



Kent Academic Repository

Kulkarni, Prakash, Porter, Lauren, Chou, Tsui-Fen, Chong, Shasha, Chiti, Fabrizio, Schafer, Joseph W., Mohanty, Atish, Ramisetty, Sravani, Onuchic, Jose N., Tuite, Mick F. and others (2025) *Evolving concepts of the protein universe*. *iScience*, 28 (3). ISSN 2589-0042.

Downloaded from

<https://kar.kent.ac.uk/110638/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1016/j.isci.2025.112012>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Perspective

Evolving concepts of the protein universe

Prakash Kulkarni,^{1,2,3,*} Lauren Porter,⁴ Tsui-Fen Chou,^{3,5} Shasha Chong,⁶ Fabrizio Chiti,⁷ Joseph W. Schafer,⁴ Atish Mohanty,¹ Sravani Ramisetty,¹ Jose N. Onuchic,^{8,9} Mick Tuite,¹⁰ Vladimir N. Uversky,^{11,*} Keith R. Weninger,¹² Eugene V. Koonin,^{4,*} John Orban,^{13,14,*} and Ravi Salgia^{1,*}

¹Department of Medical Oncology, City of Hope Medical Center, Duarte, CA, USA

²Department of Systems Biology, City of Hope Medical Center, Duarte, CA, USA

³Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA

⁴National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

⁵Proteome Exploration Laboratory, Beckman Institute, California Institute of Technology, Pasadena, CA, USA

⁶Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA

⁷Department of Experimental and Clinical Biomedical Sciences “Mario Serio”, University of Florence, Florence, Italy

⁸Center for Theoretical Biological Physics, Rice University, Houston, TX, USA

⁹Department of Physics and Astronomy, Rice University, Houston, TX, USA

¹⁰Kent Fungal Group, School of Biosciences, Division of Natural Sciences, University of Kent, CT2 7NJ Canterbury, UK

¹¹Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, USA

¹²Department of Physics, North Carolina State University, Raleigh, NC, USA

¹³W. M. Keck Laboratory for Structural Biology, University of Maryland Institute for Bioscience and Biotechnology Research, Rockville, MD, USA

¹⁴Department of Chemistry and Biochemistry, University of Maryland, College Park, MD, USA

*Correspondence: pkulkarni@coh.org (P.K.), vuffersky@usf.edu (V.N.U.), koonin@ncbi.nlm.nih.gov (E.V.K.), jorban@umd.edu (J.O.), rsalgia@coh.org (R.S.)

<https://doi.org/10.1016/j.isci.2025.112012>

SUMMARY

The protein universe is the collection of all proteins on earth from all organisms both extant and extinct. Classical studies on protein folding suggested that proteins exist as a unique three-dimensional conformation that is dictated by the genetic code and is critical for function. In this perspective, we discuss ideas and developments that emerged over the past three decades regarding the protein structure-function paradigm. It is now clear that ordered (active/functional) and disordered/denatured (and hence inactive/non-functional) represent a continuum of states rather than binary states. Some proteins can switch folds without sequence change. Others exist as conformational ensembles lacking defined structure yet play critical roles in many biological processes, including forming membrane-less organelles driven by liquid-liquid phase separation. Numerous diverse proteins harbor segments with the potential to form amyloid fibrils, many of which are functional, and some possess prion-like properties enabling conformation-based transfer of heritable information. Taken together, these developments reveal the remarkable complexity of the protein universe.

INTRODUCTION

Proteins are responsible for the great majority of enzymatic, structural, transport, and signaling functions in every cell. Even the simplest bacterium or archaeon encodes hundreds of structurally and functionally distinct, diverse proteins, whereas complex multicellular organisms, plants, and animals typically encode tens of thousands of proteins, and this proteome is further expanded several folds through alternative splice forms and post-translational modifications. As such, proteins constitute a universe defined as the collection of all proteins from all organisms on earth, both extinct and extant.¹ The protein universe is often described in terms of hierarchical models, although hierarchy is not necessarily its intrinsic property. Nonetheless, proteins are organized into tightly connected networks of homologous domains, suggesting that all proteins originated from a small number of ancestral folds.^{2,3} Deep evolutionary re-

constructions indicate that the last universal cellular ancestor (LUCA) that is thought to have lived about 4 billion years ago already encoded at least several hundred proteins with diverse folds indicating that pivotal events in protein evolution occurred at very early stages of life's history, antedating the emergence of “modern” cells.⁴

Proteins have been studied extensively across scales and disciplines to understand how amino acid sequence relates to a protein's 3D structure and how that structure relates to its function. Observations dating back to the mid-1950s and later confirmed by X-ray crystallography, site-directed mutagenesis, and solution NMR, helped to firmly establish that most proteins have well-defined 3D structures. These observations led Anfinsen to postulate the “thermodynamic hypothesis”,⁵ which states that the protein's native conformation is comprised of the totality of interatomic interactions which is determined by the amino acid sequence in a given environment.



However, classic work by Kauzmann (1959)⁵ explicitly refers to an alternative view of the protein folding problem: “According to all that we know now about protein structure, we have good reason to believe that disorder might be introduced into a protein in small increments.” In fact, by the 1970s, it was also recognized that proteins display complex behaviors and that protein structure can change in response to different aqueous environments.^{6,7} Nonetheless, for the purposes of protein structure analyses, it was generally held that proteins exist as binary states: ordered/ folded (active) or denatured/unfolded (inactive).⁸ Subsequently, this concept was amended by the notion of intermediates between the folded and unfolded states. By the late 1980s and early 1990s, the energy landscape theory of protein folding and the funnel concept were developed.^{9,10} Per this theory, protein folding represents the progressive organization of partially folded structures in the conformational ensemble as that ensemble proceeds to the natively folded structure.¹¹ As a result of evolution, proteins fold on rugged funnel-like landscapes biased toward the native structure.^{12,13} This funnel landscape theory captured how folding is related to function by showing not only how the precise structure can be attained but also how structural excitations can be involved in protein functionality.

The idea that excited states in the funnel landscape can have functional relevance became a major research topic, and in further major developments started in the 1990s, it became clear that proteins need not always be highly structured to be functional as presciently suspected by Kauzmann back in 1959.⁵ Indeed, it is now well recognized that a large fraction of the proteomes of organisms across all three domains of life is comprised of intrinsically disordered proteins/peptides (IDPs) and many more proteins with ordered domains contain intrinsically disordered regions (IDRs) that, by definition, lack rigid 3D structure yet are functional.¹⁴ Furthermore, it is now clear that some ordered proteins can switch folds and gain new function¹⁵ and that regions in certain folded proteins (and even entire proteins) can “unfold”, in a transition from order to disorder, in response to physical or chemical stimuli.¹⁶

Studies on fold switching in natural and experimentally designed systems showed that transitions occur when states have diminished stability.^{17,18} -Fold switching involves flexible regions in one conformer or the other, and a new binding surface that when exposed in the alternate fold, can result both in stabilization of the alternative state as well as a change of the biological function. Detailed studies of IDPs/IDRs showed that, although IDPs can transition from disordered conformational ensembles to folded structures either prior to or upon binding to their cognate targets, many IDPs remain largely disordered even as they interact with the targets, and yet others exhibit dynamic excursions and stochastically switch conformational states while remaining disordered.¹⁹ Thus, IDPs may be only marginally unstable although they can be tipped to populate certain preferred conformations. The latter may populate a wide spectrum from a slight change in the ensemble conformation to a fully folded structure to gain functionality. Such changes in the structural ensemble of IDPs are conceptually similar to the fold switching events observed in marginally stable folded proteins in response to mutations or environmental triggers leading to new functions (Figure 1).²⁰

Additional discoveries in the late 1990s further changed the canonical perception of protein structure/function when it became evident that, at least *in vitro*, most proteins or individual domains within them can convert into fibrils with all the characteristics of amyloid that accumulates in a variety of human and animal disorders.²² Consistent with this remarkable observation, it has been shown that fully soluble peptides and proteins, which have specific functions in this soluble state, can form amyloid fibrils representing distinct, well-defined functional states that are not necessarily pathological.²³

The discovery of prions, a sub-class of amyloids in which protein aggregation becomes self-perpetuating and, at least, in well-studied cases, infectious, further challenged prevailing wisdom by demonstrating the existence of protein-based inheritance.^{24–26} Although originally met with disbelief, this form of inheritance has been found in diverse organisms in all three kingdoms of life, is now accepted as a *bona fide* epigenetic mechanism.²⁴ Moreover, a groundbreaking study revealed that protein-based inheritance may not be particularly rare or restricted to amyloid-based prions.²⁶ This study identified a new class of prion-like proteins that are capable of causing heritable, protein-based “molecular memories” that are sustainable over hundreds of generations yet are distinct from the classical amyloid prions.²⁶

Although previously observed with certain ordered proteins, a more recent notable discovery demonstrated how IDPs can undergo liquid-liquid phase separation (LLPS) via multivalent protein-protein interactions, thereby driving the biogenesis of numerous membrane-less organelles (MLOs).²⁷ Indeed, bioinformatics predictions suggest that the vast majority of IDPs harbor the propensity for LLPS. In this perspective, we discuss how protein structure and folding are fundamental to the inherent ability of proteins to switch between conformations, revealing essential but previously underappreciated features of the protein universe.

