Invited Review

Under-Folded Proteins: Conformational Ensembles and Their Roles in Protein Folding, Function, and Pathogenesis

Vladimir N. Uversky[1,2]

[1] *Department of Molecular Medicine, USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612*

[2] *Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino 142292, Moscow Region, Russia*

Received 20 April 2013; revised 21 May 2013; accepted 30 May 2013
Published online 10 June 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22298

ABSTRACT:

For decades, protein function was intimately linked to the presence of a unique, aperiodic crystal-like structure in a functional protein. The two only places for conformational ensembles of under-folded (or partially folded) protein forms in this picture were either the end points of the protein denaturation processes or transiently populated folding intermediates. Recent years witnessed dramatic change in this perception and conformational ensembles, which the under-folded proteins are, have moved from the shadow. Accumulated to date data suggest that a protein can exist in at least three global forms–functional and folded, functional and intrinsically disordered (nonfolded), and nonfunctional and misfolded/aggregated. Under-folded protein states are crucial for each of these forms, serving as important folding intermediates of ordered proteins, or as functional states of intrinsically disordered proteins (IDPs) and IDP regions (IDPRs), or as pathology triggers of misfolded proteins. Based on these observations, conformational ensembles of under-folded proteins can be classified as transient (folding and misfolding intermediates) and permanent (IDPs and stable misfolded proteins). Permanently under-folded proteins can further be split into intentionally designed (IDPs and IDPRs) and unintentionally designed (misfolded proteins). Although intrinsic flexibility, dynamics, and pliability are crucial for all under-folded proteins, the different categories of under-foldedness are differently encoded in protein amino acid sequences. © 2013 Wiley Periodicals, Inc. Biopolymers 99: 870–887, 2013.

Keywords: *protein folding; protein misfolding; intrinsically disordered protein; conformational ensemble*

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

INTRODUCTION

According to the classical structure-function paradigm, a specific function of a protein is determined by its unique 3D structure, which can be considered as an aperiodic crystal. For a small globular protein, all the information on how to gain functional 3D structure is encoded in its amino acid sequence.[1,2] This hypothesis represents a foundation of the "one sequence-one structure-one function" model, which is the cornerstone of modern structural biology.[1–3] This structural rigidity of ordered proteins determined their ability to form crystals, which allowed the X-ray-based determination of 3D-structure of many proteins down to the atomic resolution.[4] The protein misfolding phenomenon, when due to the

Correspondence to: Vladimir N. Uversky; e-mail: vuversky@health.usf.edu

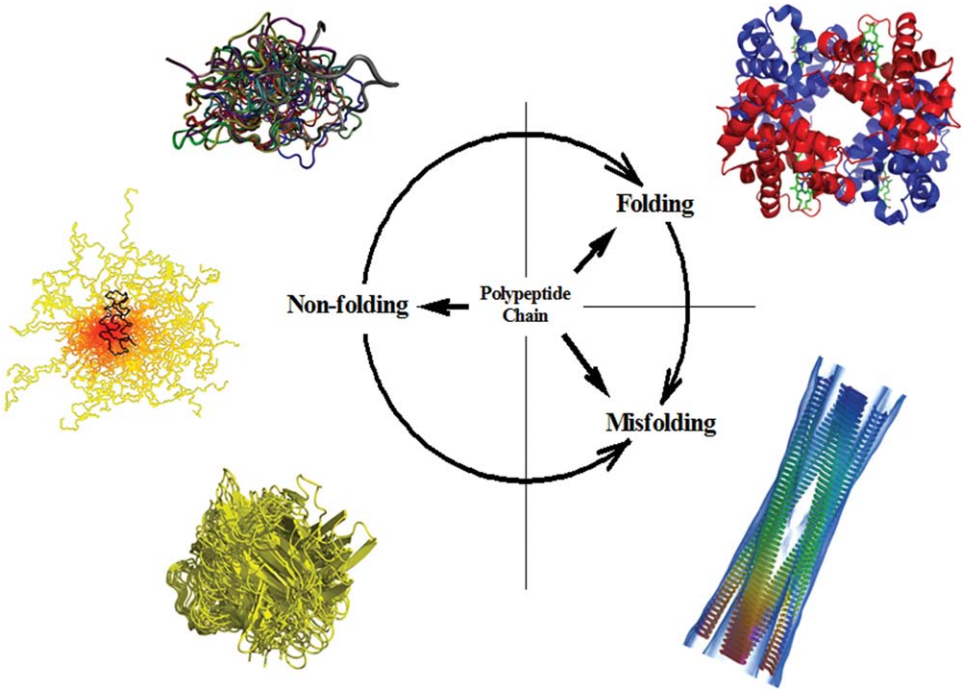

FIGURE 1 Fate of a newly synthesized polypeptide chain in a cell.

effect of environmental factors or because of the genetic defects (mutations), a polypeptide chain has lost its capability to gain a proper functional 3D structure (i.e., became misfolded), and which has multiple detrimental consequences (such as lost of function, gain of toxic function, aggregation, disbalance in proteostasis, potential cell death, etc.) that constitute molecular basis of various conformational diseases, seems to support this concept.[5–7] However, the recent revelation of countless examples of intrinsically disordered proteins (IDPs) and hybrid protein containing ordered domains and IDP regions (IDPRs) has cast doubt on the general validity of the structure-function paradigm and revealed an intriguing route of functional disorder.[8–26]

These findings clearly show that there are at least three globally different forms accessible to a protein in a living cell [functional and folded, functional and intrinsically disordered (nonfolded), and nonfunctional and misfolded], and that the propensities to be in one of these forms are encoded in the protein's amino acid sequence.[19,27,28] Therefore, a polypeptide chain is constantly facing a choice between three potential routes, nonfolding, folding, and misfolding, with the last two representing competitive routes to higher structural order (see Figure 1).[19,27] For a single-chain protein, folding, nonfolding, and misfolding pathways represents a choice of each individual molecule, whereas unproductive protein aggregation/fibrillation (that frequently follows protein misfolding and is often associated with the pathogenesis of several diseases) and functional oligomerization, and formation of various functional high order complexes is a fate of the ensemble of molecules.

Multiple factors, originating from the peculiarities of protein amino acid sequence and/or features of protein environment, might affect the choice between folding, misfolding, and nonfolding. At given environmental conditions, the primary selection between folding and nonfolding is determined only by the amino acid composition. For example, an abnormally highly charged polypeptide with low overall hydrophobicity will not fold, giving rise to an extended IDP (also known as natively unfolded protein), whereas a polypeptide chain with a balanced distribution of polar and hydrophobic residues will choose the folding path at the identical conditions. However, some changes in the amino acid sequence (point mutations) may favor the misfolding pathway for both the natively unfolded and the natively folded proteins.

