ion binding[66]—all of which could be evolutionarily early-evolved functions. As one example, resistance to copper toxicity would have been necessary in early Earth environments.[75] It is, therefore, possible that PrDs may be essential to the function of many proteins involved in copper binding and may have facilitated the evolution of copper tolerance in early microbial life.

Recently Garai et al. analyzed the functional divergence of prion candidates in plants.[76] They found the highest density of prion candidates in green algae. They concluded that the prion phenomenon of aggregation had been conserved from chlorophytes to angiosperms, possibly offering an advantage during the evolution of plants. Like bacteria, unicellular algae might have benefited from the physical properties of amyloid-based prions contributing to cell–cell adhesion of their biofilms.[77] Rice also had a high number of cPrPs, with the candidate prions significantly enriched in transposons/retrotransposons. The authors of this analysis suggested the role of candidate prions in stress response and memory pathways in the plant kingdom.[76] These observations suggest the role of plant prions in adaptation to changing environments—an adaptation that has been proven in the case of yeast.[78,79] It is quite possible that the role of prions in facilitating adaptation could be an evolutionarily ancient one, even if individual prion proteins themselves are not evolutionarily conserved.

Putative PrDs rich in Q/N similar to those found in yeast are often detected in large numbers in eukaryotic proteomes. For example, in the genomes of *Drosophila melanogaster*, *Plasmodium falciparum*, *Helobdella robusta*, and *Dictyostelium*, more than 20% of proteins contain cPrDs.[41,80,81] GO annotations of prion candidates in *Plasmodium* point to their role in the regulation of gene expression similar to those described by Garai et al. in plants.

Taken together, previous work suggests that the prion phenomenon is widespread but can be randomly distributed even among closely related species. Functions associated with prion proteins vary widely, but functions associated with survival and regulation seem to

be more common than others. This tendency can be observed in organisms belonging to different kingdoms within Eukarya as well as across different domains of life, indicating a substantial consistency in the conservation of prion-related functions across different evolutionary epochs. So far, the analysis of prions in proteomes has been mostly focused on selected taxa, and the evolutionary conservation of prion proteins is mostly studied within taxonomic groups. Researchers agree that some prions can be very evolutionarily old, but the question of "how old?" is rarely asked. Identifying a subset of the potentially earliest evolved prions on Earth by looking for similar sequences and functions in a large sampling of all three domains of life could be a valuable step toward understanding the roles of prions in evolution—as well as their roles in contemporary organisms including the pathology of devastating amyloid-associated disorders. To this end, we analyzed all high-quality proteomes available in UniProt using prion-prediction software to prepare a list of cPrPs whose functions are common to all domains. We discuss unifying features of these candidate prions in an effort to elucidate the primeval roles of prions and their associated functions. And we present a framework to help guide future experimental verification of candidate prions with conserved functions to understand their role in the early stages of evolution and potentially in the origins of life. Experimental verification and detailed analysis of these candidates could reveal the possible role(s) of prions at the very early stages of evolution, including the time of the last universal common ancestor (LUCA).

2 | METHODS

2.1 | Computational identification of prion candidates

Reference proteomes for Archaea, Bacteria, and Eukarya were downloaded from the UniProt database[82]; accessed June 1, 2021) including their stored Gene Ontology (GO) annotations.[82,83] Proteomes possibly of low quality were excluded by retaining only those ranked as "standard" based on UniProt's "Complete Proteome Detector" algorithm. This resulted in 1151 archaeal proteomes containing 2 182 921 proteins; 2836 bacterial proteomes holding 10 939 944 proteins; and 925 eukaryal proteomes holding 14 857 695 proteins. The command-line version of PLAAC,[45,59] installed from https://github.com/whitehead/plaac on September 9, 2020, was used to identify proteins with cPrDs in each proteome individually with default settings other than "-a 0." Unless otherwise noted, those with a COREscore >0 (other than "NA") were considered to contain a candidate prion domain (cPrD). This COREscore is the sum of the individual log-likelihood ratios for each amino acid residue contained within the PrD.

2.2 | GO enrichment analysis

GO enrichment analyses were performed to identify enrichment (i.e., over-representation) or purification (i.e., under-representation) of frequencies of GO annotations in proteins with cPrDs compared with all proteins scanned. This was performed with the goatools v0.8.12[84] "find_enrichment.py" script with default settings, which includes propagating counts to parents and was done for each domain of life separately. Statistical significance was defined as those with Benjamini–Hochberg false-discovery rates of ≤0.05. Plots of enrichment values were generated with R v3.6.3, with the ggplot2 package, v3.3.5.[85] The Venn diagram was generated with the R package ggvenn v0.1.9 (Linlin Yan, 2022; https://CRAN.R-project.org/package=gg).

2.3 | Phylogenomic tree construction

Phylogenomic trees for each domain were produced with GToTree v1.6.11[86] using the prepackaged single-copy gene sets for archaea (76 target genes) and bacteria (74 target genes), and a universal set (16 target genes) for eukarya.[87] Within GToTree, target genes were identified with HMMER3 v3.2.2,[88] individually aligned with muscle v3.8.1551,[89] trimmed with trimAl,[90] and concatenated before phylogenetic estimation with FastTree2 v2.1.10.[91] The trees were initially visualized and edited through the Interactive Tree of Life site.[92]

2.4 | KO annotation of prion candidates

In addition to utilizing the GO annotations that were available with the UniProt data, functional annotation of cPrPs based on Kegg Orthology (KO) terms[93] was performed with KOfamScan v1.3.0.[94]

2.5 | Additional data and code availability

All of the UniProt data accessed on June 1, 2021 and utilized in this work, along with annotated code and additional data files too large to include here, (such as annotation and sequence information for all proteins), and all supplemental tables are available at https://figshare.com/projects/Zajkowski_et_al_2022_3-domain_prion_data_and_code_repository/133155. Additionally, Conda was utilized for program installation and environment management, and bit v1.8.42[95] and TaxonKit v0.9.0[96] were utilized heavily for helper scripts and manipulating taxonomy IDs.

3 | RESULTS AND DISCUSSION

3.1 | Computational identification and phylogenetic distribution of prion candidates

3.1.1 | Archaea

A total of 1151 archaeal proteomes containing 2 182 921 proteins were scanned with PLAAC (Table S1). Of these, 728 proteomes had at

least 1 cPrD detected (~63%). Of all proteins scanned a total of 2191 contained a cPrD (~0.1%). The frequency of identified cPrDs within each archaeal proteome was ~1.90 ± 2.78 (mean ± 1 SD), with a median of 1 (Table S2). Normalized per 1000 proteins per proteome (permil), it was 1.12 ± 1.67, with a median of 0.6. The distribution of those with at least one cPrD detected spans the archaeal proteomes scanned, as depicted by the blue tips in Figure 1. Those with a higher density of cPrDs (meaning when normalized to per 1000 proteins, "permil") included members of the classes Methanomicrobia (≥10 permil) and Methanobacteria and candidatus Micrarchaeota (≥15 permil; Figure 1; Table S2).

3.1.2 | Bacteria

For bacteria, 2836 proteomes containing 10 939 944 proteins were scanned (Table S1). Of these, 2547 proteomes had at least 1 cPrD detected (~90%). Of all proteins scanned, a total of 15 861 contained a cPrD (~0.15%). Counts of identified cPrDs in bacterial proteomes averaged ~5.59 ± 5.97 (mean ± 1SD) with a median of 4. Normalized, this was ~1.39 ± 1.40 with a median of 1.04 per 1000 proteins (permil). Those with relatively higher permil values included a member of the genus *Candidatus deianiraea* of the Alphaproteobacteria class (≥10 permil) and members of the genus *Mycoplasma* within the Phylum Tenericutes (≥15 permil; Figure 2; Table S2).

3.1.3 | Eukarya

A total of 925 Eukarya proteomes containing 14 857 695 proteins were scanned with PLAAC (Table S1). Of these, 907 proteomes had at least 1 cPrD detected (~98%). Of all proteins scanned, a total of 210 509 contained a cPrD (~1.4%). The frequency of cPrDs in eukaryal proteomes was ~228 ± 154.4 (mean ± 1SD) with a median of 189. Normalized to per 1000 proteins (permil) this was ~16.31 ± 12.24, with a median of 12.43 (Table S2). There were generally many more cPrDs identified in the Eukarya domain compared with Archaea or Bacteria. This could be partially explained by the fact that the PLAAC algorithm was trained on known eukaryal prions and might carry a bias toward the amino acid composition of eukaryal prions. Therefore, it should not currently be concluded this is a true trend and not a methodology bias. One quickly apparent trend in terms of the density of cPrDs in Eukarya is that there are consistently lower permil values in the Phylum Chordata (Figure 3; Table S2).

3.2 | Analysis of functions of prion candidates

Using PLAAC, we identified thousands of proteins bearing prion-like amino acid compositions (1151 in Archaea, 15 861 in Bacteria, and 210 509 in Eukarya). Testing all of these proteins if they are true prions experimentally is virtually impossible. Because we are

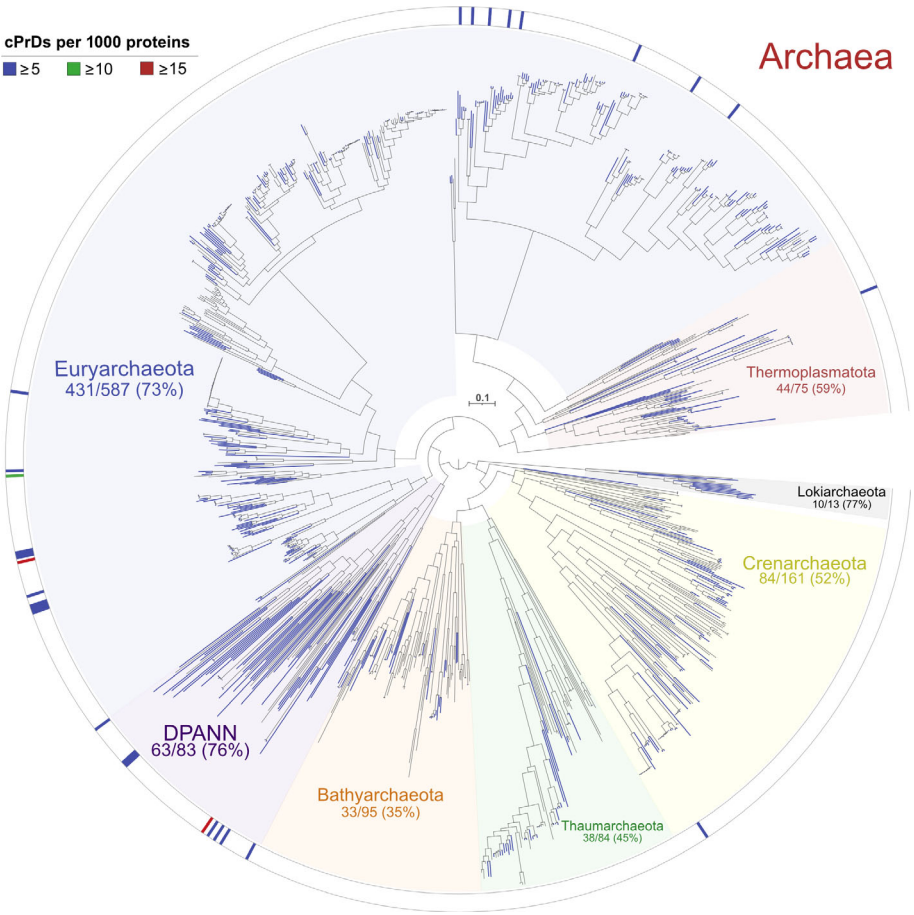


FIGURE 1 Phylogenomic tree of analyzed archaeal proteomes with their distribution of cPrDs overlain. Tips with at least 1 cPRD identified are colored blue. The outer ring depicts those with greater than 5, 10, or 15 cPrDs per 1000 proteins within a proteome.