INTRINSICALLY DISORDERED YET FUNCTIONAL PROTEINS

Two characteristic features of many IDPs/IDRs are low mean hydropathy and relatively high net charge, which are critical prerequisites for the lack of rigid 3D structure in proteins under physiological conditions. Some IDPs can transition to order under conditions such as low pH or heating that result in the denaturation or unfolding of ordered proteins while other IDPs lose function when folded, and the activation of yet other IDPs involves unmasking cryptic disorder.²⁸ These properties suggest that the order/disorder equilibrium in IDPs is subtle and that, many IDPs are very close to this boundary. It has been proposed that IDPs are “edge of chaos” systems that operate in a state between order and randomness (chaos), where complexity is at its maximum.¹⁹ Furthermore, consistent with the edge-of-chaos concept, analysis of protein structures solved in different conditions and functional states revealed that hundreds of protein fragments, dubbed “dual personality fragments”, exist both as disordered and ordered states. These fragments display singular features that differentiate them from both folded proteins and IDPs/IDRs and are frequent targets of regulation, either by allostery or by post-translational modifications.²⁹

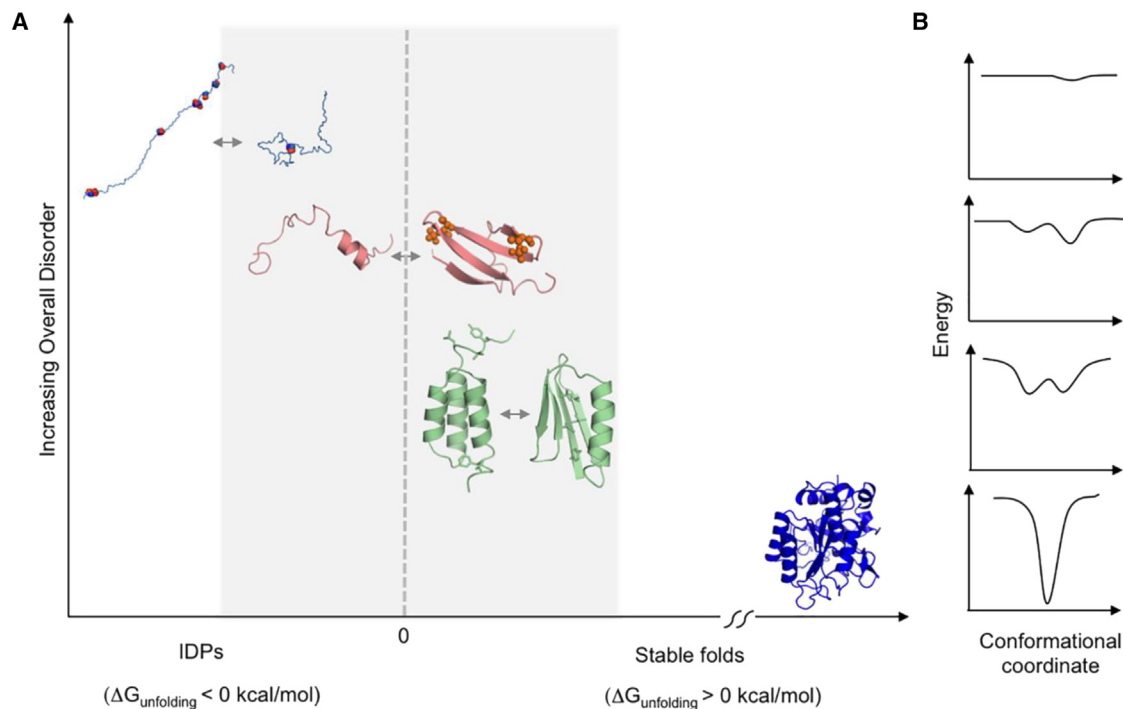


Figure 1. Proteins on the brink of stability can undergo a continuum of order/disorder transitions

(A) Examples of transitions from top left to bottom right: transition between the extended and collapsed disordered states of prostate associated gene 4 (PAGE4), modulated by phosphorylation; 111 disorder-to-order transition of 4E-BP2 induced by phosphorylation; 23 order-to-order fold switching between GA98 and GB98, triggered by single amino acid changes or ligand binding.²¹ In contrast, stable proteins such as subtilisin (shown in dark blue) do not undergo such changes.

(B) Approximate energy well diagrams for each protein from PAGE4 (top) to subtilisin (bottom) (from the study by Kulkarni P. et al.²⁰ with permission).

It has also been demonstrated that IDPs can interact with picomolar affinity with partner proteins even while being fully disordered and highly dynamic.³⁰ Furthermore, some IDPs/IDRs can fold differently upon interaction with different partners, thereby gaining distinct structures in a context-dependent manner.^{31,32} Many IDPs/IDRs can also form “fuzzy complexes” possessing high levels of disorder in bound states.^{33–35}

As a result of all these discoveries, our view of protein interactions has come a long way from the concept of rigid structures. It is now understood that proteins form conformational ensembles of proteins which manifest in multiple modes of interactions between partner proteins. This context-dependence, referred to as multiplicity of binding modes, is achieved by sampling multiple minima of the interaction energy landscape.³⁶ This situation has been observed in many protein systems where multiple conformations are necessary. Initially, multiplicity of binding modes was discussed in the context of dual funnel basins for functional reasons,^{37–39} but in many cases, multiple stable basins may exist.⁴⁰ The possibility of multiple structures for the same protein is particularly notable in the case of IDPs.

Physiologically, IDPs participate in a plethora of biological functions where multiple partner interactions and high-specificity/low-affinity interactions are crucial. Thus, IDPs are involved in several key processes, such as transcription, translation, and signaling. IDPs are also critical to several higher-order phenomena, such as regulation of the cell division cycle, circadian

rhythms, and phenotypic plasticity (see recent reviews).^{14,41,42}

Here, we focus on those aspects that have contributed to the conceptual advancement of the IDP field in the past few years, namely IDP conformational noise and LPS and discuss their implications. We also discuss fold-switching proteins, which are conformationally heterogeneous like IDPs, but differ in stability, selection, and biological scope.¹⁷

Noise, that is, variation in cell-to-cell characteristics that occurs in populations of isogenic cells, at the levels of transcription, translation, post-translational modification, protein folding, and various signaling pathways, is now a well-recognized concept in biology.⁴³ However, noise caused by IDP conformational dynamics has remained under-appreciated. A living cell is a complex adaptive system, a dissipative microenvironment that operates far from thermodynamic equilibrium. Operationally, a living cell is akin to a computer. Its hardware and software constitute the “wetware” attuned to logic gates allowing the cell to “make decisions”.¹⁹ In this metaphor, IDPs represent critical components of the wetware. They are “wired” to form the cell’s protein interaction network that is scale-free and minimally frustrated.⁴⁴ The design and functionality of this network to a large extent are governed by the IDPs whose expression is tightly regulated. Because IDPs exist as ensembles of states populating an energy landscape, where the individual states are separated by relatively low energy barriers, or multiple conformations that are further influenced by specific post-translational modifications such as

phosphorylation, IDP conformational dynamics manifests itself as conformational noise.⁴⁵ This noise can potentially amplify transcriptional and translational noise to stochastically switch cellular phenotypes. Dysregulation of IDPs in response to intrinsic or extrinsic perturbations results in “rewiring” of the network and influences cell-fate decisions.^{19,45} Furthermore, such dysregulation of IDPs can trigger pathological states including cancer and neurodegenerative diseases.⁴⁶ Given that ~90% of transcription factors harbor IDRs, conformational noise is likely to comprise an integral part of transcriptional noise, underscoring its fundamental importance.^{45,47}

DISORDER, LLPS, AND MLOs

Probably the most graphic illustration of the emergent behavior based on protein intrinsic disorder is the LLPS phenomenon that has attracted considerable attention in recent years.⁴⁸ Although LLPS *per se* has been known since the 19th century, and it was also well known that folded proteins can undergo LLPS, the phenomenon has found new applications in the realm of biological functions of the IDPs. LLPS is thought to be the molecular mechanism underlying the emergence of numerous biomolecular condensates or MLOs, which represent distinct compartments in the cell cytoplasm or nucleoplasm that are not bounded by membranes, have a composition different from the surrounding cellular environment, and play a crucial role in spatiotemporal organization of the intracellular space. Nucleoli, stress granules, P bodies, pericentriolar material, paraspeckles, and germline P granules are all typical MLOs.^{49,50}

MLOs obviate the necessity to facilitate and regulate molecular interactions via reversible and controllable sequestration of interacting molecules in specialized compartments.⁵¹ MLOs are open dynamic systems; they represent highly pliable complex networks that exist at the edge of chaos. They are characterized by unique structural and compositional dynamics. Furthermore, their multicomponent nature and polyfunctionality facilitate the fine regulation of many biological processes. MLOs are exquisitely sensitive to changes in external conditions. Even minor changes can result in constituent proteins aggregating irreversibly, transitioning to a gel-like state, or forming amyloid fibrils.⁵² The structural dynamics and fluidity of MLOs, their liquid-like characteristics, as well as their multifunctionality and responsiveness to various intra- and extracellular signals all originate from the specific IDPs/IDRs that serve as LLPS drivers.⁵³ In line with these considerations, MLOs are enriched in IDPs and proteins containing IDRs,⁵¹ and their biogenesis was shown to be controlled by IDP concentrations, post-translational modifications, binding partners, and environmental perturbations.

Besides being important for cellular organization, MLOs or biomolecular condensates have been reported to play a role in many cellular processes, including transcriptional regulation, chromatin organization, and DNA damage repair.^{27,54} At the molecular level, the formation of functionally relevant condensates is often driven by dynamic, multivalent, selective interactions between IDRs.^{27,55,56} An increasing number of reports have documented that disease-associated mutations within IDPs can lead to dysregulation of their multivalent interactions and condensate formation, thereby contributing to the pathology of many dis-

eases, including cancers and neurodegenerative disorders.⁴⁶ Many IDPs that carry disease-associated mutations are involved in transcriptional regulation, e.g., fusion of oncogenic transcription regulators and polyQ expansion proteins associated with neurodegenerative diseases.⁵⁷ Targeting these pathological condensates suggests potentially new and exciting therapeutic strategies.