Importantly, for a given polypeptide chain, a chosen fate is not a final one and a choice may be further modulated by the environmental pressure (Figure 1).[19] For example, IDPs may be forced to fold or misfold via the modification of their environment (addition of natural binding partners, changes in properties of solvent, etc.), whereas a destabilizing environment may push an ordered protein to the misfolding route. Alternatively, the presence of chaperones may reverse the misfolding route and effectively dissolve small aggregates.[29]

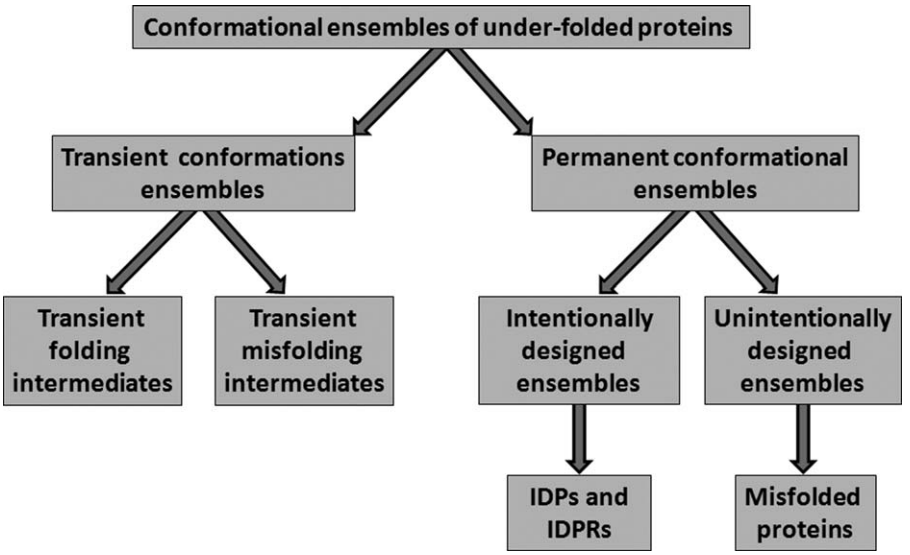

FIGURE 2 Diversity of conformational ensembles of under-folded proteins.

Another important point is that the pathological misfolding of extended IDPs to some extent resembles the process of normal protein folding and assembly, that is, it represents a way from a simple, flexible, and disordered conformation (mostly structure-less polypeptide chain) via somehow more ordered partially folded intermediate(s), to a complex and rigid structure, for example, amyloid fibril. However, the pathological misfolding of a rigid globular protein involves a step of transient disordering and formation of a partially unfolded intermediate, which is followed by the subsequent increase in the order originated from the formation of specific protein aggregates.

Besides discussed above considerations, these recent developments re-emphasized the biological importance of the under-folded protein conformations (or partially folded protein species). Such under-folded proteins do not have unique well-defined 3D structures existing instead as collapsed or extended dynamically mobile conformational ensembles. In classical structure-to-function paradigm, under-folded entities without unique structure were mostly of academic interest, since they would be typically found at the end of denaturation processes under the highly nonphysiological conditions or as transiently populated folding intermediates. However, in a new view of correlations between protein structure, function, and dysfunction, one can find important implementations of under-folded states for each of the major protein forms, functional and folded, nonfunctional and misfolded, and functional and intrinsically disordered. Here, under-folded protein states serve as important folding intermediates of ordered proteins, or as functional states of IDPs and IDPRs, or as pathology triggers of some misfolded proteins. Based on their origin, conformational ensembles of under-folded proteins can be classified as transient (folding and misfolding intermediates) and permanent (IDPs and stable misfolded proteins; see Figure 2). Permanently under-folded proteins can further be split into intentionally designed (IDPs and IDPRs) and unintentionally designed (misfolded proteins). These different categories of under-foldedness are differently encoded in protein amino acid sequences and play different roles in protein life. Sections below contain brief discussions of various roles of conformational ensembles in protein folding, misfolding, and nonfolding.

CONFORMATIONAL ENSEMBLES AND PROTEIN FOLDING

The ability of ordered proteins to adopt their functional highly structured states in the intracellular environment during/after biosynthesis on the ribosome is one of the most remarkable evolutionary achievements of biology. In this view, protein folding is taken as crucial continuation of protein biosynthesis process, where the information encoded in the DNA/mRNA nucleotide sequence is read step-by-step, and the corresponding amino acids are gathered one after another into the polypeptide chain that eventually folds into unique functional structure. In other words, during these processes, the one-dimensional information encoded in the DNA nucleotide sequence is sequentially transformed into the one-dimensional information of the protein amino acid sequence, which codes for the peculiarities of protein folding, that is, a specific way of

gaining unique three-dimensional structure. As the interactions between remote amino acid residues play a crucial role in protein folding, this process obviously deviates from the linear information transduction. Therefore, protein folding can be regarded as a second part of the genetic code, as the protein amino acid sequence contains information about its functional 3D structure.

Many proteins have rigid globular structures in aqueous solutions and are functional only in this state. The native state of these proteins is a unique conformation, which is entropically unfavorable since it has significant restrictions of the conformational freedom. However, the unfolded state of a polypeptide chain is entropically favorable, representing a dynamic ensemble of a large number of conformations originating from the main chain rotational isomerization around *F* and *Y* angles. Therefore, the possibility of a given polypeptide chain to fold into a compact state is determined by its ability to form numerous intramolecular contacts of different physical nature, to compensate the free energy increase due to the decrease in the entropy component.[30]

The first direct evidence that all the information necessary for a given polypeptide chain to fold into a unique tertiary structure is encoded in protein's amino acid sequence was obtained by Anfinsen's group,[1] who showed that the reduced and urea-denatured ribonuclease A was able to completely restore its native structure and functional state after the removal of the denaturant and the reducing agent. Later, the capability to regain the native structure in vitro was demonstrated for a variety of proteins. In recent years, our understanding of the mechanisms of the protein self-organization process has increased dramatically.[19,27,31–36] It is recognized now that only some amino acid residues are crucial for protein folding. Therefore, proteins with very low sequence identity/homology can have similar structures, whereas a single amino acid replacement can significantly affect the rate of protein folding, or in some extreme cases, can completely halt the correct protein folding.[27]