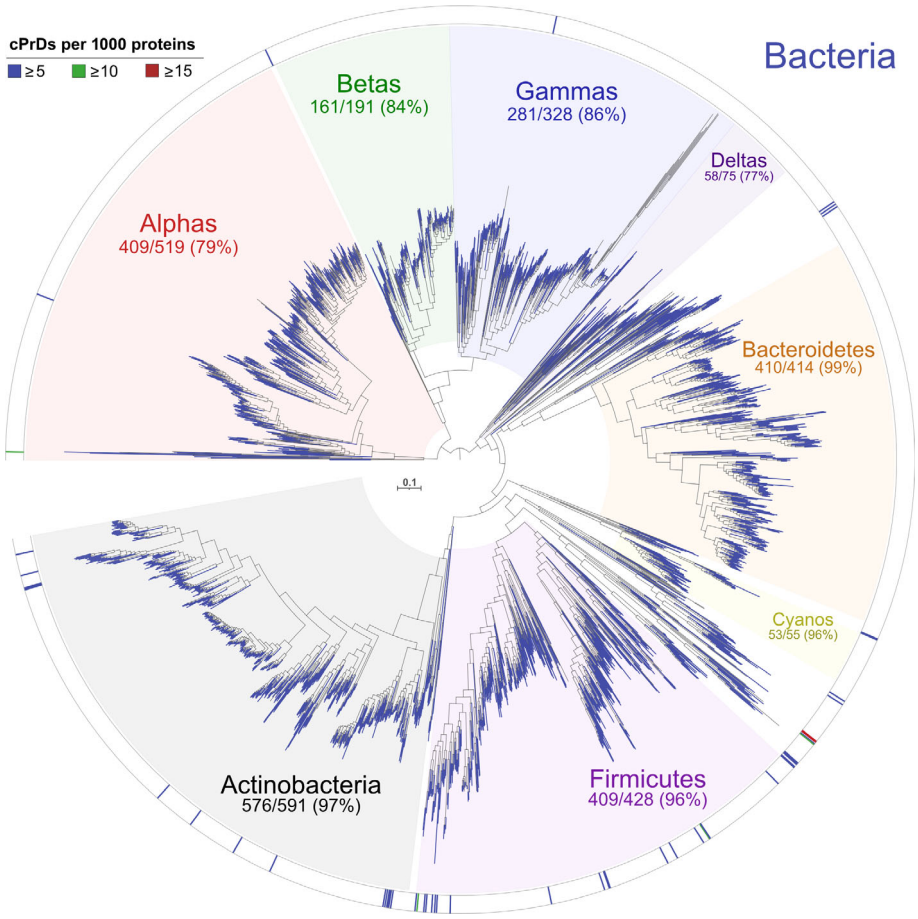


FIGURE 2 Phylogenomic tree of analyzed bacterial proteomes with their distribution of cPrDs overlain. Tips with at least 1 cPRD identified are colored blue. The outer ring depicts those with greater than 5, 10, or 15 cPrDs per 1000 proteins within a proteome.

interested in finding if prion formation could be an evolutionarily conserved phenomenon, we devised a line of reasoning that would direct us to a smaller pool of PLAAC-predicted proteins that would be of a reasonable size to analyze experimentally. We decided to identify functions of proteins predicted to harbor prion-like domains (candidate prions or cPrPs) to see if any span across all three domains of life. Conservation of function among proteins harboring these domains across all three domains could hint at their conserved role and, with it, the conservation of prions themselves if these proteins turn out to be true prions after experimental verification. To facilitate the identification of evolutionarily conserved cPrPs' functions within our dataset, we leveraged the Gene Ontology (GO) annotations associated with all proteins in the UniProt database. GO assigns information to a protein in the context of 3 "namespaces"—molecular function (MF), biological process (BP), and cellular component (CC)—through a highly curated process involving both manual and automated methods.[73] In addition to namespaces, GO terms are stored in a hierarchical structure of parent–child relationships and denoted as specific "depth" levels. Within a given namespace, a lower depth level (as in a lower number, for example, a depth of 1) will typically denote a broader, less specific annotation than a GO term with a depth level of 2. GO annotations can be thought of more as protein-domain level functional annotations than as full-protein functional annotations. To complement the GO annotations in our analysis of functions associated with cPrDs, we also annotated cPrDs with Kyoto Encyclopedia of Genes and Genomes (KEGG)[97] KEGG Orthology (KO) terms—which are more akin to full-protein functional annotations.

3.2.1 | Enriched GO terms in prion candidates

Within each domain of life, we tested for enrichment or purification (see Section 2) of specific GO terms in our identified cPrD-containing proteins as compared with all of the proteins scanned. In the archaeal domain, 51 GO terms were found to be significantly enriched (meaning these specific GO terms were more likely to be found in a protein with an identified candidate prion domain than in a protein without one), and 345 were significantly purified (meaning these specific GO terms were more likely to be associated with a protein that did not have an identified cPrD; based on a Benjamini–Hochberg false-discovery rate of ≤0.05; Table S3). For Bacteria, 248 were found to be significantly enriched, and 1243 were significantly purified (Table S4). And in the Eukarya, 2601 GO terms were found to be significantly enriched, and 6413 significantly purified (Table S5).

First, we focused on enriched GO terms (those more likely to be found in proteins with cPrDs than in proteins without cPrDs) to pursue functions that are consistently associated with candidate prions. Within the GO molecular function namespace, there were two GO terms that were found enriched in all three domains: helicase activity

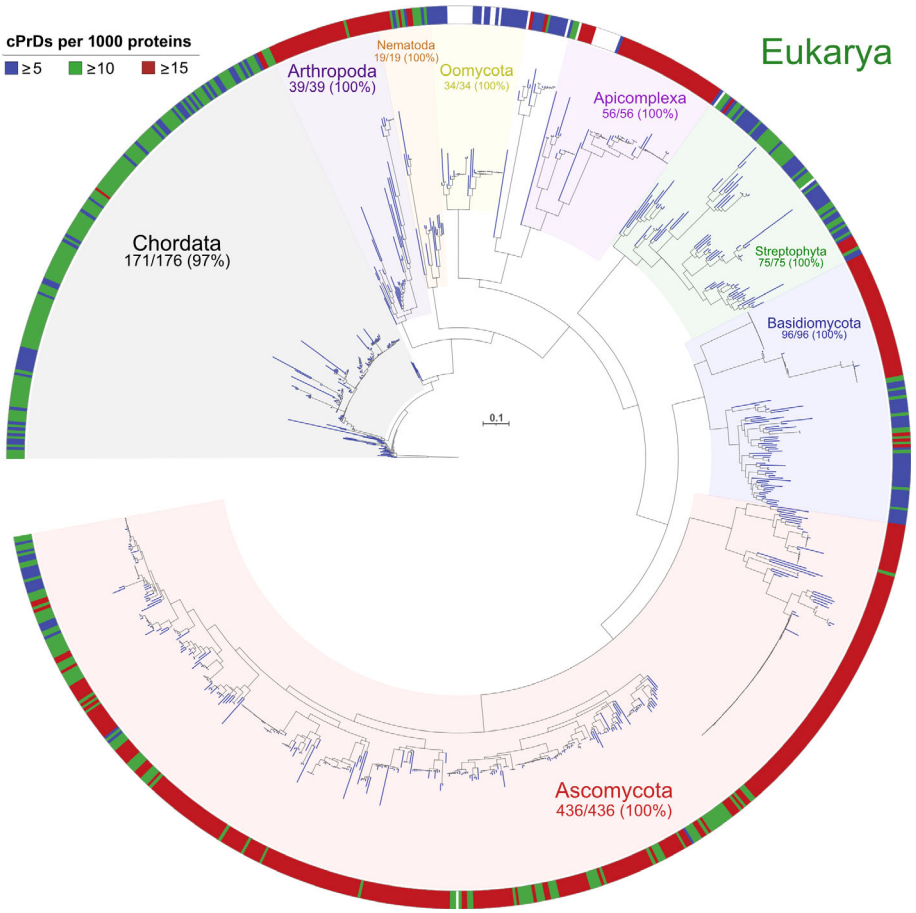


FIGURE 3 Phylogenomic tree of analyzed eukaryal proteomes with their distribution of cPrDs overlain. Tips with at least 1 cPrD identified are colored blue. The outer ring depicts those with greater than 5, 10, or 15 cPrDs per 1000 proteins within a proteome.

TABLE 1 GO terms that are statistically more likely to be found with cPrDs than without in multiple domains.