SAME SEQUENCE BUT ALTERNATE FOLDS: FOLD-SWITCHING PROTEINS

In 2008, the protein XCL1 was found to independently, and reversibly, interconvert between two conformations.⁵⁸ In the 15 years since this initial discovery, the list of fold-switching “exceptions” to Anfinsen’s postulate has continued to grow. Now, fold switchers are a recognized class of proteins with unique biophysical characteristics and are estimated to comprise as much as ~4% of the Protein Data Bank.¹⁵ Examples of other notable naturally occurring fold switchers include: (1) KaiB, an integral component of the cyanobacterial circadian clock,⁵⁹ (2) RfaH, a member of the universally conserved NusG family of transcription factors,⁶⁰ (3) Lysenin, a pore-forming toxin that switches folds upon membrane binding,⁶¹ and (4) ORF9b, a virulent SARS-CoV-2 protein that was found to fold-switch upon binding to human TOM70.⁶² To date, nearly 100-fold switchers have been experimentally characterized, several of which are implicated in diseases such as cancer, viral infection, bacterial infection, and autoimmune disorders.⁶³

The increasing number of identified fold-switching proteins and their health implications have broadened researchers’ interests in the fundamental biophysical principles governing this phenomenon. Although fold switching is not prohibited by physical or chemical laws, a protein must assume a biologically active fold, and stable alternatives would act as kinetic traps.⁶⁴ Such traps are expected to be avoided by fold-switching proteins via transformations that do not pass through fully unfolded states. Zuber et al.²¹ analyzed the mechanism of the fold-switching transformation in RfaH. RfaH is among the most dramatic fold-switchers, with 50 residues of its C-terminal domain reversibly interconverting between an α -hairpin and a β -barrel. This transformation occurs as a result of binding to opsDNA, and the refolding path of RfaH does not sample a fully unfolded state, reinforcing predictions made on how fold-switching mechanisms would avoid kinetic traps.²¹

Fold switching is also being exploited to better understand how amino acid sequences encode both protein structure and function. Ruan et al.⁶⁵ engineered mutational pathways between three common folds: S6 (a component of the 30S ribosomal subunit), GA, and GB, corresponding to α/β -plait, homeodomain-like 3α , and ubiquitin-like β -grasp folds, respectively (Figure 2A). Critical points in sequence space were identified where a single mutation changes the conformation and function of a protein. The existence of these critical points demonstrates that unique conformations are not necessarily separated by high barriers in the folding landscape.⁶⁶ Furthermore, Solomon et al. engineered a single sequence that transitions between the GA (3α) fold and the S6 (α/β -plait) fold in response to temperature changes within a relatively narrow range.¹⁸ These protein engineering studies

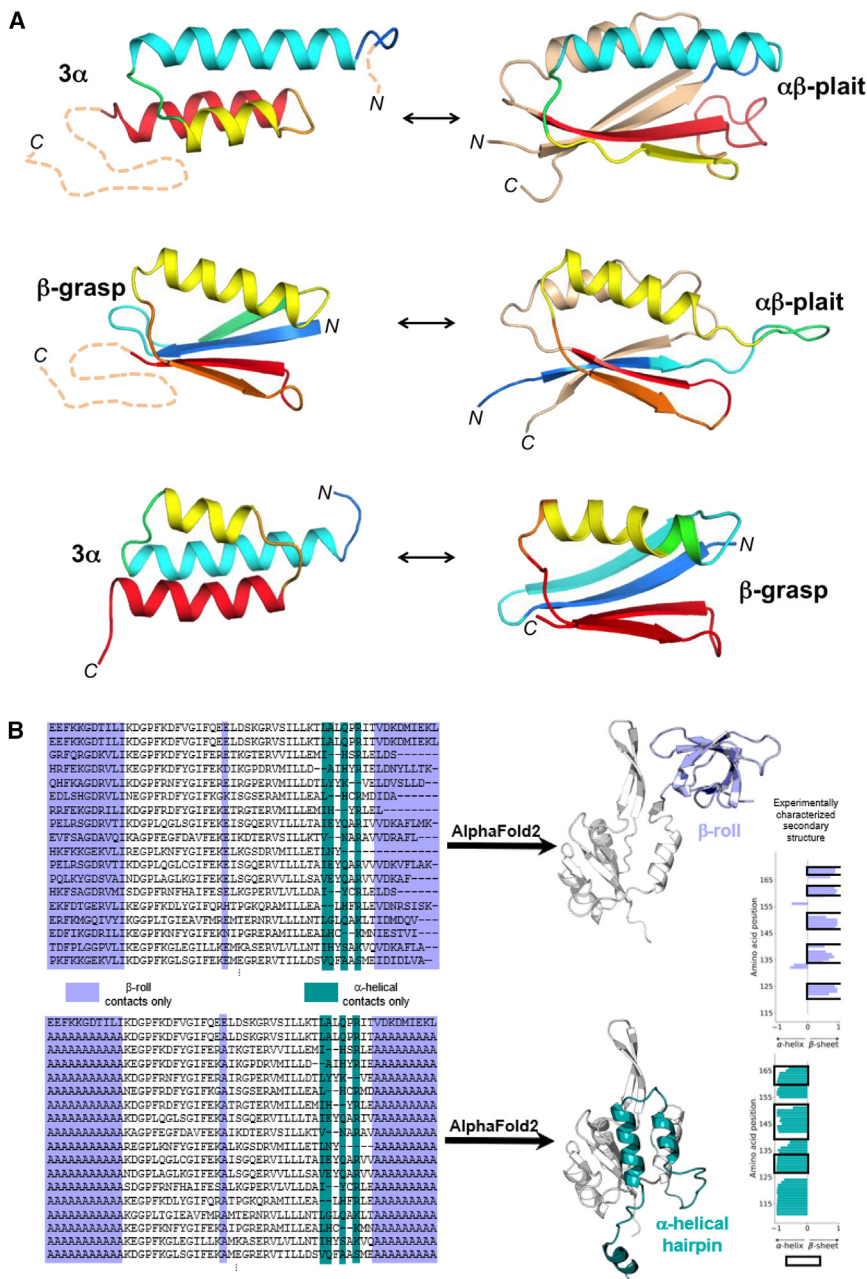


Figure 2. Fold-switching proteins

(A) Summary of designed switches between 3 α , $\alpha\beta$ -plait, and β -grasp fold topologies that have been characterized structurally. Color-coding indicates how the corresponding regions change in the alternative fold. The dashed lines show regions that are disordered in one of the states but not the other.

(B) Top, a NusG N-terminal (NGN) fold (light gray) and a C-terminal β -roll fold (lavender) are predicted from a deep input MSA (region corresponding to the CTD shown) generated from the sequence of a NusG protein with $\leq 29\%$ aligned identity to its homologs with experimentally determined structures. Predicted β sheets in the C-terminal domain that agree closely with the β sheets predicted from nuclear magnetic resonance experiments are shown with black boxes surrounding lavender bars. Bottom, a NusG N-terminal (NGN) fold (light gray) and a C-terminal α helical hairpin fold (teal) are predicted from a modified input MSA of the NusG homolog in which columns predicted to form only β -roll contacts are changed to alanine. Predicted α helices in the C-terminal domain that agree with the α helices predicted from nuclear magnetic resonance experiments are shown with black boxes surrounding teal bars. Protein structures were generated with PyMOL43.

Machine learning algorithms, such as AlphaFold2, have revolutionized predictions of protein structure but miss nearly all (85/93) alternate conformations of fold-switching proteins using default settings and shallow random sequence sampling.⁶⁸ Given that structure prediction by AlphaFold2 is based on evolutionarily conserved patterns in protein sequences, this discrepancy might be explained by fold switching being an infrequent evolutionary quirk.⁶⁹ Alternatively, machine learning algorithms might miss coevolutionary signatures of alternative protein folds. To investigate this, Schafer and Porter developed alternative contact enhancement (ACE) to apply coevolutionary algorithms to known fold-switching proteins.⁷⁰ They

demonstrate the importance of disordered regions in remodeling ordered states. They also imply a naturally occurring mechanism of fold-switching evolution. Recent work has uncovered a stepwise mutational path that switched the helix-turn-helix fold of bacterial response regulators to a winged helix,⁶⁷ demonstrating that natural proteins have also switched folds by stepwise mutation over evolutionary history.

Because fold-switching proteins challenge the single-fold paradigm that has shaped much of protein structure prediction, considerable interest has focused on predicting both conformations of fold-switching proteins from sequence.

calculated coevolutionary couplings between amino acid pairs from multiple sequence alignments (MSAs) that represent protein superfamilies (deep MSAs that contain diverse-yet-homologous sequences) and subfamilies (shallower MSAs that preferentially retain sequences highly similar to the target protein). Applying this method to all known fold-switching proteins for which deep MSAs could be generated, they found unique coevolutionary signals for both conformations in all cases. These results suggest that fold switching confers an advantage that can be selected for during evolution, and that prediction of fold-switching proteins should be possible using strategies that