For a very long time, one of the most essential questions in protein science was how an unstructured polypeptide folds into a unique native protein with specific biological function in a reasonable period of time despite the fact that there is an astronomically large number of possible conformational states.[37] To resolve this problem, a framework model of protein folding (also known as sequential mechanism of protein folding) was proposed by Oleg Ptitsyn in 1973 (see Figure 3).[38] According to this model, the folding of a globular protein from its unfolded state represents a multistage process accompanied by the formation of several folding intermediates (each is represented as specific conformational ensemble) with the increasing level of structural complexity. The first stage results in the formation of the fluctuating secondary structure elements. These elements then collapse to form a compact but highly dynamic intermediate with the native-like secondary structure, where the backbone movements are mostly restricted but the mobility of the side chains is still high. At the final stage, the unique 3D-structure is formed by restricting the side chain mobility.[38] Therefore, partially folded species with increasing degree of structural complexity were proposed to serve as universal folding intermediates. In a due time, one of these folding intermediate later named the 'molten globule (MG) state'[39] was found in a test tube.[40] Other partially folded intermediates [e.g., premolten globule (pre-MG) and highly ordered MG] were later found.[19]

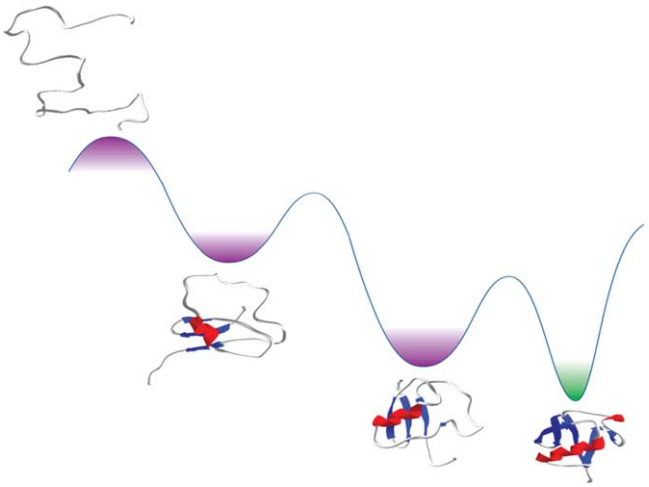

FIGURE 3 An oversimplified representation of a protein folding landscape.

According to the current view, protein folding is a more complex process, where the transition from the unfolded state to the uniquely folded native state can be realized via different pathways that are determined by the protein's energy landscape.[41,42] This complex landscape shows the dependence of the free energy on all the coordinates determining the protein conformation. Since the free energy of unfolded polypeptide chain represents a large "hilly plateau" describing the dynamic ensemble of a large number of conformations, and since the number of conformational states accessible by a polypeptide chain is reduced while approaching the native state, the resulting energetic surface is known as the "energy funnel" model. The conformational ensemble of unfolded conformations is separated from the entrance to the folding funnel by high energetic barrier(s) corresponding to the transitional state(s).[30] This barrier is of great importance for the proper protein functioning, as its existence guarantees the structural identity of all the native protein molecules. The ability of native globular proteins to form crystals is the major proof of this hypothesis.[27]

It is now generally accepted that protein folding involves discrete pathways with distinct intermediate steps. In this view, the role of under-folded protein species is in helping proteins to fold. As a result, the experimental and theoretical studies on protein folding were traditionally centered on the search for and structural characterization of partially folded intermediates as a route to defining pathways of protein folding.[43] There is considerable support for the idea that equilibrium partially folded conformations of a protein molecule can be good models for transient kinetic intermediates in protein folding.[19,44–63] Therefore discovery and structural characterization of such equilibrium conformational ensembles is believed to considerably facilitate the description of the structural properties of short-lived kinetic (transient) intermediates. The fact that the partially folded forms at moderate guanidinium hydrochloride (GdmHCl) or urea concentrations usually can be obtained only in mixture with native and/or unfolded forms, whereas the acid forms of globular proteins can be studied in pure state, makes partially folded conformations induced by extremely low (or extremely high) pH values very attractive targets for such structural studies.[40]

Close look at the Figure 3, which shows an oversimplified framework model (which can be taken as one of the vertical slices through the folding funnel), indicates that the protein folding process can be considered as a set of conformational transitions between several intermediate states. One should keep in mind, however, that in this context the term "intermediate state" has a very loose meaning since none of these partially folded forms represents a specific state with unique structure, but each of these forms should be considered as a dynamic conformational ensemble. Therefore, the majority of experimental techniques used for the structural characterization of these ensembles provide observables that are by definition statistical averages over the ensemble of conformations accessible to a protein. Since proteins are evolutionary edited random polypeptides,[52,64,65] the understanding of the common physicochemical principles underlying the protein folding process relies on the delineation of the common polymer roots and their impact on the protein structures.

The traditional way of such an analysis is a determination of the correlation between different physical characteristics of a polymer (e.g., its molecular density) and its length. Implementation of such analysis to a set of proteins in a variety of conformational states established a correlation between the hydrodynamic dimensions and the length of polypeptide chain.[17–19,27,61,62] The analyzed protein categories included ordered globular proteins with nearly spherical shapes; equilibrium MG; pre-MG; denaturant-unfolded proteins without crosslinks in the presence of strong denaturants (8*M* urea or 6*M* GdmHCl); and extended IDPs (native coils and native pre-MG). In all the cases, a correlation between the apparent molecular density (determined as $\rho = M/(4\pi R_S^3/3)$, where M is a molecular mass and R_S is a hydrodynamic radius of a given protein) and molecular mass was observed that gave rise to a set of the standard equations, $R_S = K_h M^{\varepsilon}$ (here, K_h is a constant related to the persistence length and ε is a scaling factor that depends on solvent quality), for a number of conformational states of a polypeptide chain.[17–19,61] Therefore, for a given conformational state, parameters K_h and ε were invariable over a wide range of chain lengths suggesting that the effective protein dimensions in a variety of conformational states can be predicted based on the chain length with an accuracy of 10%.[17–19,61] Thus, regardless of the differences in the amino acid sequences and biological functions, protein molecules behave as polymer homologues in a number of conformational states.

DIVERSITY OF CONFORMATIONAL ENSEMBLES INVOLVED IN PROTEIN FOLDING PROCESS

The unique 3D structure of a globular protein is stabilized by a set of noncovalent interactions of different nature. These include hydrogen bonds, hydrophobic interactions, electrostatic interactions, Van der Waals interactions, etc. Complete (or almost complete) disruption of all these interactions can be achieved in concentrated solutions of strong denaturants (such as urea or GdmHCl). Here, an initially folded and highly ordered molecule of a globular protein unfolds, that is, transforms into a highly disordered random coil-like conformation.[66–69] However, environmental changes can decrease (or eliminate) only some noncovalent interactions, whereas the remaining interactions could stay unchanged (or even could be intensified). Very often, a globular protein will lose its biological activity under these conditions, thus becoming denatured.[69] It is important to remember that denaturation is not necessarily accompanied by the unfolding of a protein, but rather might result in the appearance of various partially folded conformations with properties intermediate between those of the folded (ordered) and the completely unfolded states. In fact, globular proteins exist in at least four different equilibrium conformations: folded (ordered), MG, pre-MG and unfolded.[19,51,52,58–60,62,70] The ability of a globular protein to adopt different stable partially folded conformations each of which represents a specific conformational ensemble is believed to be an intrinsic property of a polypeptide chain.