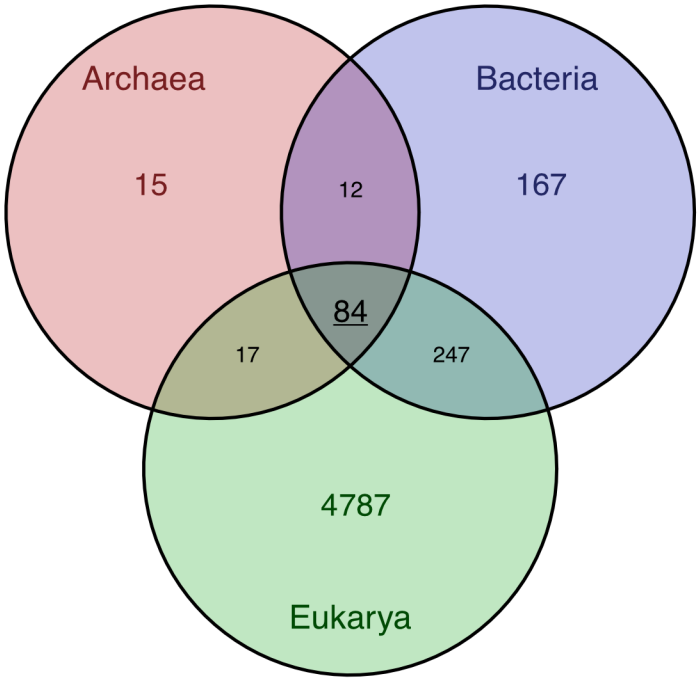
Domains with shared enriched GO terms	Shared enriched molecular function GO terms
Archaea, Bacteria, and Eukarya	Helicase activity (GO:0004386)
	Calcium ion binding (GO:0005509)
Archaea and Bacteria	Cysteine-type endopeptidase activity (GO:0004197)
	Serine-type endopeptidase activity (GO:0004252)
	Helicase activity (GO:0004386)
	Hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553)
	Calcium ion binding (GO:0005509)
	ATP-dependent activity, acting on DNA (GO:0008094)
	Hydrolase activity, acting on glycosyl bonds (GO:0016798)
	Carbohydrate-binding (GO:0030246)
	Endo-1,4-beta-xylanase activity (GO:0031176)
	Xylanase activity (GO:0097599)

(GO:0004386) and calcium ion binding (GO:0005509). For the biological process and cellular component namespaces, more broad-level terms were the only ones enriched: macromolecule catabolic process (GO:0009057) and the root term of cellular component (GO:0005575).

There were 10 molecular functions GO terms that were found enriched in both Bacteria and Archaea (without eukaryotes) (Table 1). Within the GO namespace biological process, biological adhesion, and metabolic process were found enriched. Protein ubiquitination (GO:0016567) and xylan catabolic process (GO:0045493) were the most specific descriptions found (depth 9). Within the cellular component namespaces in both Bacteria and Archaea, we found integral components of the membrane (GO:0016021) and extracellular region (GO:0005576) to be enriched (Table S6).

3.2.2 | Shared GO terms in prion candidates across domains

Among GO annotations that overlap all three domains of life, not focusing on enriched or not, we identified 84 that were common to all three domains (54 molecular functions, 22 biological processes, and 8 cell components; Figure 4).



Biological Process
GO:0000272 | polysaccharide catabolic process
GO:0005975 | carbohydrate metabolic process
GO:0006139 | nucleobase-containing compound metabolic process
GO:0006260 | DNA replication
GO:0006281 | DNA repair
GO:0006308 | DNA catabolic process
GO:0006313 | transposition, DNA-mediated
GO:0006355 | regulation of transcription, DNA-templated
GO:0006457 | protein folding
GO:0007155 | cell adhesion
GO:0007156 | homophilic cell adhesion via plasma membrane
GO:0008152 | metabolic process
GO:0009058 | biosynthetic process
GO:0009253 | peptidoglycan catabolic process
GO:0009405 | obsolete pathogenesis
GO:0016567 | protein ubiquitination
GO:0016998 | cell wall macromolecule catabolic process
GO:0030163 | protein catabolic process
GO:0032259 | methylation
GO:0045493 | xylan catabolic process
GO:0051258 | protein polymerization
GO:0055085 | transmembrane transport

Cellular Component
GO:0005576 | extracellular region
GO:0005615 | extracellular space
GO:0005618 | cell wall
GO:0005737 | cytoplasm
GO:0005886 | plasma membrane
GO:0016020 | membrane
GO:0016021 | integral component of membrane
GO:0032153 | cell division site

Molecular Function	
GO:0003676 | nucleic acid binding	GO:0008168 | methyltransferase activity
GO:0003677 | DNA binding	GO:0008199 | ferric iron binding
GO:0003678 | DNA helicase activity	GO:0008233 | peptidase activity
GO:0003697 | single-stranded DNA binding	GO:0008234 | cysteine-type peptidase activity
GO:0003700 | DNA-binding transcription factor activity	GO:0008237 | metallopeptidase activity
GO:0003796 | lysozyme activity	GO:0008270 | zinc ion binding
GO:0003824 | catalytic activity	GO:0009055 | electron transfer activity
GO:0003896 | DNA primase activity	GO:0016301 | kinase activity
GO:0003924 | GTPase activity	GO:0016491 | oxidoreductase activity
GO:0003993 | acid phosphatase activity	GO:0016740 | transferase activity
GO:0004129 | cytochrome-c oxidase activity	GO:0016746 | acyltransferase activity
GO:0004176 | ATP-dependent peptidase activity	GO:0016757 | glycosyltransferase activity
GO:0004180 | carboxypeptidase activity	GO:0016779 | nucleotidyltransferase activity
GO:0004197 | cysteine-type endopeptidase activity	GO:0016787 | hydrolase activity
GO:0004252 | serine-type endopeptidase activity	GO:0016788 | hydrolase activity, acting on ester bonds
GO:0004386 | helicase activity	GO:0016798 | hydrolase activity, acting on glycosyl bonds
GO:0004519 | endonuclease activity	GO:0016810 | hydrolase activity, acting on carbon-nitrogen bonds
GO:0004553 | hydrolase activity, hydrolyzing O-glycosyl compounds	GO:0016874 | ligase activity
GO:0004866 | endopeptidase inhibitor activity	GO:0016887 | ATP hydrolysis activity
GO:0004888 | transmembrane signaling receptor activity	GO:0030246 | carbohydrate binding
GO:0005096 | GTPase activator activity	GO:0031176 | endo-1,4-beta-xylanase activity
GO:0005506 | iron ion binding	GO:0043565 | sequence-specific DNA binding
GO:0005507 | copper ion binding	GO:0046872 | metal ion binding
GO:0005509 | calcium ion binding	GO:0051082 | unfolded protein binding
GO:0005524 | ATP binding	GO:0051536 | iron-sulfur cluster binding
GO:0005525 | GTP binding	GO:0051537 | 2 iron, 2 sulfur cluster binding
GO:0008061 | chitin binding	GO:0090729 | toxin activity

FIGURE 4 Shared GO terms of prion candidates across domains. The Venn diagram displays counts of GO terms in the same protein with cPrDs, including counts shared between domains (Table S7). The text depicts the 84 that are shared between all three domains in their given GO namespaces.

In all GO namespaces (molecular function, biological process, cellular component) we found a relatively low number of cPrPs detected in archaea (e.g., Figure 1). This could be explained by the low number of archaea proteomes available in the UniProt database, presumably due to difficulty in establishing laboratory conditions for culturing.

Within the GO molecular function namespace overlapping all three domains, we found 82 489 cPrPs: 533 in archaea, 3667 in bacteria, and 78 289 in eukaryotes (Table S7). There were 54 different overlapping GO molecular function annotations. Among them, the most common annotations were: serine-type endopeptidases (GO:0004252) 15 archeal, 273 bacterial, 244 eukaryotic; transcription factors (GO:0003700) 2 archeal, 8 bacterial, 7356 eukaryotic; single-stranded DNA binding proteins (GO:0003697) 1 archeal, 191 bacterial, 687 eukaryotic; kinases (GO:0016301) 1 archaeal, 49 bacterial, 569 eukaryotic.

Within the GO biological process namespace overlapping all three domains, we found 10 073 cPrPs: 254 in archaea, 1209 in bacteria, and 8610 in eukaryotes (Table S7). There were 22 different overlapping GO biological process annotations. Among them, the most common annotations were: protein ubiquitination (GO:0016567) 133 archaeal, 6 bacterial, 176 eukaryotic (many of the archaeal/bacterial proteins associated with this GO term were annotated with K07218, nitrous oxidase accessory protein; see Table S8); cell adhesion (GO:0007155) 38 archaeal, 259 bacterial, 180 eukaryotic; DNA replication (GO:0006260) 1 archaeal, 225 bacterial, 82 eukaryotic; DNA-templated regulation of transcription (GO:0006355) 6 archaeal, 18 bacterial, 5351 eukaryotic; DNA repair (GO:0006281) 6 archaeal, 127 bacterial, 1345 eukaryotic.

Within the GO cellular component namespace overlapping all three domains, we found 34 084 cPrPs: 924 in archaea, 6615 in bacteria, and 26 545 in eukaryotes (Table S7). There were 22 different overlapping GO cellular component annotations. Among them, the most common annotations were: integral component of membrane (GO:0016021) 842 archaeal, 5059 bacterial, 13 569 eukaryotic; plasma membrane (GO:0005886) 3 archaeal, 406 bacterial, 1915 eukaryotic; cell wall (GO:0005618) 3 archaeal, 49 bacterial, 170 eukaryotic.

3.3 | Overlap of enriched GO terms of prion candidates across different domains of life

We hypothesized that some functions associated with prions are conserved, along with prion domains, across different domains of life—allowing for the possibility that prion domains are essential to these specific GO-term functions. To pursue this hypothesis, we first focused on enriched GO term functions that were shared by cPrPs across all three domains of life. cPrPs with conserved associated functions across all three domains (if experimentally verified to be true prions) give slightly stronger credence to the possibility that prion domains may be truly conserved alongside specific protein functional domains (or a subset of specific functional domains) rather than being a product of convergent evolution. Though, importantly, convergent evolution is certainly still not eliminated as a possibility. To expand the search beyond just what is shared among all three domains, we also focused on what is shared just between Bacteria and Archaea, as these targets may still give us insight into distantly related functions associated with PLAAC-detected prion-like domains.

3.3.1 | Overlap of enriched GO terms assigned to prion candidates across three domains of life

We identified two molecular functions GO terms that were found enriched in all three domains: helicase activity (GO:0004386) and calcium ion binding (GO:0005509; Table 2, Figure 5).

TABLE 2 Numbers of candidates as a function of PLAAC CORE Score for enriched GO terms common to candidate prions in all three domains of life (with 42 being an arbitrary binning cutoff, see Section 2).

GO term	Domain	Numbers of candidates
GO:0005509 calcium ion binding	Archaea	193
	Bacteria	117
	Eukarya	2243
GO:0004386 helicase activity	Archaea	41
	Bacteria	206
	Eukarya	2814

Note: Table S9 holds PLAAC results and sequences for just these candidates.