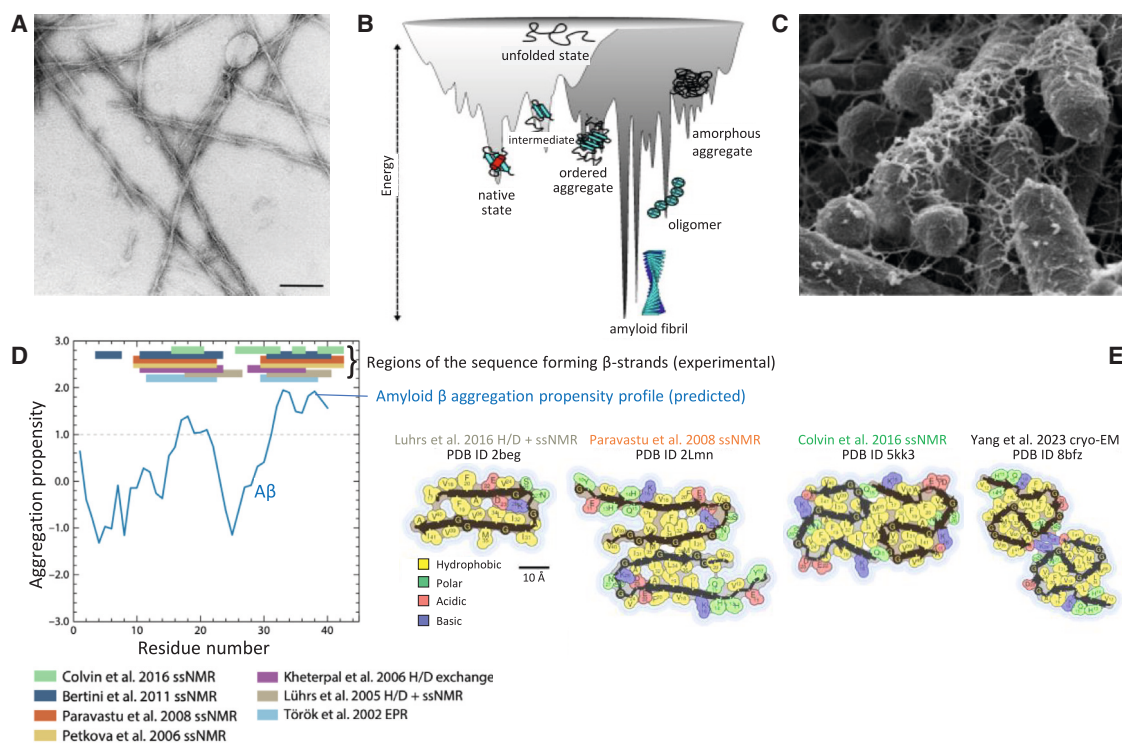


Figure 3. Amyloid-forming proteins

(A) Amyloid-like fibrils formed from an SH3 domain. Readapted with permission from the study by Watson J. L. et al.⁷²
 (B) Energy funnel illustrating the protein folding (left) and the protein aggregation (right) components. Readapted with permission from the study by Parui S. et al.⁷⁴
 (C) Curli of bacteria visible as filamentous structures as an example of functional amyloids. Readapted with permission from the study by Olsen A. et al.⁷⁷
 (D) Comparison between predicted protein aggregation profile (blue) and regions of the sequence found experimentally to form β -strands in amyloid fibrils (colored bars). Readapted with permission from the study by Westermark P. et al.⁷⁵
 (E) The four structures on the right are cross-sectional structures of amyloid fibrils along with their PDB entries and references, as extracted from the amyloid atlas. Readapted with permission from the study by Fowler D.M. et al.⁷⁸

rely on structural information derived from coevolution when enough sequence information is available.

Dual-fold coevolution paves the way for machine learning algorithms to predict 2-folds from one sequence more consistently. Accordingly, *in silico* alanine mutations introduced into the MSAs of dynamic single-fold proteins led to successful predictions of structural ensembles in several cases.⁷¹ This latter approach was tested on a NusG variant with 29% sequence identity to its closest homolog with known structure from Protein Data Bank. AlphaFold2 successfully predicts both its α -hairpin and the β -barrel conformations. These predictions are consistent with amino-acid-specific secondary structure calculated from NMR assignments⁷⁰ (Figure 2B). There has also been success using a class of algorithms broadly called diffusion models to predict/engineer proteins.⁷² To explore diffusion models' ability to sample the conformational ensemble, Jing et al. developed EIGENFOLD.⁷³ Although its accuracy is competitive for single-fold predictions, it is not yet able to accurately capture both conformations of fold-switching proteins. Nevertheless, with these and other notable computational developments,⁷⁴ it is expected that methods will be developed in the near future that will predict two 3D structures from one amino acid sequence consistently and accurately.

CONVERSION OF NATIVE PROTEINS INTO AMYLOID FIBRILS

The ability of certain proteins to convert from a soluble state—either folded or intrinsically disordered—into well-defined aggregates known as amyloid fibrils is now well documented. Amyloid fibrils are characterized by a fibrillar morphology, a β sheet, with parallel, in register β strands perpendicular to the fibril axis, and the ability to bind specific dyes such as Congo red and thioflavin T/S, have been known for over a century. Until 1998, however, it was thought that such fibrils were formed by only about 20 proteins which were known to be the main constituents of the amyloid deposits associated with well-defined pathological states, such as Alzheimer's disease (AD), light chain amyloidosis (AL), familial amyloid polyneuropathy (FAP or ATTRv-PN), and several others.⁷⁵ In 1998, however, the SH3 domain from the p85 α subunit of bovine phosphatidylinositol 3-kinase was serendipitously found by Dobson and collaborators to form amyloids⁷⁶ (Figure 3A). These amyloid fibrils were morphologically, structurally, and tinctorially indistinguishable from those associated with disease. When this discovery was reported, many in the field believed it was a peculiar behavior of the SH3 domain associated with its native all- β fold. However, the following year, a full-length

α + β protein, human muscle acylphosphatase (mAcP), was found to undergo a similar conversion.²² These findings led to the novel concept that many, if not all, proteins can be induced to form amyloids *in vitro* under appropriate conditions, which found many experimental confirmations in the following 25 years, with over 100 proteins converted into *bona fide* amyloid-like fibrils including all- α proteins.

Amyloid formation is a distinct type of conformational switch of a protein from its native state (either folded or unfolded) to the amyloid fold dominated by a β sheet. Any organism must actively combat protein aggregation that would ultimately result in amyloid fibril formation because protein assemblies, particularly the protein oligomers formed early in the amyloidogenesis process, can be toxic. The protein homeostasis network, composed (at least) of the translational machinery, molecular chaperones, the ubiquitin-proteasome system (UPS) and the autophagy-lysosomal pathway (ALP), is now thought to prevent this deleterious, potentially pathological switch.⁷⁹ The amyloid state is thermodynamically the most stable conformational state of a protein, and it is exactly where a protein ends up going if it is not attended by external factors that kinetically impede aggregation^{80,81} (Figure 3B). Moreover, there is not a unique amyloid fold that a specific protein can adopt, but multiple ones, known as polymorphs, which exacerbates the problem of thermodynamic stability and makes amyloid formation more favorable entropically^{80,81} (Figure 3B).

However, in addition to representing a pathological entity, amyloid fibrils can also constitute an ensemble of alternative functional states for proteins. The first evidence of a functional amyloid was obtained in 2002, only four years after the amyloid fold was discovered as a common structural state of proteins.⁸² In that work, the *Escherichia coli* CsgA curlin subunit was found to assemble *in vivo* and *in vitro* into fibrous structures, known as curli, with amyloid properties. Curli are anything but deleterious for bacteria; indeed, they have well defined, specific functions that allow bacteria to colonize inert surfaces and mediate binding to host proteins⁸² (Figure 3C). Other amyloid-associated functions have also been discovered in eubacteria, archaea, protists, fungi, plants, arthropods, and animals that is, across the tree of life.⁸³ The first case of a human protein was found in 2006, when the intralumenal domain of Pmel17, a membrane protein of the melanosomes present in melanocytes, was shown to form fibrous striations in the form of amyloid fibrils on which the melanin polymers responsible for protecting human skin from damaging UV radiation are able to grow.⁷⁸ Subsequently, a number of hormone peptides were found to accumulate as amyloids in exocytic vesicles, the pituitary secretory granules, before they are released as soluble peptides into the extracellular space.²³ In this case, amyloid has a specific storage function, allowing the compaction of hormones before they are released in a controlled manner. The receptor-interacting serine/threonine-protein kinase 1/3 (RIP1/RIP3) also forms amyloid fibrils that mediate programmed cell necrosis induced by the tumor necrosis factor (TNF), extending the range of amyloid functions beyond the boundaries of structural roles to include signaling.⁸⁴ Other cases then followed, making it clear that the amyloid switch represents an important functional resource in nature to generate new biological functions.

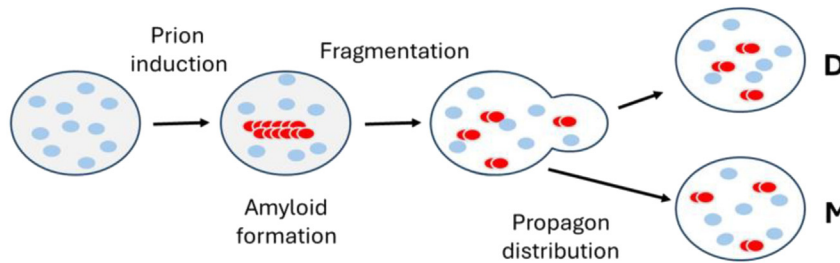
From a mechanistic standpoint, it is now clear that amyloid fibril formation occurs through a nucleated process where oligomers or even monomers that possess an appropriate conformation act as nuclei (primary nucleation) and recruit new monomers/oligomers to grow into a fibril with β sheet (elongation). Fibrils can also break into shorter yet growing fragments or catalyze the formation on their surface of new oligomers acting as nuclei (secondary nucleation), thereby augmenting their potential to propagate and diffuse.⁸⁵ It soon became clear that specific amino acid segments of a peptide/protein sequence with a high β sheet propensity, high hydrophobicity, and low net charge promoted amyloid fibril formation and formed the β strands stacking along the fibril axis as shown here for A β (Figure 3D). In addition, mutations within these segments had major effects on the rate of amyloid formation.^{83,86} With the advent of dedicated applications of solid-state nuclear magnetic resonance (ssNMR) spectroscopy, and particularly cryo-electron microscopy (cryo-EM), many amyloid folds have been resolved to a high resolution which was unimaginable a few years ago. An amyloid atlas has recently been created,⁸⁷ listing all such structures, along with representative images and references (<https://people.mbi.ucla.edu/sawaya/amyloidatlas/>). Polymorphism is a major feature of amyloids as the same peptide/protein can form several different amyloid structures (Figure 3E). Nevertheless, all of the structures present a β sheet structure, with parallel, in-register, β strands perpendicular to the fibril axis.⁸⁷ Moreover, from a statistical survey of these structures, it is clear that amyloid-forming segments consistently have a high β sheet propensity, high hydrophobicity, and low net charge.

PRIONS: PROTEINS AS MEDIATORS OF TRANSGENERATIONAL INHERITANCE

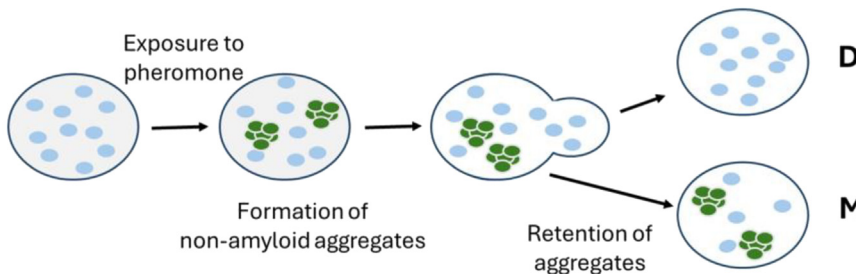
Within the ever-expanding family of amyloid-forming proteins, a remarkable subclass has emerged: prion-forming proteins, which add a paradigm-breaking dimension to the protein universe. When these diverse proteins refold into their amyloid form, they can mediate the flow of heritable information, which can be vertically transmitted across many generations. Prions thus define a previously unsuspected, unique form of protein-only epigenetic inheritance. Furthermore, at least one prion is linked to several potentially fatal neurodegenerative diseases in humans and other mammals.