Conformational Ensembles of Unfolded States

The unfolded state represents the starting point of the protein folding reaction. This state represents an ensemble of

rapidly interchanging conformations, some of which are extended, and some more compact. It is possible that when stabilizing interactions occur they induce a more populated ensemble of chain conformations, and, if such structures exist in the unfolded state, they would probably guide the folding process and function as folding-initiation sites.[71] In fact, theoretical studies revealed that small preferences for native-like interactions in the unfolded state will substantially increase the probability of reaching the native state.

Coming back to the polymer roots, under conditions known as "ideal" or "θ-conditions," that is, when the attractions of the macromolecular segments are balanced by those with the solvent, the density of macromolecules is expected to follow $M^{-0.5}$, thereby, the $R_S = lN^{0.5}$ with l being a statistical chain length and N being a number of amino acid residues in a protein.[69,72,73] Here, the polymer is assumed to be in a random coil conformation, and its conformational behavior can be described with the Gaussian statistics.[72] Further, in a good solvent, the macromolecular coil is expanded due to the prevalence of the repulsive interactions between polymer segments, the molecular dimensions change more significantly with increasing chain length, $R_S = (l^2B)^{0.2}\ N^{0.6}$, where B is the second virial coefficient that characterizes the pair collisions of the monomer units of the polymer chain.

Based on the mentioned above analysis of the protein molecular density in various conformations and its length it has been concluded that the "fully" unfolded states induced by the GdmHCl or urea provide $\varepsilon = 0.54$ and 0.52, respectively.[69,72,73] Given that these ε-values are <0.6, it appears that the unfolded polypeptide chains under these conditions exhibit features of macromolecular coils in θ-solvents. Recently, this conclusion was further supported by the examination of the correlation between the denatured-state radii of gyration, R_g, of 26 proteins and their polypeptide lengths ranging from 16 to 549 residues.[74] This analysis revealed that the dimensions of most chemically denatured proteins scale with polypeptide length by means of the power-law relationship with a best-fit exponent, 0.598 ± 0.028, coinciding closely with the 0.588 predicted for an excluded volume random coil. Based on these observations it has been concluded that the mean dimensions of the chemically denatured proteins are effectively indistinguishable from the mean dimensions of a random-coil ensemble.[74]

However, the values of the hydrodynamic dimensions, which Tanford measured for the unfolded proteins[69] correspond better to a model where 20% of the residues are located in the collapsed structures.[75] In agreement with these observations, more recent analysis showed that presence of ~20% α-helix generated the unfolded state with the experimentally observed radii of gyration.[76] Furthermore, it has been pointed out that the inclusion of "knots" of collapsed structure into the random coil model would not have a great influence on the hydrodynamic dimensions of a coil.[77] In fact, analysis of model systems where several proteins of known structure were used to computationally generate disordered conformers by varying backbone torsion angles at random for ~8% of the residues, with the remaining ~92% of the residues being remained fixed in their native conformations, revealed that despite this extreme degree of imposed internal structure, the analyzed conformational ensembles had end-to-end distances and mean radii of gyration that agree well with random-coil expectations.[77]

Such theoretical evaluations are supported by rich experimental observations, where noticeable residual structure is seen in unfolded proteins even under the most severe denaturing conditions, such as high concentrations of strong denaturants. Among the illustrative examples of well-characterized unfolded globular proteins with considerable residual structure are staphylococcal nuclease,[78–85] the α-subunit of tryptophan synthetase,[86,87] fragment of the protein 434,[88–90] human fibroblast growth factor 1,[91] the SH3 domain,[92,93] barstar,[94] barnase,[95] the WW-domain,[96] BPTI,[97,98] chymotrypsin inhibitor 2,[99] human carbonic anhydrase II[100–102] apomyoglobin,[103] lysozyme,[104] photoactive yellow protein,[105] the *Escherichia coli* outer membrane protein X,[106] the N-terminal domain of enzyme I from *Streptomyces coelicolor*,[107] bovine and human α-lactalbumins,[108] protein eglin C,[109] intestinal fatty acid binding protein,[110] yeast alcohol dehydrogenase,[111] HIV-1 protease,[112] "Trp-cage" miniprotein TC5b,[113] *Bacillus licheniformis* β-lactamase,[114] hyperthermophilic ribosomal protein S16,[115] thermophilic ribonucleases H,[116] and ubiquitin[117] among many other examples.

Therefore, the existence of profound residual structure might be a general characteristic of unfolded polypeptide chain under the aggressively denaturing conditions.[118–122] Therefore, unfolded states of proteins exhibit behavior that is not random coil in nature, which is not surprising considering the complexity of polypeptides. In fact, it has been pointed out that a total lack of intraresidue interactions would be unexpected in the unfolded state, because certain (e.g., hydrophobic) side chains have noticeable affinity for each other in an unfolded protein,[102,123] and some secondary structure elements could be expected within unfolded protein due to the preferential distribution of F and Y angles.[124–126] All this considerably restricts the conformational space of the unfolded polypeptide chain. Thus, it seems most likely that the polypeptide chains under the "strong denaturing conditions" are still below the critical point (bad solvent conditions), and can be easily transformed to the compact state.

Conformational Ensembles of Nonglobular Pre-MG States

When the thermodynamic quality of the solvent worsens, the binary interactions between the monomers become mainly attractive.[61] As a result, the probability of many-body interactions increases, which leads to the increase in the molecular density and partial collapse of the polymer chain. For ordered protein originally unfolded by high concentrations of strong denaturants, this typically correlates with transition to lower denaturant concentrations, and many globular proteins can form a specific compact partially folded conformation, a pre-MG state under the appropriate conditions.[53,58–62,127–133] This conformational ensemble is characterized by considerable secondary structure, although much less pronounced than that of the MG. The pre-MG state is considerably less compact than the MG, but it is still more compact than the random coil of similar molecular mass. Individual molecules within the pre-MG conformational ensemble contain some hydrophobic clusters, as evidenced from their increased propensity to interact with the hydrophobic fluorescent probes, such as 8-anilinonaphthalene-1-sulfonate, ANS.