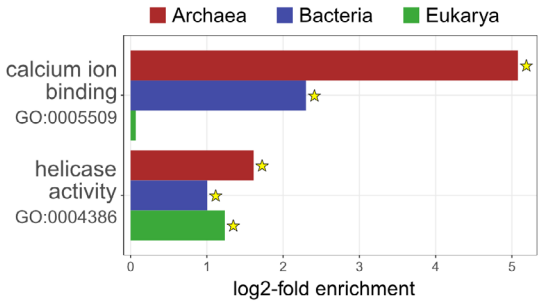


FIGURE 5 Log2-fold enrichment values for the two GO terms found to be enriched in cPrPs in all three domains. Those with stars had Benjamini–Hochberg adjusted-*p*-values ≤4E−8; for Eukarya, GO:0005509 had a BH adjusted *p*-value of .073.

We also looked at the KO annotations associated with cPrPs that held enriched GO annotations. KO annotations are more akin to whole protein-level annotations; while GO annotations are more akin to protein-domain level annotations. The group of prion candidates annotated with the calcium ion binding GO term (GO:0005509) was annotated with a variety of different KO functions across Archaea, Bacteria, and Eukarya, but the one consistent KO function that occurred in all three domains was K13974, a calcium-binding protein (Table S10; full KO annotation information for all cPrPs is available in our figshare repository linked in Section 2).

Among candidates annotated with helicase activity (GO:0004386), the majority of bacterial prion candidates were annotated as K03628, transcription termination factor Rho. Transcription termination factor Rho was originally identified as a cPrP by Iglesias et al.[74] and was later shown to form a self-perpetuating prion state in *Escherichia coli*.[68] For Archaea, the most common KO annotation within this group of proteins was K11927, ATP-dependent RNA helicase RhlE—which was also found among bacterial prion candidates. In Eukarya, multiple different helicases showed up as prion candidates including ATP-dependent RNA helicase DDX6/DHH1 (K12614), DDX5/DBP2 (K12823), and DDX3X (K11594; Table S10). All these helicases are good candidates for experimental verification of prion

properties. This observation is further strengthened by the fact that DDX5 was recently shown to form cytoplasmic aggregates in the brains of old killifish and mice. When studied in a yeast-based heterologous system, DDX5's prion-like domain allowed these aggregates to propagate across many generations.[98]

Prion proteins are known to be involved in stress response, and helicases are particularly good candidates for actuators of prion-based response as they can integrate diverse inputs and activate diverse outputs. Helicases provide immediate access to genetically complex traits by influencing the activity of many genes at the same time—in consequence enabling phenotypic variation that might be necessary for survival in a changing environment. The fact that helicases are one of only two groups of enriched functions of candidate prions shared by all three domains of life supports the idea that prion mechanisms may be essential to population-level survival and hence might be evolutionarily conserved.

Interestingly calcium ion binding prion candidates could also play a role in stress response but in a very different way than helicases. Ca^{2+} ion serves as a messenger, transmitting signals from the cell surface to its interior. One of the functions of calcium-binding proteins is to regulate the amount of free Ca^{2+} in the cytosol of the cell. Therefore, proteins that regulate the number of free Ca^{2+} are critical in the proper functioning of Ca^{2+} signaling. Many prions were shown to condensate in the form of gel droplets or amyloids.[99] This physical aggregation allows storing more protein in a given volume compared with its non-aggregated state. Therefore, aggregation of calcium ion binding protein might be an efficient way of sequestering intracellular calcium, influencing free Ca^{2+} levels influencing cell signaling and stress response via prion-like aggregation.

3.3.2 | Overlap of enriched GO terms of prion candidates between archaea and bacteria

As mentioned above, focusing on cPrD-associated functions that span all three domains helps support the notion that these are evolutionarily conserved across all of known life, but it also limits our scope. To expand our search, we next focused on what is shared just between Archaea and Bacteria, as these targets too may provide a window into distantly related but conserved functions associated with candidate prion domains (see Table 3). When we searched for common GO annotations in these groups, we found an overlap of GO terms in the biological processes namespace involved in adhesion, metabolic process, and protein modification (see Figure 6). Overlapping GO terms in the molecular function's namespace of prion candidates were in agreement with these above-listed GO biological process terms. For example, the carbohydrate-binding function corresponded to the adhesion process. Enzymatic functions corresponded to the metabolic process, and protein ubiquitination corresponded to protein modification. And, of course, as enriched GO terms for helicase activity and calcium ion binding were shared between all three domains, those are also shared between Bacteria and Archaea. According to our dataset, based on this approach, these processes and functions are the most conserved prion candidate functions in nature.

TABLE 3 Numbers of candidates for selected enriched GO terms common to Archaea and Bacteria.

GO term	Domain	Numbers of candidates
GO:0016567 protein ubiquitination	Archaea	133
	Bacteria	6
GO:0045493 xylan catabolic process	Archaea	6
	Bacteria	85
GO:0007156 homophilic cell adhesion	Archaea	6
	Bacteria	19
GO:0031176 endo-1,4-beta-xylanase	Archaea	6
	Bacteria	77
GO:0004197 cysteine-type endopeptidase	Archaea	3
	Bacteria	6
GO:0004252 serine-type endopeptidase	Archaea	15
	Bacteria	273

Note: Table S11 holds PLAAC results and sequences for just these candidates.

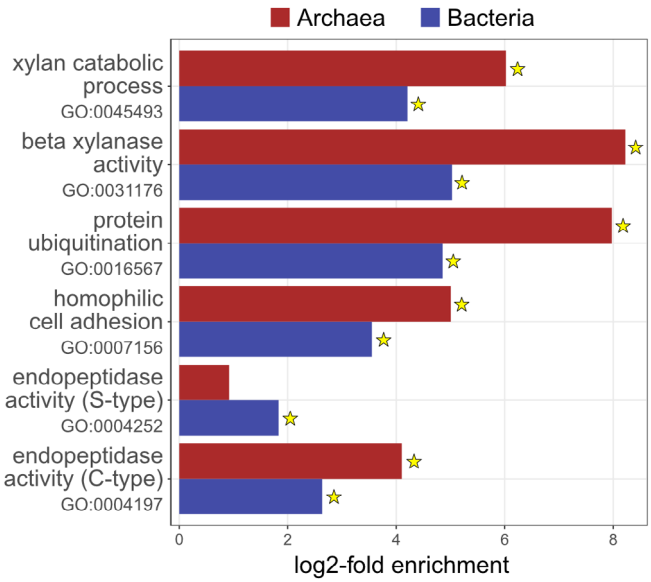


FIGURE 6 Log2-fold enrichment values for the two GO terms found to be enriched in cPrPs in both Bacteria and Archaea domains as compared with all proteins from those domains. Those with stars had Benjamini-Hochberg adjusted-*p*-values ≤.012; for Archaea, GO:0004252 had a BH adjusted *p*-value of .19. "S-type" refers to serine-type; "C-type" refers to cysteine-type.

3.3.3 | Organisms harboring enriched GO terms that overlap across all three domains of life

After identifying enriched GO terms that overlap across all three domains of life, we wanted to see if there were any trends in the distribution of candidate prions associated with these specific GO terms with regard to the phylogenetic relationships of the organisms containing them. Presumably, it is possible that a candidate prion domain (cPrD) being consistently associated with a specific GO term is either

a consequence of that organism's evolutionary history (which would present largely as monophyletic phylogenetic clades of organisms possessing the protein holding cPrD and GO term), horizontal gene transfer, or it could be due to independently evolved, yet common, characteristics (which, as with horizontal gene transfer, might present as a highly polyphyletic distribution).

For both the GO annotations calcium ion binding (GO:0005509) and helicase activity (GO:0004386) in Archaea, most candidates were detected within one monophyletic clade of Euryarchaeota, with a few others spread around members of the DPANN group and within the Thaumarchaeota (Figures S1 and S2). Considering the possibility of the ancient origin of prions, one might expect to discover cPrPs annotated with these two GO terms within the superphylum Asgard-suspected progenitors of Eukaryota.[100] The lack of candidate prions annotated with enriched GO terms overlapping all three domains (calcium ion binding and helicases) in Asgard archaea might be explained by the fact that this group consists mostly of uncultured and relatively understudied organisms, which lowers their representation in standard databases. Indeed, the UniProt reference proteome database we were working with here only included 16 from the Asgard group, all belonging to the candidate phylum Lokiarchaeota (Table S1). There is also the potential that the retained protein similarity of potentially homologous functions was too divergent to be functionally annotated with the same GO term.

Distribution patterns in Bacteria and Eukarya were similar for both of the conserved, enriched GO terms (calcium ion binding [GO:0005509] and helicase activity [GO:0004386]). In bacteria, we found prion candidates with associated calcium ion binding annotations in all major phyla, with a higher concentration of calcium ion binding within the Alphaproteobacteria and a higher concentration of helicase activity within Actinobacteria (Figures S1 and S2). And in Eukarya, candidates with both GO terms are found roughly throughout the entire tree, with the exception being within the Chordata (Figures S1 and S2).

3.4 | Overlap of GO terms of prion candidates across all three domains of life

Next, we identified GO annotations of prion candidates common to different domains of life, whether they were statistically enriched in prion candidates over all proteins or not (meaning, now no longer focused on "enriched" as was done above). Having an enriched molecular function, biological process, or cellular component (the GO namespaces) implies that a prion candidate is more likely to be associated with a certain GO term as compared with a protein that is not a prion candidate. That view helps focus on GO terms that seem to be only found in prion candidates, but it would miss those that can be with or without a candidate prion domain. But even without looking at enrichment, the proteins identified by PLAAC are valid prion candidates. If their GO annotations are found to be common across all three domains of life, they may still represent some of the oldest prions on Earth (again, assuming that some of these candidates would be verified experimentally to behave as prions, as we have seen before; e.g., Reference 66). In this section, we focused on such prion candidates whose GO terms were not necessarily enriched but were detected in all three domains of life.

3.4.1 | Overlap of molecular function GO terms of prion candidates across all three domains of life

Among GO molecular functions of cPrPs that overlap across all three domains of life, the best represented in our dataset were transcription factors, kinases, DNA binding, peptidases, ribonucleotide binding, and metal binding (Figure 4; Table S7).