The first prion-forming protein to be described was mammalian PrPc. Its transmissible conformer (designated PrPSc) has been linked to several slow-manifesting, transmissible, fatal neurodegenerative diseases, typified by Creutzfeldt Jakob disease (CJD) in humans, bovine spongiform encephalopathy (BSE) in cows, and Scrapie in sheep.⁸⁸ The original hypothesis proposed by Prusiner to explain the emergence and propagation of such a protein-only infectious agent postulated a form of self-templated structural conversion giving rise to highly stable and transmissible polymeric assemblies of PrPSc.⁸⁸ The prion assemblies of PrPSc then act effectively as “proteopathic agents” that can spread within the infected host and can also be transferred between individuals of the same species. Prions are thus a unique class of infectious agents that challenges the accepted nucleic acid-based paradigm of heredity and infection.

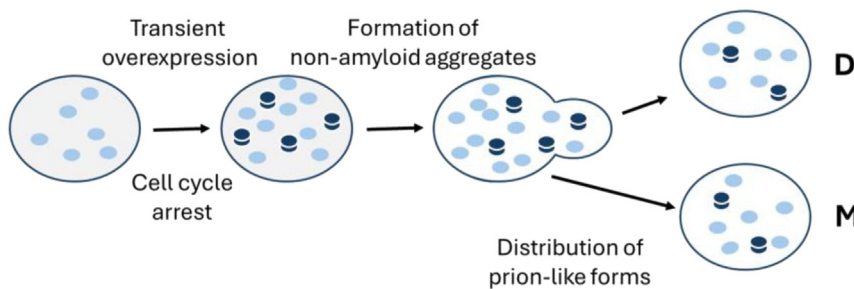
A Prion



B Mnemons



C Prion-like determinants



After discovering and validating PrPSc as a prion, several fungal proteins were identified, in the yeast *Saccharomyces cerevisiae* (Ure2p and Sup35p) and one in the filamentous fungus *Podospira anserina* (HET-s) that showed prion-like properties⁸⁹ i.e., can switch via a self-templating mechanism to one or more alternative amyloid conformations for transmissible prion seeds (Figure 4A). These seeds, sometimes referred to as “propagons”, are then transmitted from mother to daughter cells during mitosis and to haploid spores following meiosis.

Such atypical protein behavior leading to the transgenerational epigenetic inheritance of specific phenotypes in fungi is not restricted to prions. For example, non-transmissible high molecular weight “superassemblies” of the Whi3p protein—referred to as a mnemonic⁹⁰—is retained by the mother cell in response to an aborted mating event, giving rise to a form of molecular memory not inherited by daughter cells (Figure 4B). There are also 50 or more low abundance, intrinsically disordered proteins which can generate a range of different, stably inherited phenotypes in yeast when transiently over expressed and share many of the properties ascribed to prions (Figure 4C).²⁶

Figure 4. Three types of protein-based inheritance described in the yeast *Saccharomyces cerevisiae*

(A) Prions that can be generated *de novo* by overexpression of the prion-forming protein. The resulting amyloid fibrils (red) are fragmented by Hsp104 into smaller elements called propagons that can be transferred to daughter cells at mitosis. (B) Mnemons are generated in cells in response to a chemical signal. The only example described in detail so far is the Whi3-based mnemons (ref.⁹⁰), high-molecular-weight forms of the protein (in dark blue) that are retained by the mother cell and not passed on to daughter cells. These forms are not amyloid but require Hsp70 or Hsp104 for their maintenance.

(C) The prion-like proteins described by Chakrabortee et al.,²⁶ which are induced *de novo* by overexpression of a typically low-abundance protein, although the nature of the resulting altered forms of the protein (yellow) remains to be defined but are not amyloid. Depending on the protein, they either require Hsp104, Hsp90, or Hsp70 to maintain the associated phenotype. In all three cases, the cells in green show the altered phenotype. Reproduced with permission from Tuite M.F. et al.²⁴

The inheritance of the prion-linked phenotype in fungi requires that the elongating amyloid fibrillar forms of the protein be fragmented prior to cell division by the coordinated action of three chaperones. In the case of prions, three chaperones, Hsp40, Hsp70, and Hsp104 mediate such fragmentation to generate the essential seeds.⁹¹ A similar process mediated by a functionally related trio of chaperones appears to occur in mammals as well⁹¹ and Hsp104 has also

been implicated in the formation of the other forms of protein-based epigenetic memory but here the role of chaperones is less clear.

Unlike PrPSc, some fungal prions can confer environment dependent, beneficial traits on their host rather than being detrimental or lethal.⁹² Prions have also been reported in viruses, bacteria, insects, and plants and so might represent a rare but universal mechanism for heritable modulation of protein function and diversity.⁹³

What enables a protein to switch to a transmissible prion form can be gleaned in part by interrogation of the primary amino acid sequence.⁹⁴ Examination of various prion-forming proteins suggests that they harbor one or more discrete, low-complexity regions (“domains”) that are essential for the switch to the heritable amyloid form. These prion-forming domains (PrDs) in yeast are enriched in the polar amino acids glutamine and asparagine (i.e., are QN-rich).⁹⁵ However, such a QN-rich region is absent in PrPc and in several other prion-forming proteins, suggesting that QN residues are not essential for prion formation *per se* but likely contribute to the intrinsically disordered structure.²⁶ The less

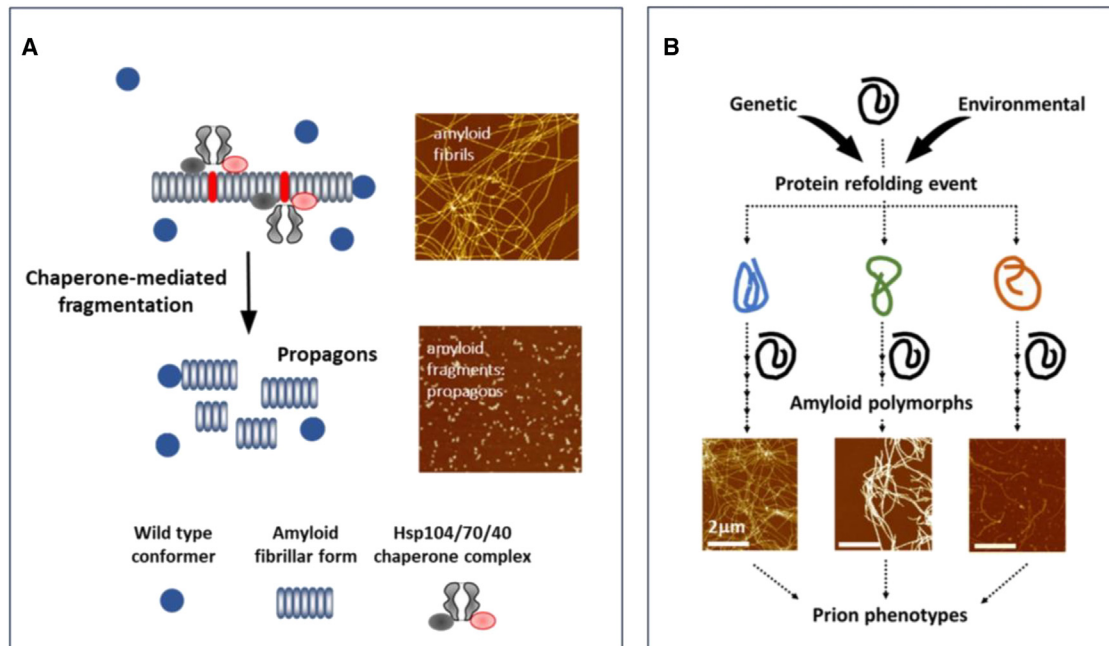


Figure 5. Prions: Proteins mediating transgenerational inheritance

(A) Certain proteins can refold to take up an amyloid form, triggered either by changes in the cellular environment or because of a mutational change in the protein's sequence. As illustrated, the resulting amyloid can take up one of several different structures (amyloid polymorphs) which in turn can give rise to distinct prion-mediated phenotypes.

(B) The transmissible form of the prion amyloid is generated by fragmentation of the amyloid fibrillar form of the protein. This fragmentation is mediated by the coordinated action of at least three different molecular chaperones and the resulting transmissible fragments are referred to as prion seeds or propagons. The atomic force microscopic (AFM) images refer to different mutant forms of the yeast Sup35 prion-forming domain (A) or the wild-type form of the same protein (B). AFM images kindly provided by Drs Wei-feng Xue and Ricardo Marchante, University of Kent.

prevalent hydrophobic and aromatic amino acids could play important roles in the assembly of prion aggregates.⁹⁶ Remarkably, several hundred proteins in eukaryotes contain sequences showing PrD characteristics and are referred to as “prion-like domains” (PrLDs). Besides possibly defining prion-forming proteins, PrLDs might be important for the formation of a variety of complex, dynamic biomolecular condensates, such as stress granules.⁹⁷

One of the most enigmatic properties of prion-forming proteins is their propensity to refold into an ensemble of alternatively folded transmissible amyloid polymorphs and their respective assembly states⁹⁸ (Figure 5A). Prion polymorphs show both structural and functional diversity and can give rise to different outcomes in terms of pathology for PrPSc in animals (prion strains) or phenotype linked to fungal prions (prion “variants”). This novel type of diversity can occur in response to genetic or environmental triggers and arise without underlying genomic changes. This suggests that there is a common thread with IDP conformational noise although mutations within the prion protein can also lead to different conformational forms of the protein (Figure 5B). Species-specific differences in prion protein sequence may also substantially reduce the potential for inter-species transfer of prions, giving rise to an all-important species barrier.⁹⁹

Prion propagation involves a self-templating mechanism whereby the prion form imposes its structure on the natively

folded or partially unfolded form of the same protein. Individual monomers are sequentially added to the ends of the growing fibrils to produce the characteristic fibrillar prion assemblies (Figure 5B). In the case of PrP, this templated conformational switch involves a major rearrangement of the PrPc structure from a largely α helical to a densely packed β -strand-rich configuration, highlighting the connection to fold-switching. However, prion fold switching is driven by multimerization, which is not required for many other fold-switching proteins.