Analysis of hydrodynamic data reveals that the molecular dimensions of pre-MGs follow the chain length as $R_S = 0.6M^{0.40}$.[17–19,61,62] The fact that for this state, $\varepsilon = 0.40$ is noticeably smaller than ε of 0.50 expected for the random coil, indicates the bad solvent conditions and suggests that this conformation exhibits behavior, which is typical for squeezed macromolecular coils. Furthermore, the pre-exponential term K_h of 0.6 observed for the pre-MGs is significantly larger than K_h values retrieved for the unfolded species (typically, in a range of 0.2–0.3), suggesting the existence of multiple bodies interactions inside the polypeptide chain.[17–19,61,62] Therefore, any small variations in the protein environment, that is, changes in the thermodynamic quality of the solvent, or changes induced by the proton transfer, interactions with a ligand, fluctuations of temperature, etc., can trigger the transition of the compact protein molecule to the more rigid MG or native states.[72]

Conformational Ensembles of MG States

The MG state of a globular protein is typically described as a conformational ensemble of compact denatured molecules that have no (or has only a trace of) rigid tertiary structure but possesses well-developed secondary structure. Small-angle X-ray scattering analysis shows that the MG has a globular structure typical of folded globular proteins.[58,134–137] 2D-NMR, coupled with hydrogen-deuterium exchange, shows that the MG is characterized not only by the native-like secondary structure content, but also by the native-like folding pattern.[138–145] A considerable increase in the accessibility of a protein molecule to proteases is noted as a specific property of the MG.[146,147] The transformation into this intermediate state is accompanied by a considerable increase in the affinity of a protein molecule to ANS and this behavior is a characteristic property of the MGs.[148,149] Finally, on the average, the hydrodynamic radius of the MG is increased by no >15% compared with that of the folded state, which corresponds to the volume increase of ~50%.[150]

The theory of the "coil-globule" transition predicts that the overall dimension of a polymer globule, R_S, changes with the chain length, N, as $R \sim (C/B)^{1/3}N^{1/3}$. Here, B and C are the second and the third virial coefficients, which characterize the pair collisions and tree-body interactions of the monomer units of the polymer chain.[72] The density of the globules is expected to show no changes with the increasing chain length, owing to $\rho = N/R^3 = (-B/C)$. These results are in excellent agreement with the data obtained for the MG conformational ensembles of proteins, for which the parameter K_h has a value of 0.9 (which reflects the larger probability of three-body interactions within the members of this conformational ensemble defined by the compact but flexible nature of the MGs) and ε equals to 0.33.[17–19,61,62]

Nature of Structural Transitions Between Different Conformational Ensembles

Conformational ensembles of partially folded intermediates of globular proteins are highly dynamic, suggesting that individual molecules within these ensembles are characterized by low conformational stability. This is reflected in low steepness of the transition curves describing their unfolding induced by strong denaturants (MG unfolding) or even in the complete lack of the sigmodal shape of the unfolding curves (pre-MG unfolding). Such behavior is in a strict contrast to the solvent-induced unfolding of ordered globular proteins, which is known to be a highly cooperative process and for many small globular proteins represents an all-or-none transition where a cooperative unit includes the whole molecule, that is, no intermediate states can be observed within the transition region. Often, urea- or GdmHCl-induced unfolding of globular proteins involves at least two cooperative steps: the ordered state to MG (N ↔ MG) and the MG to unfolded state (MG ↔ U) transitions.[51,52,151–154] Therefore, the steepness of urea- or GdmHCl-induced unfolding curves depends strongly on whether a given protein has a rigid tertiary structure (i.e., it is ordered) or is already denatured and exists as a MG conformational ensemble.[155,156]

The slope of the transition curve at its middle point is proportional to the change of the thermodynamic quantity conjugated with the variable provoking the transition, that is, to the

difference in the numbers of denaturant molecules "bound" to the initial and final states in the urea-induced or GdmHCl-induced transitions, Δv_{eff}. The slope of a phase transition in small systems depends on the system's dimensions[157,158]: in the case of first-order phase transition, the slope increases proportionally to the number of units in a system,[157] whereas the slope of the second-order phase transition is proportional to the square root of this number.[158] Therefore, it is possible to distinguish between phase and nonphase intramolecular transitions by measuring whether their slopes depend on molecular weight. Based on these premises, the dependence of slopes of solvent-induced N ↔ U, N ↔ MG, and MG ↔ U transitions in globular proteins (measured in terms of the corresponding Δv_{eff} values) on protein molecular mass (M) was analyzed.[155,156] For small proteins, cooperativity of unfolding N ↔ U transition increased with M, suggesting that their denaturant-induced unfolding exhibited the characteristics of an all-or-none transition, that is, an intramolecular analogue of first-order phase transition in macroscopic systems.[155,156] Similar behavior was also observed for the denaturant-induced N ↔ MG and MG ↔ U transitions suggesting that these two denaturant-induced transitions in small globular proteins can also be described in terms of the all-or-none transitions.[155,156]

Finally, the pre-MG and the MG were shown to be separated by an all-or-none phase transition, reflecting the fact that these partially folded intermediates represent discrete phase states.[156,159] Importantly, several structural elements of pre-MGs may occupy native-like positions.[62] The existence of such a state substantially reduces any search through the conformational space, ensuring rapid folding. Given that this state might comprise a specific native-like core with burial of hydrophobic residues, the transition from pre-MG to the MG state or to the ordered state would not require significant energy changes and could occur quite easily. An oversimplified representation of folding energy profile for the framework model with the corresponding energy barriers separating various conformational ensembles populated by a protein molecule during its folding is shown in Figure 3.

CONFORMATIONAL ENSEMBLES AND PROTEIN NONFOLDING

Structural Heterogeneity of IDPs/IDPRs

It is recognized now that a considerable number of biologically active proteins are not completely rigid, but possess some amount of disorder under the physiological conditions.[15,19,20,23,24,27] These IDPs or hybrid proteins with ordered domains and IDPRs cannot be adequately described without being considered as conformational ensembles. Contrarily to conformational ensembles transiently populated during protein folding, conformational ensembles of IDPs/IDPRs describe native functional states of these proteins. Structurally, IDPs are highly diverse and some compact IDPs contain noticeable secondary structure and behave as native MGs, whereas other IDPs are extended and possess little residual structure (i.e., these IDPs behave as native coils or native pre-MGs).[11,17,19,20] However, it was emphasized recently that intrinsic disorder can have multiple faces, can affect different levels of protein structural organization, and whole proteins, or various protein regions can be disordered to a different degree.[160] Therefore, instead of being grouped into a few discrete classes (e.g, native MGs, native pre-MGs, and native coil) structures of IDPs might be described by a complex structural spectrum with a great variety of potential structural classes and subclasses, or even can be visualized as a continuous spectrum of differently disordered conformations extending from fully ordered to completely structure-less proteins, with everything in between them.[160] Furthermore, even a single polypeptide chain can encode for a highly heterogeneous protein molecule that contains variously ordered regions, that is, possess diverse sets of foldons, inducible foldons, semifoldons, nonfoldons, and unfoldons.[161] In this view, foldon represents an independent cooperative foldable unit that can fold independently from the rest of the protein.[162] Foldon concept is derived from the analysis of ordered proteins, folding of which can be described as the stepwise assembly of the foldon units, with previously formed foldons guiding and stabilizing subsequent foldons to progressively build the native protein.[163–166] Since some regions of an IDP are spontaneously folded, other can fold (at least in part) at interaction with binding partners, still other are always in semifolded state, whereas some regions do not fold at all, an IDP can be described as a modular assembly of foldons, inducible foldons, semifoldons, and nonfoldons.[160] Furthermore, some IDPs contain unfoldons, that is, parts of protein structure that has to undergo order-to-disorder transition in order to make protein active.[160]