One prime example of this is DNA-binding transcription factor activity (GO:0003700). Transcription regulation is commonly implicated with prion biology of confirmed prion proteins, such as Ure 2—[*URE3*$^{+}$] prion,[101] Mot3—[*MOT3*$^{+}$] prion,[45] and Sfp1—[*ISP*$^{+}$] prion.[102,103] Binding to nucleic acids is one of the most fundamental activities of life and plays a role in each step of the central dogma of biology. When we looked at KO annotations of prion candidates that fall under GO:0003700, we found that archaea and eukaryotes share the KO term K21042, "HCMV protein UL11" (Table S12), which is also present in viruses. The candidate among archaea belongs to the genus *Thermoproteus*, a thermophile whose protein contains the marR-type HTH domain, which is responsible for antibiotic resistance. Production of this protein may be a response to the presence of organisms in the environment that produce antibiotic-like substances (e.g., fungi). The same protein of the UL11 type in eukaryotes has the FHA domain that takes part in an ancient and widespread mechanism of regulation based on the phospho-dependent assembly of protein complexes.[104] In bacteria, prion candidates annotated as DNA-binding transcription factors (GO:0003700) consisted mostly of RNA polymerase sigma-70 factor, ECF subfamily (K12888), which is the major factor that initiates transcription in bacteria.

Single-stranded DNA binding (GO:0003697) has been identified as common among prions not only in our dataset of prion candidates but also by Harrison.[105] Single-strand binding proteins are exceptionally important for maintaining the stability of an organism's genome, as they are involved in key processes taking place in the nucleus, such as DNA replication, repair, and recombination. Among single-stranded DNA binding cPrPs, we found 190 annotated as K03111, single-strand DNA-binding protein, but only in bacteria. In eukaryotes, we identified three other KO descriptions: K21390, adhesion defective protein 2, which is responsible for the transcriptional regulation of cell adhesion; K12888, heterogeneous nuclear ribonucleoprotein U, intra-nuclear proteins that take part in many processes of the nucleus metabolic pathway including organization of chromatin, regulation of telomere length, transcription, and alternative mRNA splicing; and K13184, ATP-dependent RNA helicase A, a multifunctional protein involved in processes such as DNA replication, post-transcriptional regulation of RNA, mRNA translation, and silencing. All KO term groups identified single-stranded DNA binding proteins that play a particularly important role in keeping DNA and RNA functioning

properly. As for the archaea, the only prion candidate from this GO did not have any KO term assigned.

Kinase activity (GO:0016301) was another GO annotation that overlapped all three domains. Within this GO annotation, we found diverse kinase-related proteins based on their KO annotations, including activators of two-component systems (e.g., K20340), serine/threonine kinase activators (e.g., K08286), and activators of tyrosine kinases (eK23453)—with annotations virtually only being ascribed to eukarya (Table S12). These have a diverse repertoire of functions ranging from regulating cell death, cell migration, and cell adhesion, to general transcriptional repression related to circadian rhythm.

Another group of cPrPs that overlapped all three domains were serine-type endopeptidases (GO:0004252). Peptidases are active enzymes that cleave peptide bonds in proteins and peptides by hydrolysis, and serine-type endopeptidases fall into a class of peptidases that are characterized by the presence of a serine residue in the active site of the enzyme. They are of extremely widespread occurrence from prokaryotes to vertebrates and exhibit diverse functions.[106] When we looked at KO annotations of prion candidates that fall under the serine-type endopeptidases (GO:0004252) in bacteria. we found many annotated as K08372—putative serine protease PepD (Table S12). Most PepD peptidases have PDZ domains that recognize and process misfolded proteins at the cell membrane, leading to the activation of signaling pathways and the establishment of a feedback loop that can facilitate bacterial adaptation.[107,108] This observation is consistent with the hypothesis that prion formation might be an ancient mechanism facilitating adaptation. Among endopeptidases in archaea, we found ATP-dependent Lon protease. This protease is found both in mitochondria and bacteria and shares high similarities between the organelle and the domain.[109] Among eukaryotic prion candidates annotated with the GO term serine-type endopeptidases (GO:0004252), we found rhomboid proteases (e.g., K19225; Table S12) that are common in all domains of life and are implicated in various functions including cell signaling, quorum sensing, and homeostasis.[110] Recently, Rhomboid Protease RHBDL4 was shown to cleave amyloid precursor protein (APP) a key molecule in the etiology of Alzheimer's disease.[111-113]

Another GO description overlapping all three domains of life was ribonucleotide binding (GO:0032553), which, similarly to the DNA-binding transcription factor activity (GO:0003700), is central to biology because ribonucleotides are primary sources of energy for biochemical reactions. Ribonucleotide binding is a parent term of four GO annotations that also was associated with candidate prions in our dataset in all three domains: ATP binding (GO:0005524), ATP hydrolysis activity (GO:0016887), GTP binding (GO:0005525), and GTPase activity (GO:0003924).

The last major group of annotations common to all domains (regardless of enrichment) identified in our dataset was annotated as metal-binding proteins (GO:0046872), which play essential roles in a wide range of structural and catalytic functions. Similar to the above-mentioned functions, these are also central to all biology. Child terms of GO:0046872 that are also common to all domains were zinc ion binding (GO:0008270; 1 archeal, 94 bacterial, 10 781 eukaryotic), iron ion binding (GO:0005506; 1 archeal, 9 bacterial, 38 eukaryotic), copper ion binding (GO:0005507; 25 archeal, 17 bacterial, 88 eukaryotic), and ferric iron-binding (GO:0008199; 1 archeal, 1 bacterial, 1 eukaryotic). KO annotations for the cPrPs holding these shared functions were highly varied (Table S12).

3.4.2 | Overlap of biological process GO terms of prion candidates across all three domains of life

When we analyzed GO terms within the biological process namespace that overlapped across three domains of life, we found that for archaea, the greatest number of prion candidates were annotated with protein ubiquitination (GO:0016567; 133 proteins; Table S7). For bacteria, two GO terms were more common than others: cell adhesion (GO:0007155; 259 proteins) and DNA replication (GO:0006260; 225 proteins). For eukaryotes, the greatest number of prion candidates were annotated as DNA-templated regulation of transcription (GO:0006355; 5351 proteins; Table S7).

Another GO annotation brought to our attention was DNA repair (GO:0006281), which was previously noted as an abundant biological process among candidate prions by Iglesias et al.[74] This annotation was clustered by these authors in the group stimulus–response process with some candidates annotated specifically as a cellular response to DNA damage stimulus (GO:0006974).

Other GO terms that were detected in our analysis as well as others,[74] were grouped under a common category of invasion and virulence (GO:0009405, obsolete pathogenesis, and GO:0000272 polysaccharides catabolic process), and under another broad category of nucleotide metabolism (GO:0006353, DNA-templated transcription termination and GO:0006260, DNA replication).

One of the major groups of prion candidates identified by Iglesias et al.[74] was annotated as GO:0009056, catabolic process. Many catabolic processes were also present in our dataset: GO:0000272, polysaccharide catabolic process; GO:0006308, DNA catabolic process; GO:0045493, xylan catabolic process; GO:0030163, protein catabolic process; GO:0000272, polysaccharide catabolic process; GO:0009253, peptidoglycan catabolic process; and GO:0016998, cell wall macromolecule catabolic process. The last two GO annotations (GO:0009253 and GO:0016998) have also been described as common annotations of bacterial prion candidates by Harrison.[105]

A high number of candidates in archaea and bacteria candidates were annotated as GO:0007155 cell adhesion (38 archaeal, 259 bacterial), which might be related to the fact that both groups of organisms often form biofilms.

3.4.3 | Overlap of cellular component GO terms of prion candidates across all three domains of life

In our dataset the majority of cPrPs were annotated as an integral component of membranes (GO:0016021). The localization of multiple cPrPs on the periphery of the cell was noticed by other authors as

well.[64,105] In addition to those, we found a surprisingly high number of cPrPs annotated as proteins occupying extracellular regions—not attached to the cell surface (GO:0005576). So far as we are aware, no extracellular proteins are known to form prions. A large number of prion candidates having been annotated as extracellular proteins correlates with the fact that PLAAC-predicted prion-like domains are often found to aggregate as amyloids. Amyloid fibrils are often found in the extracellular matrix of bacteria,[114] contributing to biofilm formation. It has even been suggested that this (meta) function of amyloids may have arisen many times independently during the evolution

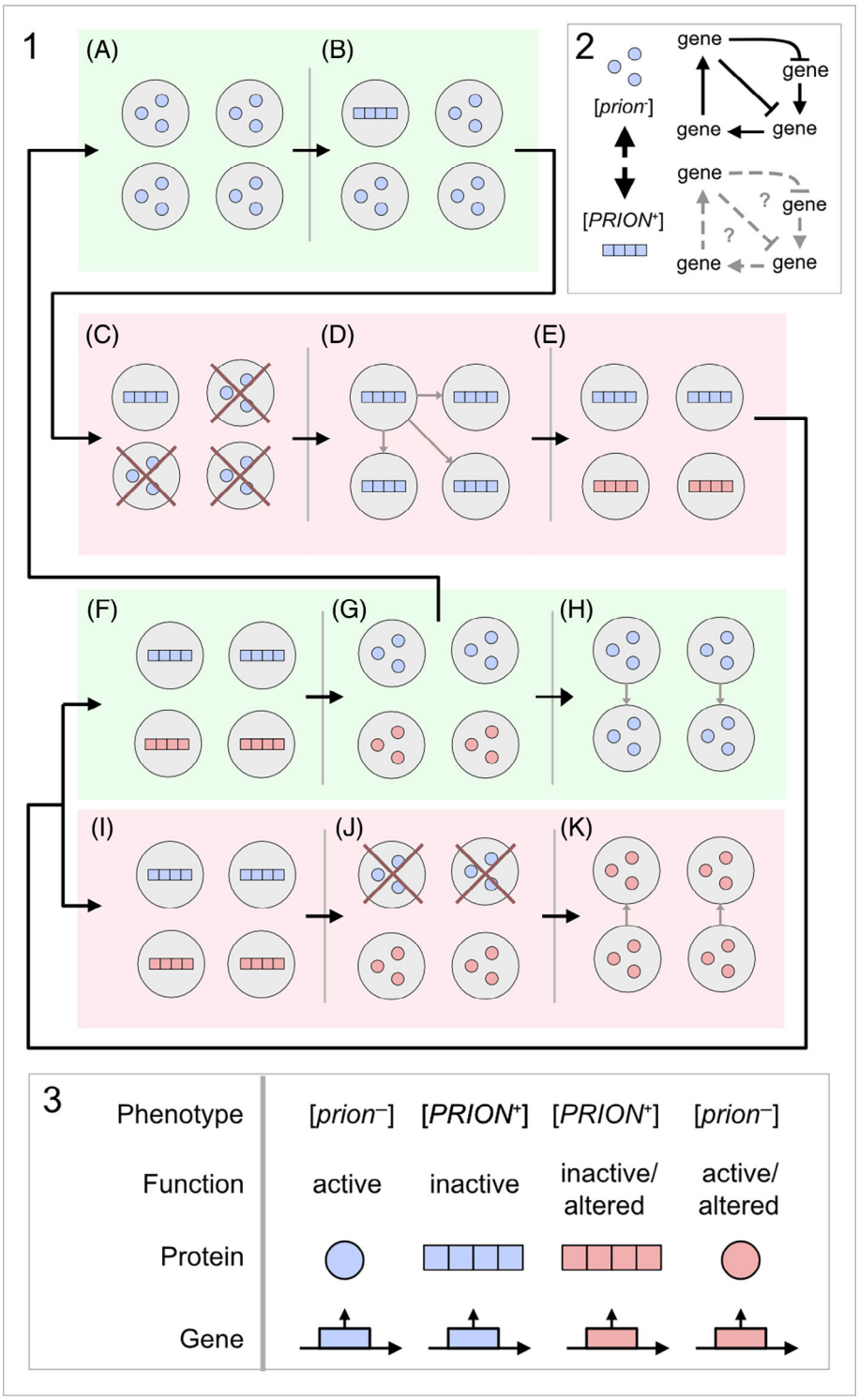


FIGURE 7 Legend on next page.

of bacteria.[115] Interestingly, some extracellular proteins were also found to form amyloid fibrils in archaea,[116] adding to the universality of this function.