Although the prion-templating mechanism was first proposed over three decades ago, critical details of how prions and their structural variants are formed and propagated are still lacking. There has been considerable debate about the atomic-level structural changes that occur during prion formation, and although a structure for one of the transmissible fungal prions—HET-s of *Podospora anserina*—was solved over a decade ago,¹⁰⁰ only now are high-resolution atomic structures of *ex vivo* but infectious forms of PrPSc are starting to appear.¹⁰¹ These emerging studies support the earlier proposal that PrPSc as well as other amyloids have a distinct, in parallel, in-register intermolecular β sheet (PIRIBS) architecture, with all the structural variants forming the same cross- β interactions, but with differences in, for example, the position of the β -turns.¹⁰² The availability of high-resolution prion structures, particularly of different prion strains and variants, can be expected to soon shed light on how prions variously impact their host and how

they are propagated. For example, do different conformers of a given prion protein interact differently with natural binding partners, whether proteins or nucleic acids? These concepts might also apply to major neurodegenerative diseases such as Parkinson's and Alzheimer's disease, where evidence is beginning to emerge of both transmissibility and strain diversity, suggesting that transmission might occur via a prion-like mechanism.¹⁰³ Lastly, it is worth noting that the capacity to form polymorphs is not limited to amyloid-based prions and that, these polymorphs appear to contribute to adaptive diversification of phenotypes in natural settings.¹⁰⁴

IMPACT AND IMPLICATIONS OF THE NEW PICTURE OF THE PROTEIN UNIVERSE

The discoveries of the last 30 years, in particular, demonstration of the widespread and diverse biological roles of IDPs, amyloid formation, and fold switching, amount to a paradigm shift, revealing major aspects of protein structure and dynamics that had not been fully appreciated previously. These discoveries make it clear that proteins can exist in a continuum of states from ensembles of interconverting conformers to highly stable structures. Proteins are now being looked at from new angles, such as nonlinear dynamics.¹⁹ Conformational dynamics of ordered/structured proteins has been studied previously, to understand protein folding, catalysis and allostery. However, ordered proteins have not been extensively explored from a dynamical systems perspective that is essential for the study of IDPs given the huge repertoire of conformations sampled by these proteins. Thus, the IDP conformational search process involves a higher-dimensional phase space compared to that of an ordered protein. Folding of ordered proteins is generally considered to occur in relatively simple, “minimally frustrated” or “funnel-like” energy landscapes, but the landscapes of IDPs are far more complex, “highly frustrated”, or “weakly funneled”, with numerous local free energy minima separated by low barriers. Fold-switching proteins can be considered as intermediate between typical ordered proteins and IDPs, with at least two minima reachable with comparable probabilities. Retrospectively, the study of IDPs, fold-switchers, and amyloid states seems to prompt a second look at the landscapes of protein folding in general, suggesting that these landscapes could be more rugged and complex than previously suspected.¹⁰⁵

The idea that IDPs represent edge-of-chaos systems at the boundary between order and randomness (chaos) with maximum complexity, spurred new theoretical approaches. For example, Waddington's epigenetic landscape has been invoked as a paradigm in which the landscape represents a high-dimensional state space, and cell fates are determined by stable attractors. This analogy shed new light on how IDP conformational dynamics influences cell fate decisions.¹⁹

With the advent of powerful artificial intelligence (AI)-based approaches implemented in programs, such as AlphaFold2 and ESMFold, the structures of hundreds of millions of ordered proteins can be predicted with high accuracy.^{105–107} However, IDPs, amyloids, as well as fold-switching proteins still present the prediction methods with major challenges that need to be addressed.⁶⁸

Until recently, ordered proteins had been the targets of choice for developing effective therapeutics because of their well-defined docking sites.¹⁰⁸ However, IDPs,¹⁰⁸ amyloid plaques, or their precursor oligomers¹⁰⁹ as well as fold-switching proteins¹⁰² are now increasingly being investigated as potential novel drug targets in several diseases. The discovery that numerous peptides and proteins harbor a latent potential to convert from their native functional states into intractable amyloid aggregates not only further underscores the importance of the protein structure continuum but also provides new insight into mechanisms of several devastating human diseases.⁸¹ The capability of IDPs/IDRs to form multivalent interactions and undergo LLPS represents a unique molecular mechanism underlying the emergence and biogenesis of numerous biomolecular condensates and MLOs, which are now believed to serve as means of organization and control of intracellular space.⁵²

CONCLUSIONS AND FUTURE DIRECTIONS

The discovery of the widespread and major biological importance of IDPs, fold-switching proteins, amyloids, and prions, calls for new thinking on protein structure and folding. Some IDPs are almost completely disordered, but they are not random coils and exhibit conformational preferences that have major impact on their function.¹¹⁰ The study of IDPs and fold-switching proteins and amyloids does not refute Anfinsen's postulate. Rather, these observations suggest that perhaps, the concept of sequence-determined protein folding should be extended to accommodate these substantial variations exhibited by numerous proteins. It remains true that a protein structure is encoded in its sequence, but at least for fold-switching proteins, IDPs, and amyloid-forming proteins, the same sequence gives rise to different structures depending on environmental conditions. The latest findings increasingly show that this structural malleability is far from being a rare exception but rather is a widespread phenomenon in the protein universe. A common belief that may be considered a corollary of Anfinsen's postulate is that protein function involves only small conformational shifts, whereas major structural changes represent misfolding that often leads to deleterious protein aggregation. The latest findings on fold-switching proteins and amyloid formation dispel this conviction by showing that both alternative folds of the same protein and aggregates can be functional as well as pathological. The prions that propagate amyloid formation across generations are the ultimate manifestation of these phenomena. Anecdotal evidence of this greater complexity associated with the protein structure/function paradigm was obtained many years ago, and findings of the last few years made it clear that these are not anomalies but major features of the protein universe.

ACKNOWLEDGMENTS

We thank Dr. Ariane Helou, California Institute of Technology, for her thoughtful comments on the manuscript and expert edits. To all our peers whose work we were unable to cite due to space limitations, we extend our sincere apology. Work in the Chiti laboratory was supported by the Regione Toscana (Bando Ricerca Salute 2018, PRAMA project) and Ministero dell'Università e Ricerca (projects PNRR PE8 Age-IT and PRIN 2020PBS5MJ). The work done by

Joseph W. Schafer and Lauren L. Porter was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health (LM202011). Work in the Orban laboratory was supported by NIH R01 GM62154 and GM141290 funding, in the Weninger laboratory by a NIH grant GM132263, and that in the Salgia laboratory by a NIH/NCI Cancer Center Support Grant P30CA033572-40 and a DOD grant HT9425-23-1-0581. Shasha Chong is supported by the NIH/NCI under award number P30CA016042, Pew-Stewart Scholar Award, Searle Scholar Award, The Shurl and Kay Curci Foundation Research Grant, Merkin Innovation Seed Grant, The Mallinckrodt Research Grant, The Margaret E. Early Medical Research Trust Grants, and The Alex's Lemonade Stand Foundation Innovation Grant under award number 1260879.