Amino Acid Code for Intrinsic Disorder

The absence of unique structures in IDPs/IDPRs together with all their functional and structural peculiarities is encoded in their amino acid sequences. In fact, there are significant differences between the ordered proteins/domains and IDPs/IDPRs at the level of their amino acid sequences.[11,24,167] Some of the highly disordered proteins were shown to have low sequence complexity, assuming that the sequences of IDPs may be essentially degenerated. However, it was later established that the distributions of the complexity values for ordered and disordered sequences overlapped, suggesting that low sequence

complexity did not represent the only characteristic feature of IDPs.[168] Overall, the sequences of the IDPs are characterized by noticeable amino acid compositional biases.[167,169] For example, extended IDPs were shown to be specifically localized within a unique region of the charge-hydrophobic phase space, being highly charged and possessing low hydropathy.[24] Furthermore, in comparison with ordered proteins, IDPs/IDPRs are characterized by noticeable biases in their amino acid compositions, containing less of so-called "order-promoting" residues (cysteine, tryptophan, isoleucine, tyrosine, phenylalanine, leucine, histidine, valine, asparagines, and methionine, which are mostly hydrophobic residues which are commonly found within the hydrophobic cores of foldable proteins) and more of "disorder-promoting" residues (lysine, glutamine, serine, glutamic acid, and proline, which are mostly polar and charged residues, which are typically located at the surface of the foldable proteins).[11,23,24,167,170,171]

Natural Abundance of Intrinsic Disorder

Support for the biological significance of protein intrinsic disorder phenomenon is given by the extremely wide distribution of these proteins among all kingdoms of life.[11,24,172–176] For example, an analysis of completed proteomes of 3,484 species from three main kingdoms of life (archaea, bacteria, and eukaryotes) and viruses revealed that the evolution process is characterized by the unique patterns of changes in the protein intrinsic disorder content.[176] For example, viruses are characterized by the widest spread of the disorder content in their proteomes, with the number of disordered residues ranging from 7.3% in human coronavirus NL63 to 77.3% in *Avian carcinoma virus*.[176] For several organisms from all kingdoms of life, a clear correlation was seen between their disorder contents and habitats. In multicellular eukaryotes, there was a weak correlation between the organism complexity (evaluated as a number of different cell types) and the overall disorder content. Although for both the prokaryotes and eukaryotes, the disorder content was generally independent of the proteome size, it showed sharp increase associated with the transition from the prokaryotic to eukaryotic cells.[176] This suggested that the increased disorder content in eukaryotic proteomes might be used by nature to deal with the increased cell complexity due to the appearance of the various cellular compartments.[176]

Polymer Physics of Extended IDPs

Application of the polymer physics formalism to the two classes of extended IDPs (native coils and native pre-MGs) revealed that "salted water" of typical "physiological" buffer that contains 100–150 m*M* NaCl does not represent for them a poor solvent, since these proteins are essentially noncompact under these conditions and do not possess globular structure. In other words, these solvent conditions do not force polymer segments to interact specifically with each other and, thus, do not force them to be effectively excluded from the solvent. The hydrodynamic analysis of extended IDPs revealed that the molecular dimensions of extended IDPs follow the chain length as $R_S = 0.28M^{0.49}$ or $R_S = 0.6M^{0.40}$ for the native coils and native pre-MGs, respectively. This suggests that native coils belong to the class of relatively extended unfolded conformations. Importantly, these coils show the largest K_h and the smallest ε values between different unfolded conformations of a polypeptide chain, suggesting that native coils under the physiological conditions are in considerably worsened solvent conditions in comparison with the globular proteins in the urea or the GdmHCl solutions (lowest ε value), which gives rise to the increased probability of multiple body interactions (highest K_h value). However, the molecular dimensions of native pre-MG IDPs follow the exactly same chain length dependence as conformational ensembles of pre-MGs detected as folding intermediates of ordered globular proteins. Thus, these proteins may exhibit structural features of a squeezed polymer coil.

Functions of Intrinsically Disordered Conformational Ensembles and Function-Related Structural Transitions

Highly dynamic conformational ensembles of IDPs and IDPRs are involved in countless biological activities, since the lack of rigid globular structure under physiological conditions represents a considerable functional advantage for IDPs/IDPRs.[8,10,11,13–15,17–20,23,25,26,177–183] Numerous vital cellular processes, such as the regulation of transcription and translation, and the control of cell cycle are dependent on the IDPs/IDPRs during (reviewed in Refs. 8,11,13–15,17–20,23,25). The common theme of protein disorder-based functionality is recognition, and IDPs/IDPRs are frequently involved in complex protein-protein, protein-nucleic acid, and protein-small molecule interactions. Some of these interactions can induce a disorder-to-order transition in the entire IDP or in its part.[11,13,14,18,20–22,24–26,172,184–187] In other words, some IDPs/IDPRs undergo binding-promoted functional folding at least in some of their parts. Furthermore, intrinsic disorder opens a unique capability for one protein to be involved in interaction with several unrelated binding partners and to gain differently folded bound structures.[183,188] Some IDPs/IDPRs can form highly stable complexes, whereas others are involved in signaling interactions where they undergo constant "bound-unbound" transitions, thus acting as dynamic and sensitive "on-off" switches.[21] Several IDPs/IDPRs were