3.5 | Could ancient prions facilitate adaptation?

Prions have been hypothesized to facilitate the generation of diverse responses to changes in environmental or cellular conditions by regulating gene expression patterns.[117] In yeast, prion [*GAR*$^+$] regulates whether cells are metabolic specialists or generalists,[15] prion [*SMAUG*$^+$] regulates sporulation,[118] and prion [*ESI*$^+$] influences the activity of multiple genes by regulating the expression of subtelomeric regions in response to environmental stresses such as antifungal drugs.[26] In Bacteria, prion-aggregation of protein Rho influences the expression of hundreds of genes, dramatically influencing cell phenotypes, potentially rendering prion-harboring cells better suited to rapid changes in the environment.[68] Some of the functions of cPrPs identified in our dataset that overlap across all three domains of life (namely transcription factors, DNA binding, and peptidases) are particularly well suited to generating diverse responses to changes in their environment. We suspect that these potential prions can provide a rapid means for organisms to explore the evolutionary landscape, and, given their presence across all three domains, they may be some of the most ancient functions associated with prions (if experimentally verified).

Perhaps one of the most efficient ways of generating diverse responses to changes in environmental or cellular conditions is by changing signal response and regulation pathways. Candidate prions with associated kinase-related functions are therefore intriguing to be found across all three domains of life in our dataset (Figure 4; GO:0016301). Generally, the potential role of candidate prions in signal transduction and regulation suggests they may have played a critical role in the evolution of some of the complex pathways that have allowed life to expand into a wide range of environmental niches. Indeed, a question that has plagued the study of kinases is how such complexity could have evolved and how orthologous signaling proteins diverge.[119] Suspending the pressures of natural selection acting on a protein for several generations following a duplication event would allow for a sufficient number of step-wise mutations to accumulate, leading to the large mutational jumps required for preventing cross-talk between newly diverged kinases and their ancestors. Because prion evolution can, in part, be dictated by both DNA and protein heritability,[120] they may also provide a partial solution to the question of how newly evolved signaling proteins prevent cross-talk with their very similar ancestors. If a duplicated signaling protein becomes functionally inert due to being aggregated into fibrils in a prionic form, a larger number of mutations are allowed to accumulate, independent of selection. Because these prionic forms can be inherited by daughter cells via the phenomenon of seeding, this state of non-selective mutagenesis can carry on in daughter cells for many generations before the aggregated prion state ends and function is restored. This mechanism allows for large mutational "jumps" that could lead to large enough divergence to prevent cross-talk with orthologous ancestral signaling proteins—being one example of how proteins could accumulate many step-wise mutations without selection acting on individual mutations along the way (see Figure 7). Mechanisms for the accelerated evolution of kinase proteins have been theorized as ways of preventing cross-talk, such as diversity-generating retroelements specific to the cyanobacterial phylum.[121] However, these retroelements have only been found in viruses and prokaryotes, where they usually target genes encoding for proteins involved in cell–cell or phage–cell attachment. Cyanobacteria are the only phylum where they have been found to target kinase genes specifically. Candidate prionic kinases, being found in all domains of life, are thus an intriguing potential mechanism for circumventing stepwise mutation. Furthermore, the kind of conformational switching that the prion state provides proteins allows for multiple activity states from the same polypeptide sequence, providing greater efficiency and economy for the proteome.

FIGURE 7 A hypothesized evolutionarily advantageous, prion-based mechanism of adaptation to new environments. **1**—(A) Population of cells in a given environment (green background). (B) Spontaneous acquisition of prion phenotype. Squares represent amyloid fibrils. (C) Change of the environment (red background). Survival of the population is possible because it is phenotypically heterogeneous. The phenomenon is known as biological "bet-hedging"—discussed in the text. In the new environment, the prion phenotype undergoes positive selection. (D) Reversion to non-prion phenotype is counter-selected. Prion phenotype dominates the population. (E) Because the protein product is locked in prion aggregate (inactive in this example), the gene coding for prion protein experiences less evolutionary pressure, and more mutations are tolerated. As a result, the population diverges. There are now two genetically different but phenotypically identical populations of cells. The diversity is at the level of genes, but for simplification, we marked the aggregates in two different colors. (F) The environment reverses to the original state. Prion phenotype no longer experiences environmental selection and, with time, will be reduced. (G) Prion is lost due to natural selection, and phenotype reverses to one that depends on the active form of the prion protein. The mutations acquired in point F can either experience negative selection pressure or, if benevolent, will help adapt to the new environment when encountered again in the future. (H) The original population is restored back to the state from point A. (I) An alternative scenario in which the environment does not reverse to the original state and selection pressure on prion phenotype continues. (J) Reversion to non-prion phenotype is still counter-selected, but the extended time during which new mutations can accumulate eventually leads to the emergence of protein that facilitates survival in the new environment—even when not aggregated in the form of prion. (K) Eventually, the population survives even when the prion phenotype is lost. The population is now adapted to the new environment. **2**—Prion conversion influences regulatory gene cascades, influencing the expression of many genes at once, providing a rapid means for organisms to explore the evolutionary landscape. **3**—Relationship between gene, protein, function, and phenotype explained with symbols used in panel 1.

Signal response is evolutionarily ancient as it is essential for multicellularity and would even be needed in LUCA to mount a simple metabolic response to changes in nutrients or increases in environmental toxicity (e.g., metal concentration, pH, etc.). While the role of prions in allowing for the greater divergence of signaling proteins remains to be explored, it is clear that they are important for this evolutionarily ancient mechanism of species adaptation.

4 | CONCLUSIONS

In proteomes of organisms representing three different domains of life, we identified cPrDs and then analyzed the functional annotations associated with the proteins harboring them. Helicases and calcium ion binding proteins were enriched among prion protein candidates in all three domains of life—meaning these specific GO terms were more likely to be associated with a protein with an identified candidate prion domain than in a protein without one. Beyond just those enriched in candidate prions, we identified numerous functional annotations that are associated with candidate prions in all three domains. Some of the most represented included peptidases, transcription factors, single-stranded DNA binding proteins, and kinases; functions that are fundamental to the proper functioning of cells.

The role of prions in responses to changes in environmental conditions has been described as a process called biological "bet-hedging" that can facilitate the survival of the resultant phenotypically heterogeneous population.[122,123] In this model, the accumulation of prionic forms of proteins provides some phenotypic change that allows the cell, and in turn, some subset of the population, to potentially be better suited to a new environment. Aggregation of this prion form of the protein also typically renders the protein functionally inert with regard to its previous, more traditionally understood function, which could then allow for random genetic mutagenesis to occur more frequently, independent of selective pressures it otherwise would have faced. In these cases, due to chance, selection may then favor mutated progeny even once it is no longer experiencing prion aggregation (Figure 7). This mechanism could allow for adaptation and population expansion into more diverse environments and/or the functional expansion of proteins involved in essential processes such as signal transduction and regulation.

The exciting possibility that these functions are also subjected to regulation through prion formation remains to be verified experimentally. Many of the proteins identified in our study had high COREscore values, indicative of high compositional similarity to known yeast prions. Confirming that these proteins can form prions would indicate that at least some essential protein functions are accompanied by prion domains across great evolutionary distances. If that turns out to be true, the role of prions in the regulation of the fundamental cell processes could be an evolutionarily ancient one, even if individual prion domains themselves are not evolutionarily conserved.

Based on our results, we hypothesize that prions that can influence the expression of many genes at once provide a rapid means for organisms to explore the evolutionary landscape—and, if true, this might be the most ancient function of prions conserved across all domains of life.

AUTHOR CONTRIBUTIONS

Tomasz Zajkowski: Conceptualization; investigation; funding acquisition; writing – original draft; methodology; validation; visualization; writing – review and editing; formal analysis; project administration; supervision; resources. **Michael D. Lee:** Conceptualization; investigation; writing – original draft; methodology; validation; visualization; writing – review and editing; software; formal analysis; data curation; supervision; resources. **Siddhant Sharma:** Conceptualization; investigation; writing – original draft; methodology; validation; visualization; writing – review and editing; software; formal analysis; data curation; resources. **Alec Vallota-Eastman:** Investigation; writing – original draft; writing – review and editing. **Mikołaj Kuska:** Writing – original draft; investigation; writing – review and editing. **Małgorzata Malczewska:** Investigation; writing – original draft; writing – review and editing. **Lynn J. Rothschild:** Conceptualization; funding acquisition; writing – original draft; validation; writing – review and editing; formal analysis; project administration; supervision; resources.

ACKNOWLEDGMENTS

We acknowledge Blue Marble Space Institute of Science (BMSIS) for organizational support. We thank Doctor Daniel Jarosz for his insightful comments during manuscript preparation. We thank the Polish Astrobiology Society and members of the Prion Task Force (www.astrobio.pl/prion), especially Magdalena Pilska-Piotrowska for help in developing the concept of domesticated amyloids and prions. S.S. would like to acknowledge the SETI Forward 2020 from the SETI Institute. Tomasz Zajkowski and Lynn J. Rothschild are sponsored by the National Aeronautics and Space Administration's Planetary Science Division Research Program and Ames Research Innovation Award (ARIA). Alec Vallota-Eastman is supported by the National Science Foundation under Grant No. INTERN NSF DCL 21-013.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/prot.26558.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Figshare at https://figshare.com/projects/Zajkowski_et_al_2022_3-domain_prion_data_and_code_repository/133155.