AUTHOR CONTRIBUTIONS

P.K. and R.S. conceptualized the manuscript; P.K., V.N.U., L.P., J.W.S., T.-F. C., S.C., F.C., M.T., and E.V.K. and wrote the first draft; J.N.O., K.R.W., J.O., F.C., E.V.K., and R.S. contributed substantially by revising the manuscript; All authors approved the submitted version and are fully accountable for every aspect of the work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Ladunga, I. (1992). Phylogenetic continuum indicates "galaxies" in the protein universe: preliminary results on the natural group structures of proteins. *J. Mol. Evol.* *34*, 358–375.
- Povolotskaya, I.S., and Kondrashov, F.A. (2010). Sequence space and the ongoing expansion of the protein universe. *Nature* *465*, 922–926.
- Dokholyan, N.V., Shakhnovich, B., and Shakhnovich, E.I. (2002). Expanding protein universe and its origin from the biological Big Bang. *Proc. Natl. Acad. Sci. USA* *99*, 14132–14136.
- Matveev, V.V. (2019). Cell theory, intrinsically disordered proteins, and the physics of the origin of life. *Prog. Biophys. Mol. Biol.* *149*, 114–130.
- Kauzmann, W.J. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* *14*, 1–63.
- Brandts, J.F. (1964). The Thermodynamics of Protein Denaturation. I. The Denaturation of Chymotrypsinogen. *J. Am. Chem. Soc.* *86*, 4291–4301.
- Brandts, J.F. (1969). In *Structure and Stability of Biological Macromolecules*, S.N. Timasheff and G.D. Fasman, eds. (Marcel Dekker), pp. 213–290.
- Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science* *181*, 223–230.
- Leopold, P.E., Montal, M., and Onuchic, J.N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA* *89*, 8721–8725.
- Onuchic, J.N., Luthey-Schulten, Z., and Wolynes, P.G. (1997). Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* *48*, 545–600.
- Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* *21*, 167–195.
- Onuchic, J.N., and Wolynes, P.G. (2004). Theory of protein folding. *Curr. Opin. Struct. Biol.* *14*, 70–75.
- Wolynes, P.G., Onuchic, J.N., and Thirumalai, D. (1995). Navigating the folding routes. *Science* *267*, 1619–1620.
- Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signaling and regulation. *Nat. Rev. Mol. Cell Biol.* *16*, 18–29.
- Porter, L.L., and Looger, L.L. (2018). Extant fold-switching proteins are widespread. *Proc. Natl. Acad. Sci. USA* *115*, 5968–5973.
- Lin, X., Eddy, N.R., Noel, J.K., Whitford, P.C., Wang, Q., Ma, J., and Onuchic, J.N. (2014). Order and disorder control the functional rearrangement of influenza hemagglutinin. *Proc. Natl. Acad. Sci. USA* *111*, 12049–12054.
- Chakravarty, D., Schafer, J.W., and Porter, L.L. (2023). Distinguishing features of fold-switching proteins. *Protein Sci.* *32*, e4596.
- Solomon, T.L., He, Y., Sari, N., Chen, Y., Gallagher, D.T., Bryan, P.N., and Orban, J. (2023). Reversible switching between two common protein folds in a designed system using only temperature. *Proc. Natl. Acad. Sci. USA* *120*, e2215418120.
- Kulkarni, P., Bhattacharya, S., Achuthan, S., Behal, A., Jolly, M.K., Kotnala, S., Mohanty, A., Rangarajan, G., Salgia, R., and Uversky, V. (2022). Intrinsically Disordered Proteins: Critical Components of the Wetware. *Chem. Rev.* *122*, 6614–6633.
- Kulkarni, P., Solomon, T.L., He, Y., Chen, Y., Bryan, P.N., and Orban, J. (2018). Structural metamorphism and polymorphism in proteins on the brink of thermodynamic stability. *Protein Sci.* *27*, 1557–1567.
- Zuber, P.K., Said, N., Hilal, T., Wang, B., Loll, B., González-Higueras, J., Ramírez-Sarmiento, C.A., Belogurov, G.A., Artsimovitch, I., Wahl, M.C., and Knauer, S. (2024). Concerted transformation of a hyper-paused transcription complex and its reinforcing protein. Preprint at bioRxiv. <https://doi.org/10.1101/2023.06.26.546588>.
- Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C.M. (1999). Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc. Natl. Acad. Sci. USA* *96*, 3590–3594.
- Maji, S.K., Perrin, M.H., Sawaya, M.R., Jessberger, S., Vadodaria, K., Rissman, R.A., Singru, P.S., Nilsson, K.P.R., Simon, R., Schubert, D., et al. (2009). Functional amyloids as the natural storage of peptide hormones in pituitary secretory granules. *Science* *325*, 328–332.
- Tuite, M.F. (2016). Remembering the Past: A New Form of Protein-Based Inheritance. *Cell* *167*, 302–303.
- Li, L., and Lindquist, S. (2000). Creating a protein-based element of inheritance. *Science* *287*, 661–664.
- Chakrabortee, S., Byers, J.S., Jones, S., Garcia, D.M., Bhullar, B., Chang, A., She, R., Lee, L., Fremin, B., Lindquist, S., and Jarosz, D.F. (2016). Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. *Cell* *167*, 369–381.e12.
- Lyon, A.S., Peeples, W.B., and Rosen, M.K. (2021). A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* *22*, 215–235.
- Jakob, U., Kriwacki, R., and Uversky, V.N. (2014). Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem. Rev.* *114*, 6779–6805.
- Zhang, Y., Stec, B., and Godzik, A. (2007). Between order and disorder in protein structures: analysis of "dual personality" fragments in proteins. *Structure* *15*, 1141–1147.
- Borgia, A., Borgia, M.B., Bugge, K., Kissling, V.M., Heidarsson, P.O., Fernandes, C.B., Sottini, A., Soranno, A., Buholzer, K.J., Nettels, D., et al. (2018). Extreme disorder in an ultrahigh-affinity protein complex. *Nature* *555*, 61–66.
- Oldfield, C.J., Meng, J., Yang, J.Y., Yang, M.Q., Uversky, V.N., and Dunker, A.K. (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genom.* *9*, S1. <https://doi.org/10.1186/1471-2164-9-S1-S1>.
- Hsu, W.L., Oldfield, C.J., Xue, B., Meng, J., Huang, F., Romero, P., Uversky, V.N., and Dunker, A.K. (2013). Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci.* *22*, 258–273.
- Sharma, R., Raduly, Z., Miskei, M., and Fuxreiter, M. (2015). Fuzzy complexes: Specific binding without complete folding. *FEBS Lett.* *589*, 2533–2542.
- Gianni, S., Freiburger, M.I., Jemth, P., Ferreira, D.U., Wolynes, P.G., and Fuxreiter, M. (2021). Fuzziness and Frustration in the Energy Landscape

- of Protein Folding, Function, and Assembly. *Acc. Chem. Res.* *54*, 1251–1259.
35. Freiburger, M.I., Wolynes, P.G., Ferreira, D.U., and Fuxreiter, M. (2021). Frustration in Fuzzy Protein Complexes Leads to Interaction Versatility. *J. Phys. Chem. B* *125*, 2513–2520.
 36. Barrera-Vilarmau, S., Teixeira, J.M.C., and Fuxreiter, M. (2022). Protein interactions: anything new? *Essays Biochem.* *66*, 821–830.
 37. Lammert, H., Schug, A., and Onuchic, J.N. (2009). Robustness and generalization of structure-based models for protein folding and function. *Proteins* *77*, 881–891.
 38. Andrews, B.T., Gosavi, S., Finke, J.M., Onuchic, J.N., and Jennings, P.A. (2008). The dual-basin landscape in GFP folding. *Proc. Natl. Acad. Sci. USA* *105*, 12283–12288.
 39. Whitford, P.C., Gosavi, S., and Onuchic, J.N. (2008). Conformational transitions in adenylate kinase. Allosteric communication reduces misligation. *J. Biol. Chem.* *283*, 2042–2048.
 40. Morcos, F., Jana, B., Hwa, T., and Onuchic, J.N. (2013). Coevolutionary signals across protein lineages help capture multiple protein conformations. *Proc. Natl. Acad. Sci. USA* *110*, 20533–20538.
 41. Bondos, S.E., Dunker, A.K., and Uversky, V.N. (2022). Intrinsically disordered proteins play diverse roles in cell signaling. *Cell Commun. Signal.* *20*, 20–26.
 42. Már, M., Nitsenko, K., and Heidarsson, P.O. (2023). Multifunctional Intrinsically Disordered Regions in Transcription Factors. *Chemistry* *29*, e202203369.
 43. Pal, S., and Dhar, R. (2024). Living in a noisy world—origins of gene expression noise and its impact on cellular decision-making. *FEBS Lett.* *598*, 1673–1691.
 44. Tripathi, S., Kessler, D.A., and Levine, H. (2023). Minimal frustration underlies the usefulness of incomplete regulatory network models in biology. *Proc. Natl. Acad. Sci. USA* *120*, e2216109120.
 45. Mahmoudabadi, G., Rajagopalan, K., Getzenberg, R.H., Hannenhalli, S., Rangarajan, G., and Kulkarni, P. (2013). Intrinsically disordered proteins and conformational noise: implications in cancer. *Cell Cycle* *12*, 26–31.
 46. Uversky, V.N. (2018). Intrinsic Disorder, Protein-Protein Interactions, and Disease. *Adv. Protein Chem. Struct. Biol.* *110*, 85–121.
 47. Kulkarni, P., Salgia, R., and Rangarajan, G. (2023). Intrinsically disordered proteins and conformational noise: The hypothesis a decade later. *iScience* *26*, 107109.
 48. Uversky, V.N. (2021). Recent Developments in the Field of Intrinsically Disordered Proteins: Intrinsic Disorder-Based Emergence in Cellular Biology in Light of the Physiological and Pathological Liquid-Liquid Phase Transitions. *Annu. Rev. Biophys.* *50*, 135–156.
 49. Hirose, T., Ninomiya, K., Nakagawa, S., and Yamazaki, T. (2023). A guide to membraneless organelles and their various roles in gene regulation. *Nat. Rev. Mol. Cell Biol.* *24*, 288–304.
 50. Darling, A.L., and Uversky, V.N. (2023). Known types of membrane-less organelles and biomolecular condensates. In *Droplets of life* (Academic Press), pp. 271–335.
 51. Darling, A.L., Liu, Y., Oldfield, C.J., and Uversky, V.N. (2018). Intrinsically Disordered Proteome of Human Membrane-Less Organelles. *Proteomics* *18*, e1700193.
 52. Antifeeva, I.A., Fonin, A.V., Fefilova, A.S., Stepanenko, O.V., Povarova, O.I., Silonov, S.A., Kuznetsova, I.M., Uversky, V.N., and Turoverov, K.K. (2022). Liquid-liquid phase separation as an organizing principle of intracellular space: overview of the evolution of the cell compartmentalization concept. *Cell. Mol. Life Sci.* *79*, 251–283.
 53. Turoverov, K.K., Kuznetsova, I.M., Fonin, A.V., Darling, A.L., Zaslavsky, B.Y., and Uversky, V.N. (2019). Stochasticity of Biological Soft Matter: Emerging Concepts in Intrinsically Disordered Proteins and Biological Phase Separation. *Trends Biochem. Sci.* *44*, 716–728.
 54. Bhat, P., Honson, D., and Guttman, M. (2021). Nuclear compartmentalization as a mechanism of quantitative control of gene expression. *Nat. Rev. Mol. Cell Biol.* *22*, 653–670.
 55. Chong, S., and Mir, M. (2021). Towards Decoding the Sequence-Based Grammar Governing the Functions of Intrinsically Disordered Protein Regions. *J. Mol. Biol.* *433*, 166724.
 56. Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., and Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* *361*, eaar2555.
 57. Figueiredo, A.S., Loureiro, J.R., Macedo-Ribeiro, S., and Silveira, I. (2023). Advances in Nucleotide Repeat Expansion Diseases: Transcription Gets in Phase. *Cells* *12*, 826.
 58. Tuinstra, R.L., Peterson, F.C., Kutlesa, S., Elgin, E.S., Kron, M.A., and Volkman, B.F. (2008). Interconversion between two unrelated protein folds in the lymphotactin native state. *Proc. Natl. Acad. Sci. USA* *105*, 5057–5062.
 59. Tseng, R., Goularte, N.F., Chavan, A., Luu, J., Cohen, S.E., Chang, Y.G., Heisler, J., Li, S., Michael, A.K., Tripathi, S., et al. (2017). Structural basis of the day-night transition in a bacterial circadian clock. *Science* *355*, 1174–1180.
 60. Burmann, B.M., Knauer, S.H., Sevostyanova, A., Schweimer, K., Mooney, R.A., Landick, R., Artsimovitch, I., and Rösch, P. (2012). An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* *150*, 291–303.
 61. Podobnik, M., Savory, P., Rojko, N., Kisovec, M., Wood, N., Hambley, R., Pugh, J., Wallace, E.J., McNeill, L., Bruce, M., et al. (2016). Crystal structure of an invertebrate cytolysin pore reveals unique properties and mechanism of assembly. *Nat. Commun.* *7*, 11598.
 62. Gordon, D.E., Hiatt, J., Bouhaddou, M., Rezelj, V.V., Ulferts, S., Braberg, H., Jureka, A.S., Obernier, K., Guo, J.Z., Batra, J., et al. (2020). Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* *370*, eaabe9403.
 63. Kim, A.K., and Porter, L.L. (2021). Functional and Regulatory Roles of Fold-Switching Proteins. *Structure* *29*, 6–14.
 64. Murzin, A.G. (2008). Biochemistry. Metamorphic proteins. *Science* *320*, 1725–1726.
 65. Ruan, B., He, Y., Chen, Y., Choi, E.J., Chen, Y., Motabar, D., Solomon, T., Simmerman, R., Kauffman, T., Gallagher, D.T., et al. (2023). Design and characterization of a protein fold switching network. *Nat. Commun.* *14*, 431.
 66. Porter, L.L. (2023). Fluid protein fold space and its implications. *Bioessays* *45*, e2300057.
 67. Chakravarty, D., Sreenivasan, S., Swint-Kruse, L., and Porter, L.L. (2023). Identification of a covert evolutionary pathway between two protein folds. *Nat. Commun.* *14*, 3177.
 68. Chakravarty, D., and Porter, L.L. (2022). AlphaFold2 fails to predict protein fold switching. *Protein Sci.* *31*, e4353.
 69. Dishman, A.F., and Volkman, B.F. (2023). Metamorphic protein folding as evolutionary adaptation. *Trends Biochem. Sci.* *48*, 665–672.
 70. Schafer, J.W., and Porter, L.L. (2023). Evolutionary selection of proteins with two folds. *Nat. Commun.* *14*, 5478.
 71. Sala, D., Engelberger, F., Mchaourab, H.S., and Meiler, J. (2023). Modeling conformational states of proteins with AlphaFold. *Curr. Opin. Struct. Biol.* *81*, 102645.
 72. Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R.J., Milles, L.F., et al. (2023). De novo design of protein structure and function with RF diffusion. *Nature* *620*, 1089–1100.
 73. Jing, B., Erives, E., Pao-Huang, P., Corso, G., Berger, B., and Jaakkola, T. (2023). EigenFold: Generative Protein Structure Prediction with Diffusion Models. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2304.02198>.