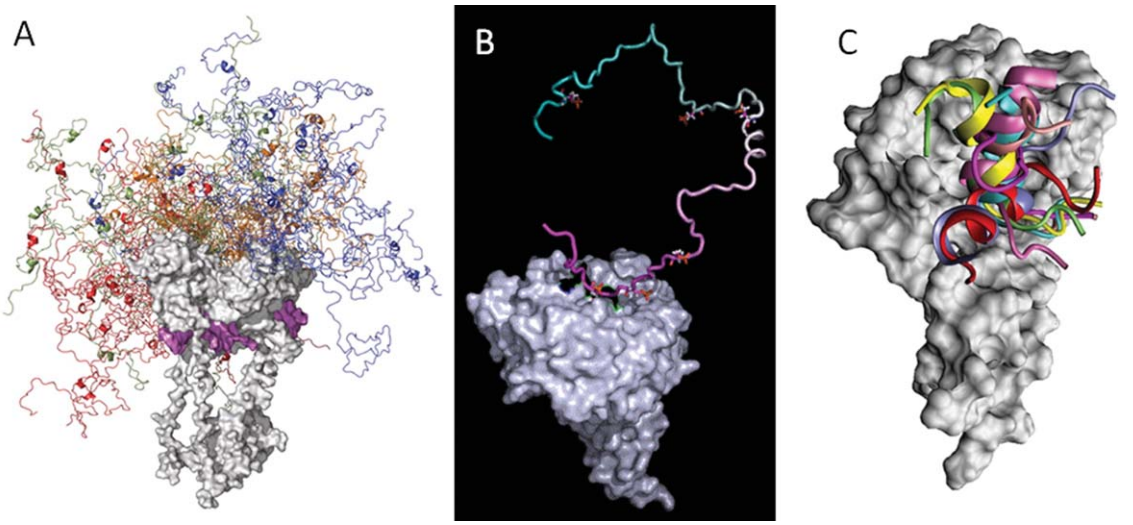

FIGURE 4 Fuzziness of protein structures and complexes. **A**: Fuzzy structure of a hybrid protein (p53 tetramer) that contains structured DNA-binding and tetramerization domains (gray space-filling models) and a disordered transactivator domain (shown as an ensemble of 20 conformations in different colors for each molecule in the tetramer). Figure is modified from Ref. 253 with permission. **B**: The NMR structure of a fuzzy complex between the cyclin-dependent kinase inhibitor Sic1 [depicted as a ribbon with color-coding from cyan (N-terminus) to magenta (C-terminus)] and the ubiquitin ligase Cdc4 (depicted as space-filling gray model). At any given moment, only one out of the nine phosphorylated sites of Sic1 interacts with a single binding site in Cdc4, generating a highly dynamic conformational ensemble of a complex described within the frames of the "polyelectrostatic" model.[254,255] **C**. Fuzzy complex of the negative regulatory domain (NRD) of p53 with dimeric S100B($\beta\beta$). According to the extensive all-atom explicit solvent simulations, NRD of p53 remains highly dynamic in the S100B($\beta\beta$)-bound state.[256]

shown to fold into different conformations depending on the peculiarities of their environments or upon interaction with different binding parters.[172,188] Although partial folding during the IDP/IDPR-based interactions is a widespread phenomenon,[184,185] there are still many other IDPs/IDPRs that are involved in the formation of the "fuzzy complexes," where they keep a certain amount of disorder in their bound forms (Figure 4).[16,21,189,190]

The range of conformational changes induced in the IDPs/IDPRs by their interaction with natural partners is very wide.[17,21] In fact, the examples of all possible conformational transitions have been described including function-induced transitions of coil to pre-MG, coil to MG, coil to ordered conformation, pre-MG to MG, pre-MG to rigid structure and MG to ordered, and rigid form.[17,18] Therefore, native proteins (or their functional regions) can exist in any of the known conformational states, ordered, MG, pre-MG, and coil.[11,23,25] Function can arise from any of these conformations and transitions between them. In other words, not just the ordered state but any of the known polypeptide conformations can be the native state of a protein.

In addition to the functional transitions toward more structured conformational ensembles, some ordered proteins possess functional dormant disorder, where these proteins are inactive when they are ordered, and become activated when they become more disordered.[161] The important features of these functional alterations are their induced nature and transient character. In other words, the function-related disordering of a protein is induced by transient alterations in its environment or by transient modification of its structure and are released as soon as the environment is restored or the modification is removed. These unusual features are important prerequisites of the protein functions relying on the induced unfolding or transient disorder mechanism.[161] In other words, functions of these proteins depends on transitions against the major stream, that is, from ordered states to dynamic conformational ensembles. Importantly, this awakening of dormant disorder phenomenon is rather abundant and different means are used by Nature to ensure such functional order-to-disorder transitions.[161] In fact, any external factor that can potentially unfold a structure of a folded protein can be used here, such as changes in pH, temperature, redox potential, light, mechanical force, membrane, interaction with ligands, protein-protein interaction, various posttranslational modifications (PTMs), release of autoinhibition due to the

unfolding of autoinhibitory domains or their interaction with nucleic acids, proteins, membranes, PTMs, etc.[161]

IDPs/IDPRs in Human Diseases

Intrinsic disorder is a tightly controlled phenomenon and there is an evolutionarily conserved tight regulation of synthesis and clearance of most IDPs,[191] giving rise to the "controlled chaos" concept.[192] This tight control is directly related to the major roles of IDPs in signaling, where, for a given signaling protein, it is crucial to be available in appropriate amounts and not to be present longer than needed.[191] However, uncontrolled chaos is frequently associated with human maladies, and as a result, intrinsic disorder is highly abundant among proteins associated with various human diseases. Since ID proteins are very common in various diseases, the "disorder in disorders" or D^2 concept was introduced to summarize work in this area[193] and concepts of the disease-related unfoldome and unfoldomics were developed.[194]

CONFORMATIONAL ENSEMBLES AND PROTEIN MISFOLDING

Molecular Mechanisms of Protein Misfolding and Protein Deposition Diseases

The sequences of proteins have evolved in such a way that their native states can be formed very efficiently even in the complex environment inside a living cell. However, under some conditions, many proteins fail to fold properly, or to remain correctly folded, giving raise to the protein this misfolding phenomenon that can eventually lead to the development of different pathological conditions, such as Alzheimer's disease, Parkinson's disease, transmissible spongiform encephalopathies, cancer, cardiovascular disease (CVD), diabetes, etc. Among the well-known structural consequence of protein misfolding, is protein aggregation leading to the development of various protein deposition diseases (frequently termed amyloidoses). Here, a specific protein or protein fragment changes from its natural soluble form into insoluble fibrils, which accumulate in a variety of organs and tissues.[34,195–199] Importantly, prior to fibrillation, amyloidogenic polypeptides may be rich in β-sheets, α-helices, or contain both α-helices and β-sheets. They may be well folded proteins or be IDPs or hybrid proteins containing differently ordered domains and differently disordered IDPRs. Despite these differences, the fibrils from different pathologies display many common properties including a core cross-β-sheet structure with continuous β-sheets formed where β-strands are running perpendicular to the long axis of the fibrils.[200]

Since all amyloid-like fibrils independent of the original structure of the given amyloidogenic proteins have a common cross-β-structure, considerable conformational rearrangements have to occur prior to fibrillation.[201] Based on the detailed analysis of structural changes preceding and accompanying amyloidogenesis, and on the structural characterization of the amyloidogenic intermediate(s) it has been concluded the amyloidogenic conformation is only slightly folded and shares many structural properties with the conformational ensembles typical for the pre-MG proteins.[201] Therefore, the general hypothesis of the molecular mechanisms of fibrillogenesis postulates that structural transformation of a polypeptide chain into the conformational ensemble of partially folded molecules represents an important prerequisite for the successful protein fibrillation.[201] However, pathways to these amyloidogenic conformational ensembles are quite different for ordered proteins and IDPs.