ORCID

Tomasz Zajkowski https://orcid.org/0000-0001-9433-6071
Michael D. Lee https://orcid.org/0000-0001-7750-9145
Siddhant Sharma https://orcid.org/0000-0003-1768-1802
Alec Vallota-Eastman https://orcid.org/0000-0002-6453-4680

Mikołaj Kuska https://orcid.org/0000-0002-7568-1926

Małgorzata Malczewska https://orcid.org/0000-0002-3610-4632

Lynn J. Rothschild https://orcid.org/0000-0003-2970-8048

REFERENCES

1. Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*. 2006;75:333-366.
2. Bratkovič H. Prions, prionoid complexes and amyloids: the bad, the good and something in between. *Swiss Med Wkly*. 2017;147: w14424.
3. Inge-Vechtomov S, Zhouravleva G, Chernoff YO. Biological roles of prion domains. *Prion*. 2007;1:228-235. doi:10.4161/pri.1.4.5059
4. Gazit E. Self assembly of short aromatic peptides into amyloid fibrils and related nanostructures. *Prion*. 2007;1(1):32-35.
5. Rufo CM, Moroz YS, Moroz OV, et al. Short peptides self-assemble to produce catalytic amyloids. *Nat Chem*. 2014;6(4):303-309.
6. Dobson CM. Protein folding and misfolding. *Nature*. 2003;426:884-890. doi:10.1038/nature02261
7. Jahn TR, Radford SE. The Yin and Yang of protein folding. *FEBS J*. 2005;272:5962-5970. doi:10.1111/j.1742-4658.2005.05021.x
8. Yoshimura Y, Lin Y, Yagi H, et al. Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. *Proc Natl Acad Sci U S A*. 2012;109(36):14446-14451.
9. Shi Y, Li X, Ding G, Yangjiang W, Weng Y, Zhijun H. Control of β-sheet crystal orientation and elastic modulus in silk protein by nanoconfinement. *Macromolecules*. 2014;47:7987-7992. doi:10.1021/ma501864g
10. Slotta U, Hess S, Spiess K, Stromer T, Serpell L, Scheibel T. Spider silk and amyloid fibrils: a structural comparison. *Macromol Biosci*. 2007;7(2):183-188.
11. Bissig C, Rochin L, van Niel G. PMEL amyloid fibril formation: the bright steps of pigmentation. *Int J Mol Sci*. 2016;17(9):1438. doi:10.3390/ijms17091438
12. Beyersdörfer T. Curli expression and biofilm formation by *Escherichia coli* isolates. 2019.
13. Garcia MC, Lee JT, Ramsook CB, Alsteens D, Dufrêne YF, Lipke PN. A role for amyloid in cell aggregation and biofilm formation. *PLoS One*. 2011;6(3):e17632.
14. Sabaté R, Gallardo M, Estelrich J. An autocatalytic reaction as a model for the kinetics of the aggregation of β-amyloid. *Biopolymers*. 2003;71:190-195. doi:10.1002/bip.10441
15. Brown JCS, Lindquist S. A heritable switch in carbon source utilization driven by an unusual yeast prion. *Genes Dev*. 2009;23(19):2320-2332.
16. Halfmann R, Lindquist S. Epigenetics in the extreme: prions and the inheritance of environmentally acquired traits. *Science*. 2010; 330(6004):629-632.
17. Jarosz DF, Khurana V. Specification of physiologic and disease states by distinct proteins and protein conformations. *Cell*. 2017;171(5): 1001-1014.
18. Wickner RB, Edskes HK, Shewmaker F. How to find a prion: [URE3], [PSI+] and [beta]. *Methods*. 2006;39(1):3-8.
19. Chakrabortee S, Byers JS, Jones S, et al. Intrinsically disordered proteins drive emergence and inheritance of biological traits. *Cell*. 2016; 167(2):369-381.
20. Dennis EM, Garcia DM. Biochemical principles in prion-based inheritance. *Epigenomes*. 2022;6(1):4.
21. Krammer C, Schätzl HM, Vorberg I. Prion-like propagation of cytosolic protein aggregates: insights from cell culture models. *Prion*. 2009;3(4):206-212.
22. Chakrabortee S, Kayatekin C, Newby GA, Mendillo ML, Lancaster A, Lindquist S. Luminidependens (LD) is an Arabidopsis protein with prion behavior. *Proc Natl Acad Sci U S A*. 2016;113(21):6065-6070.
23. Pallarès I, Ventura S. The transcription terminator rho: a first bacterial prion. *Trends Microbiol*. 2017;25(6):434-437.
24. Molina-García L, Gasset-Rosa F, Álamo M M-d, et al. Functional amyloids as inhibitors of plasmid DNA replication. *Sci Rep*. 2016; 6(1):1-8.
25. Goncharoff DK, Zhiqiang D, Li L. A brief overview of the Swi1 prion-[SWI+]. *FEMS Yeast Res*. 2018;18(6):foy061. doi:10.1093/femsyr/foy061
26. Harvey ZH, Chakravarty AK, Futia RA, Jarosz DF. A prion epigenetic switch establishes an active chromatin state. *Cell*. 2020;180(5): 928-940.
27. Li L, Lindquist S. Creating a protein-based element of inheritance. *Science*. 2000;287(5453):661-664.
28. Gomes E, Shorter J. The molecular language of membraneless organelles. *J Biol Chem*. 2019;294(18):7115-7127.
29. Franzmann TM, Alberti S. Prion-like low-complexity sequences: key regulators of protein solubility and phase behavior. *J Biol Chem*. 2019;294(18):7128-7136.
30. Ba N, Alex N, Yeh BJ, et al. Proteome-wide discovery of evolutionary conserved sequences in disordered regions. *Sci Signal*. 2012; 5(215):rs1.
31. Wright PE, Jane Dyson H. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol*. 2015;16(1):18-29.
32. Toombs JA, McCarty BR, Ross ED. Compositional determinants of prion formation in yeast. *Mol Cell Biol*. 2010;30(1):319-332.
33. Chandramowlishwaran P, Sun M, Casey KL, et al. Mammalian amyloidogenic proteins promote prion nucleation in yeast. *J Biol Chem*. 2018;293(9):3436-3450.
34. Taneja V, Maddelein M-L, Talarek N, Saupe SJ, Liebman SW. A non-Q/N-rich prion domain of a foreign prion, [Het-S], can propagate as a prion in yeast. *Mol Cell*. 2007;27(1):67-77.
35. Suzuki G, Tanaka M. Expanding the yeast prion world: active prion conversion of non-glutamine/asparagine-rich Mod5 for cell survival. *Prion*. 2013;7(2):109-113.
36. Ehsani S, Tao R, Pocanschi CL, Ren H, Harrison PM, Schmitt-Ulms G. Evidence for retrogene origins of the prion gene family. *PLoS One*. 2011;6(10):e26800.
37. Holmes DL, Lancaster AK, Lindquist S, Halfmann R. Heritable remodeling of yeast multicellularity by an environmentally responsive prion. *Cell*. 2013;153(1):153-165.
38. Westaway D, Daude N, Wohlgemuth S, Harrison P. The PrP-like proteins Shadoo and Doppel. *Top Curr Chem*. 2011;305:225-256.
39. Cloutier R, Clement AM, Lee MSY, et al. Elpistostege and the origin of the vertebrate hand. *Nature*. 2020;579(7800):549-554.
40. Harrison LB, Zhan Y, Stajich JE, Dietrich FS, Harrison PM. Evolution of budding yeast prion-determinant sequences across diverse fungi. *J Mol Biol*. 2007;368(1):273-282.
41. An L, Harrison PM. The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. *Biol Direct*. 2016; 11:32.
42. Su T-Y, Harrison PM. Conservation of prion-like composition and sequence in prion-formers and prion-like proteins of saccharomyces cerevisiae. *Front Mol Biosci*. 2019;6:54. doi:10.3389/fmolb.2019.00054
43. Ross ED, Baxa U, Wickner RB. Scrambled prion domains form prions and amyloid. *Mol Cell Biol*. 2004;24(16):7206-7213.
44. Ross ED, Edskes HK, Terry MJ, Wickner RB. Primary sequence independence for prion formation. *Proc Natl Acad Sci U S A*. 2005; 102(36):12825-12830.
45. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell*. 2009;137(1):146-158.
46. Buchan JR, Muhlrad D, Parker R. P bodies promote stress granule assembly in *Saccharomyces cerevisiae*. *J Cell Biol*. 2008;183(3): 441-455.