74. Parui, S., Brini, E., and Dill, K.A. (2023). Computing Free Energies of Fold-Switching Proteins Using MELD x MD. *J. Chem. Theor. Comput.* **19**, 6839–6847.
75. Westermark, P., Araki, S., Benson, M.D., Cohen, A.S., Frangione, B., Masters, C.L., Saraiva, M.J., Sipe, J.D., Husby, G., Kyle, R.A., and Selkoe, D. (1999). Nomenclature of amyloid fibril proteins. Report from the meeting of the International Nomenclature Committee on Amyloidosis, August 8–9, 1998. Part 1. *Amyloid* **6**, 63–66.
76. Guijarro, J.I., Sunde, M., Jones, J.A., Campbell, I.D., and Dobson, C.M. (1998). Amyloid fibril formation by an SH3 domain. *Proc. Natl. Acad. Sci. USA* **95**, 4224–4228.
77. Olsén, A., Jonsson, A., and Normark, S. (1989). Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coli*. *Nature* **338**, 652–655.
78. Fowler, D.M., Koulov, A.V., Alory-Jost, C., Marks, M.S., Balch, W.E., and Kelly, J.W. (2006). Functional amyloid formation within mammalian tissue. *PLoS Biol.* **4**, e6.
79. Sinnige, T., Yu, A., and Morimoto, R.I. (2020). Challenging Proteostasis: Role of the Chaperone Network to Control Aggregation-Prone Proteins in Human Disease. *Adv. Exp. Med. Biol.* **1243**, 53–68.
80. Jahn, T.R., and Radford, S.E. (2008). Folding versus aggregation: polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **469**, 100–117.
81. Chiti, F., and Dobson, C.M. (2017). Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68.
82. Chapman, M.R., Robinson, L.S., Pinkner, J.S., Roth, R., Heuser, J., Hammar, M., Normark, S., and Hultgren, S.J. (2002). Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **295**, 851–855.
83. Fowler, D.M., Koulov, A.V., Balch, W.E., and Kelly, J.W. (2007). Functional amyloid—from bacteria to humans. *Trends Biochem. Sci.* **32**, 217–224.
84. Li, J., McQuade, T., Siemer, A.B., Napetschnig, J., Moriwaki, K., Hsiao, Y.S., Damko, E., Moquin, D., Walz, T., McDermott, A., et al. (2012). The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis. *Cell* **150**, 339–350.
85. Knowles, T.P.J., Waudby, C.A., Devlin, G.L., Cohen, S.I.A., Aguzzi, A., Vendruscolo, M., Terentjev, E.M., Welland, M.E., and Dobson, C.M. (2009). An analytical solution to the kinetics of breakable filament assembly. *Science* **326**, 1533–1537.
86. Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C.M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805–808.
87. Sawaya, M.R., Hughes, M.P., Rodriguez, J.A., Riek, R., and Eisenberg, D.S. (2021). The expanding amyloid family: Structure, stability, function, and pathogenesis. *Cell* **184**, 4857–4873.
88. Prusiner, S.B. (1998). Prions. *Proc. Natl. Acad. Sci. USA* **95**, 13363–13383.
89. Liebman, S.W., and Chernoff, Y.O. (2012). Prions in yeast. *Genetics* **191**, 1041–1072.
90. Caudron, F., and Barral, Y. (2013). A super-assembly of Whi3 encodes memory of deceptive encounters by single cells during yeast courtship. *Cell* **155**, 1244–1257.
91. Kushnirov, V.V., Dergalev, A.A., and Alexandrov, A.I. (2021). Amyloid Fragmentation and Disaggregation in Yeast and Animals. *Biomolecules* **11**, 1884.
92. Harvey, Z.H., Chen, Y., and Jarosz, D.F. (2018). Protein-Based Inheritance: Epigenetics beyond the Chromosome. *Mol. Cell* **69**, 195–202.
93. Galkin, A.P., Sysoev, E.I., and Valina, A.A. (2023). Amyloids and prions in the light of evolution. *Curr. Genet.* **69**, 189–202.
94. Cascarina, S.M., and Ross, E.D. (2014). Yeast prions and human prion-like proteins: sequence features and prediction methods. *Cell. Mol. Life Sci.* **71**, 2047–2063.
95. Tuite, M.F. (2000). Yeast prions and their prion-forming domain. *Cell* **100**, 289–292.
96. Gonzalez Nelson, A.C., Paul, K.R., Petri, M., Flores, N., Rogge, R.A., Cascarina, S.M., and Ross, E.D. (2014). Increasing prion propensity by hydrophobic insertion. *PLoS One* **9**, e89286.
97. Fomicheva, A., and Ross, E.D. (2021). From Prions to Stress Granules: Defining the Compositional Features of Prion-Like Domains That Promote Different Types of Assemblies. *Int. J. Mol. Sci.* **22**, 1251.
98. Carta, M., and Aguzzi, A. (2022). Molecular foundations of prion strain diversity. *Curr. Opin. Neurobiol.* **72**, 22–31.
99. Igel-Egalon, A., Béringue, V., Rezaei, H., and Sibille, P. (2018). Prion Strains and Transmission Barrier Phenomena. *Pathogens* **7**, 5.
100. Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A.B., Riek, R., and Meier, B.H. (2008). Amyloid fibrils of the HET-s(218–289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319**, 1523–1526.
101. Manka, S.W., Wenborn, A., Betts, J., Joiner, S., Saibil, H.R., Collinge, J., and Wadsworth, J.D.F. (2023). A structural basis for prion strain diversity. *Nat. Chem. Biol.* **19**, 607–613.
102. Cobb, N.J., Sönnichsen, F.D., McHaourab, H., and Surewicz, W.K. (2007). Molecular architecture of human prion protein amyloid: a parallel, in-register beta-structure. *Proc. Natl. Acad. Sci. USA* **104**, 18946–18951.
103. Brundin, P., Melki, R., and Kopito, R. (2010). Prion-like transmission of protein aggregates in neurodegenerative diseases. *Nat. Rev. Mol. Cell Biol.* **11**, 301–307.
104. Itakura, A.K., Chakravarty, A.K., Jakobson, C.M., and Jarosz, D.F. (2020). Widespread Prion-Based Control of Growth and Differentiation Strategies in *Saccharomyces cerevisiae*. *Mol. Cell* **77**, 266–278.e6.
105. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596.
106. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130.
107. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.
108. Saurabh, S., Nadendla, K., Purohit, S.S., Sivakumar, P.M., and Cetinel, S. (2023). Fuzzy Drug Targets: Disordered Proteins in the Drug-Discovery Realm. *ACS Omega* **8**, 9729–9747.
109. van Dyck, C.H., Swanson, C.J., Aisen, P., Bateman, R.J., Chen, C., Gee, M., Kanekiyo, M., Li, D., Reyderman, L., Cohen, S., et al. (2023). Lecanemab in Early Alzheimer’s Disease. *N. Engl. J. Med.* **388**, 9–21.
110. Kulkarni, P., Jolly, M.K., Jia, D., Mooney, S.M., Bhargava, A., Kagohara, L.T., Chen, Y., Hao, P., He, Y., Veltri, R.W., et al. (2017). Phosphorylation-induced conformational dynamics in an intrinsically disordered protein and potential role in phenotypic heterogeneity. *Proc. Natl. Acad. Sci. USA* **114**, E2644–E2653.