47. Halfmann R, Wright JR, Alberti S, Lindquist S, Rexach M. Prion formation by a yeast GLFG nucleoporin. *Prion*. 2012;6(4):391-399.
48. Cieśla M, Makała E, Płonka M, et al. Rbs1, a new protein implicated in RNA polymerase III biogenesis in yeast saccharomyces cerevisiae. *Mol Cell Biol*. 2015;35(7):1169-1181.
49. Bourbon H-M. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Res*. 2008;36(12):3993-4008.
50. Harrison AF, Shorter J. RNA-binding proteins with prion-like domains in health and disease. *Biochem J*. 2017;474(8):1417-1438.
51. Nizhnikov AA, Antonets KS, Bondarev SA, Inge-Vechtomov SG, Derkatch IL. Prions, amyloids, and RNA: pieces of a puzzle. *Prion*. 2016;10(3):182-206.
52. Malinovska L, Kroschwald S, Alberti S. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim Biophys Acta*. 2013;1834(5):918-931.
53. Kato M, Han TW, Xie S, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell*. 2012;149(4):753-767.
54. Kroschwald S, Maharana S, Mateju D, et al. Promiscuous interactions and protein Disaggregases determine the material state of stress-inducible RNP granules. *elife*. 2015;4:e06807.
55. Minhas A, Amir FU, Ross ED, Ben-Hur A. Amino acid composition predicts prion activity. *PLoS Comput Biol*. 2017;13(4):e1005465.
56. Cascarina SM, Ross ED. Natural and pathogenic protein sequence variation affecting prion-like domains within and across human proteomes. *BMC Genomics*. 2020;21(1):1-18.
57. Angarica E, Vladimir SV, Sancho J. Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. *BMC Genomics*. 2013;14:316.
58. Harrison PM, Gerstein M. A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes. *Genome Biol*. 2003;4(6):R40.
59. Lancaster AK, Nutter-Upham A, Lindquist S, King OD. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics*. 2014;30(17):2501-2502.
60. Michelitsch MD, Weissman JS. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci U S A*. 2000;97(22):11910-11915.
61. Ross ED, Maclea KS, Anderson C, Ben-Hur A. A bioinformatics method for identifying Q/N-rich prion-like domains in proteins. *Methods Mol Biol*. 2013;1017:219-228.
62. Sabate R, Rousseau F, Schymkowitz J, Ventura S. What makes a protein sequence a prion? *PLoS Comput Biol*. 2015;11(1):e1004013.
63. Zambrano R, Conchillo-Sole O, Iglesias V, et al. PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Res*. 2015;43(W1):W331-W337.
64. Batlle C, Iglesias V, Navarro S, Ventura S. Prion-like proteins and their computational identification in proteomes. *Expert Rev Proteomics*. 2017;14(4):335-350.
65. Gil-Garcia M, Iglesias V, Pallarès I, Ventura S. Prion-like proteins: from computational approaches to proteome-wide analysis. *FEBS Open Bio*. 2021;11(9):2400-2417.
66. Zajkowski T, Lee MD, Mondal SS, et al. The hunt for ancient prions: archaeal prion-like domains form amyloid-based epigenetic elements. *Mol Biol Evol*. 2021;38(5):2088-2103.
67. Fleming E, Yuan AH, Heller DM, Hochschild A. A bacteria-based genetic assay detects prion formation. *Proc Natl Acad Sci U S A*. 2019;116(10):4605-4610.
68. Yuan AH, Hochschild A. A bacterial global regulator forms a prion. *Science*. 2017;355(6321):198-201.
69. Nan H, Chen H, Tuite MF, Xiaodong X. A viral expression factor behaves as a prion. *Nat Commun*. 2019;10(1):1-11.
70. Tetz G, Tetz V. Prion-like domains in phagobiota. *Front Microbiol*. 2017;8:2239.
71. Tetz G, Tetz V. Prion-like domains in eukaryotic viruses. *Sci Rep*. 2018;8(1):8931.
72. Angarica E, Vladimir AA, Giner A, Losilla G, Ventura S, Sancho J. PrionScan: an online database of predicted prion domains in complete proteomes. *BMC Genomics*. 2014;15:102.
73. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29.
74. Iglesias V, de Groot NS, Ventura S. Computational analysis of candidate prion-like proteins in bacteria and their role. *Front Microbiol*. 2015;6. doi:10.3389/fmicb.2015.01123
75. Puig S, Thiele DJ. Molecular mechanisms of copper uptake and distribution. *Curr Opin Chem Biol*. 2002;6(2):171-180.
76. Garai S, Citu SL, Singla-Pareek SK, Sopory CK, Yadav G. Complex networks of prion-like proteins reveal cross talk between stress and memory pathways in plants. *Front Plant Sci*. 2021;12:707286.
77. Jarvis S, Mostaert A. *The Functional Fold: Amyloid Structures in Nature*. CRC Press; 2012.
78. Itakura AK, Chakravarty AK, Jakobson CM, Jarosz DF. Widespread prion-based control of growth and differentiation strategies in *Saccharomyces cerevisiae*. *Mol Cell*. 2020;77(2):266-278.
79. Oamen HP, Lau Y, Caudron F. Prion-like proteins as epigenetic devices of stress adaptation. *Exp Cell Res*. 2020;396(1):112262.
80. Malinovska L, Palm S, Gibson K, Verbavatz J-M, Alberti S. *Dictyostelium discoideum* has a highly Q/N-rich proteome and shows an unusual resilience to protein aggregation. *Proc Natl Acad Sci U S A*. 2015;112(20):E2620-E2629.
81. Pallarès I, de Groot NS, Iglesias V, et al. Discovering putative prion-like proteins in plasmodium falciparum: a computational and experimental analysis. *Front Microbiol*. 2018;9. doi:10.3389/fmicb.2018.01737
82. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49(D1):D480-D489.
83. Gene Ontology Consortium. The gene ontology resource: enriching a gold mine. *Nucleic Acids Res*. 2021;49(D1):D325-D334.
84. Klopfenstein DV, Zhang L, Pedersen BS, et al. GOATOOLS: a python library for gene ontology analyses. *Sci Rep*. 2018;8(1):10872.
85. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer; 2016.
86. Lee MD. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics*. 2019;35(20):4162-4164.
87. Hug LA, Baker BJ, Anantharaman K, et al. A new view of the tree of life. *Nat Microbiol*. 2016;1:16048.
88. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol*. 2011;7(10):e1002195.
89. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform*. 2004;5:113.
90. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972-1973.
91. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010;5(3):e9490.
92. Letunic I, Bork P. Interactive tree of life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics*. 2007;23(1):127-128.
93. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and Protein annotation. *Nucleic Acids Res*. 2016;44(D1):D457-D462. doi:10.1093/nar/gkv1070
94. Aramaki T, Blanc-Mathieu R, Endo H, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*. 2019;36(7):2251-2252.

95. Lee M. Bit: a multipurpose collection of bioinformatics tools. *F1000Research*. 2022;11(122):122.
96. Shen W, Ren H. TaxonKit: a practical and efficient NCBI taxonomy toolkit. *J Genet Genomics*. 2021;48(9):844-850.
97. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30.
98. Harel I, Chen YR, Ziv I, et al. Identification of protein aggregates in the aging vertebrate brain with prion-like and phase separation properties. *bioRxiv*. 2002. doi:10.1101/2022.02.26.482115
99. Amaral MJ d, Freire MHO, Almeida MS, Pinheiro AS, Cordeiro Y. Phase separation of the mammalian prion protein: physiological and pathological perspectives. *J Neurochem*. 2022;166:58-75. doi:10.1111/jnc.15586
100. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*. 2017;541(7637):353-358.
101. Wickner RB. [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science*. 1994;264:566-569.
102. Rogoza T, Goginashvili A, Rodionova S, et al. Non-mendelian determinant [ISP+] in yeast is a nuclear-residing prion form of the global transcriptional regulator Sfp1. *Proc Natl Acad Sci U S A*. 2010;107(23):10573-10577.
103. Volkov KV, Aksenova AY, Soom MJ, et al. Novel non-mendelian determinant involved in the control of translation accuracy in saccharomyces cerevisiae. *Genetics*. 2002;160(1):25-36.
104. Durocher D, Jackson SP. The FHA domain. *FEBS Lett*. 2002;513(1):58-66.
105. Harrison PM. Evolutionary behaviour of bacterial prion-like proteins. *PLoS One*. 2019;14(3):e0213030.
106. Kraut J. Serine proteases: structure and mechanism of catalysis. *Annu Rev Biochem*. 2003;46:331-358. doi:10.1146/annurev.bi.46.070177.001555
107. Muley VY, Akhter Y, Galande S. PDZ domains across the microbial world: molecular link to the proteases, stress response, and protein synthesis. *Genome Biol Evol*. 2019;11:644-659. doi:10.1093/gbe/evz023
108. White MJ, He H, Penoske RM, Twining SS, Zahrt TC. PepD participates in the mycobacterial stress response mediated through MprAB and SigE. *J Bacteriol*. 2010;192(6):1498-1510.
109. Wang N, Gottesman S, Willingham MC, Gottesman MM, Maurizi MR. A human mitochondrial ATP-dependent protease that is highly homologous to bacterial Lon protease. *Proc Natl Acad Sci U S A*. 1993;90(23):11247-11251.
110. Urban S, Dickey SW. The rhomboid protease family: a decade of progress on function and mechanism. *Genome Biol*. 2011;12:231. doi:10.1186/gb-2011-12-10-231
111. Paschkowsky S, Hamzé M, Oestereich F, Munter LM. Alternative processing of the amyloid precursor protein family by rhomboid protease RHBDL4. *J Biol Chem*. 2016;291(42):21903-21912.
112. Penalva YCM, Recinto SJ, Munter L-M. Relevance of RHBDL4-mediated APP processing for Alzheimer's disease. *Alzheimers Dement*. 2021;17:e053878.
113. Urban S, Moin SM. A subset of membrane-altering agents and γ-secretase modulators provoke nonsubstrate cleavage by rhomboid proteases. *Cell Rep*. 2014;8(5):1241-1247.
114. Kosolapova AO, Belousov MV, Sulatskaya AI, et al. Two novel amyloid proteins, RopA and RopB, from the root nodule bacterium *Rhizobium leguminosarum*. *Biomolecules*. 2019;9(11). doi:10.3390/biom9110694
115. Galkin AP, Sysoev EI, Valina AA. Amyloids and prions in the light of evolution. *Curr Genet*. 2023;1-14.
116. Christensen LF, Bakmann LM, Hansen KF, et al. The sheaths of are made of a new type of amyloid protein. *Front Microbiol*. 2018;9:2729.
117. Shorter J, Lindquist S. Prions as adaptive conduits of memory and inheritance. *Nat Rev Genet*. 2005;6(6):435-450.
118. Chakravarty AK, Smejkal T, Itakura AK, Garcia DM, Jarosz DF. A non-amyloid prion particle that activates a heritable gene expression program. *Mol Cell*. 2020;77(2):251-265.
119. Capra EJ, Laub MT. Evolution of two-component signal transduction systems. *Annu Rev Microbiol*. 2012;66:325-347. doi:10.1146/annurev-micro-092611-150039
120. Wickner RB, Kelly AC. Prions are affected by evolution at two levels. *Cell Mol Life Sci*. 2016;73(6):1131-1144.
121. Vallota-Eastman A, Arrington EC, Meeken S, et al. Role of diversity-generating retroelements for regulatory pathway tuning in cyanobacteria. *BMC Genomics*. 2020;21(1):664.
122. Halfmann R, Alberti S, Lindquist S. Prions, protein homeostasis, and phenotypic diversity. *Trends Cell Biol*. 2010;20(3):125-133.
123. Jong IG d, de Jong IG, Haccou P, Kuipers OP. Bet hedging or not? A guide to proper classification of microbial survival strategies. *BioEssays*. 2011;33:215-223. doi:10.1002/bies.201000127

SUPPORTING INFORMATION

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

How to cite this article: Zajkowski T, Lee MD, Sharma S, et al. Conserved functions of prion candidates suggest a primeval role of protein self-templating. *Proteins*. 2023;91(9):1298-1315. doi:10.1002/prot.26558