Even the most tightly folded protein is never completely devoid of flexibility, and due to the conformational breathing (spontaneous structural fluctuations) the structure of a globular protein under physiological conditions typically represents a mixture of tightly folded and multiple partially unfolded conformations, with the great prevalence of the former. Therefore, in ordered, well-folded proteins, amyloidogeneity-promoting changes cannot happen spontaneously due to the strong prevalence of a stable and unique tertiary structure. Thus, destabilization of an ordered protein favoring partial unfolding and formation of conformational ensembles of partially unfolded molecules is required. Therefore, the first critical step in the fibrillogenesis of an ordered protein is its partial unfolding or destabilization leading to the formation of an amyloidogenic conformational ensemble. Presumably, such a partially unfolded conformational ensemble favors reciprocal and specific intermolecular interactions, including electrostatic attraction, hydrogen bonding, and hydrophobic contacts, which are necessary for oligomerization and fibrillation.[6,7,34,195–199,202–204] In line with this hypothesis, most mutations associated with accelerated fibrillation and protein deposition diseases were shown to destabilize the native structure, increasing the steady-state concentration of partially folded conformers.[24,195–199,205–211] However, the aggregation propensity of a protein can be significantly reduced by the stabilization of the ordered structure, for example, via specific binding of ligands.[212–214]

Contrarily to ordered proteins, IDPs are assumed well suited for amyloidogenesis, since they lack significant secondary and tertiary structure, as well as many specific intra-chain interactions. In the absence of such conformational constraints, they are expected to be substantially more conformationally flexible, and thus able to polymerize more readily than tightly packed globular proteins. Substantial evidence suggests that in fibrillation of extended IDPs, which constitute a significant fraction of known amyloidogenic proteins,[215,216] and that

do not have unique tertiary structures in their native states, one of the first steps is partial folding, that is, stabilization of conformational ensembles containing partially folded protein molecules.[217–221]

α-Synuclein as a Model Amyloidogenic IDP

In addition to point mutations, various environmental factors can promote formation of such an amyloidogenic conformational ensemble. An illustrative example of the extreme sensitivity of IDPs to their environment and ability to form amyloidogenic partially folded form is given by α-synuclein, which is a small (14 kDa), soluble, intracellular, highly conserved protein that is abundant in various regions of the brain and account for as much as 1% of the total protein in soluble cytosolic brain fractions. Structurally, purified α-synuclein is a typical extended IDP, which is, being highly unstructured under conditions of neutral pH and physiological temperature, does not represent a random coil[217] but possesses some residual secondary structure,[222] that leads to partial compaction of this protein.[217,223]

Misfolding, dysfunction, aggregation, and deposition of aggregated α-synuclein are associated with several neurodegenerative diseases collectively known as synucleinopathies, with Parkinson's disease being the most well-known example of this group of neurodegenerative disorders.[224–235] The fibrillogenesis of this protein is intensively studied, and accumulated data strongly suggest that the formation of a partially folded intermediate (possessing the major characteristics of the pre-MG) represents the critical first step of α-synuclein fibrillogenesis.[217] This conformational ensemble can be stabilized by numerous factors, such as high temperatures, low pH,[217] the presence of low concentrations of various organic solvents[236] and TMAO,[217] the presence of different metal ions,[237] various salts,[238] several common pesticides/herbicides,[239–241] heparin and other glycosoaminoglycans,[242] some polycations,[243] or as a result of a spontaneous oligomerization both in vitro and in vivo.[244] In all conditions stabilizing the pre-MG-like conformation, α-synuclein was shown to possess enhanced fibrillation propensity. Importantly, fibril formation was considerably slowed down or even completely inhibited under conditions favoring formation of more folded conformations, or by stabilization of the more unfolded form, for example, by oxidation of its methionines.[245]

Multiple Pathways of Protein Misfolding and Aggregation

Obviously, the process of amyloid fibril formation does not represent the only misfolding route. In fact, contrarily to the process of the productive protein folding resulting in the formation of a unique conformation with the specific function, the end products of misfolding may have very different appearances. The morphology of these end products depends on the particular experimental conditions, and misfolded product may appear as soluble oligomers, amorphous aggregates, or amyloid-like fibrils. Any of these three species could be cytotoxic, thus giving rise to the development of pathological conditions. The reason for such a morphological difference is potentially connected to the diversity of the conformational ensembles of partially folded forms favoring protein self-association. In fact, multiple environmental factors, such as point mutations, the decrease in pH, the increase in temperature, the presence of small organic molecules or metal ions, and other charged molecules, might induce structural rearrangements within a protein molecule, shifting equilibrium toward the partially folded conformation(s). As different factors may stabilize slightly different conformational ensembles, the formation of morphologically different aggregates is expected. This idea is illustrated by Figure 5, which represents an idealized model of amyloid fibril formation and clearly shows that fibrillation is a directed process with a series of consecutive steps, including the formation of several different oligomers.[246] In this model, various oligomers are comprised of structurally identical monomers and the formation of these oligomers constitutes productive steps of the fibrillation pathway. However, aggregation is known to induce dramatic structural changes in the aggregating protein. Therefore, monomers at different aggregation stages are not identical. In addition, recent studies clearly showed that a given protein could self-assemble into various aggregated forms, depending on the peculiarities of its environment. In fact, the typical aggregation process only rarely results in the appearance of a homogeneous product where at the end of reaction only one aggregates species (amyloid fibrils, amorphous aggregates, or soluble oligomers) is present. More often, heterogeneous mixtures of various aggregated forms are observed.[246] Furthermore, each aggregated form can have multiple morphologies and monomers comprising morphologically different aggregated forms can be structurally different. All this suggests that aggregation is not a simple reaction, but a very complex process with multiple related and unrelated pathways, which can be connected or disjoined. However, regardless of the model or pathway considered, the appearance of a large aggregate inevitably involves the formation of some small oligomeric species.[246]

Polymeric Aspects of Protein Misfolding and Aggregation

Behavior of a given polymer in a given solution is determined by the peculiarities of polymer segments–solvent

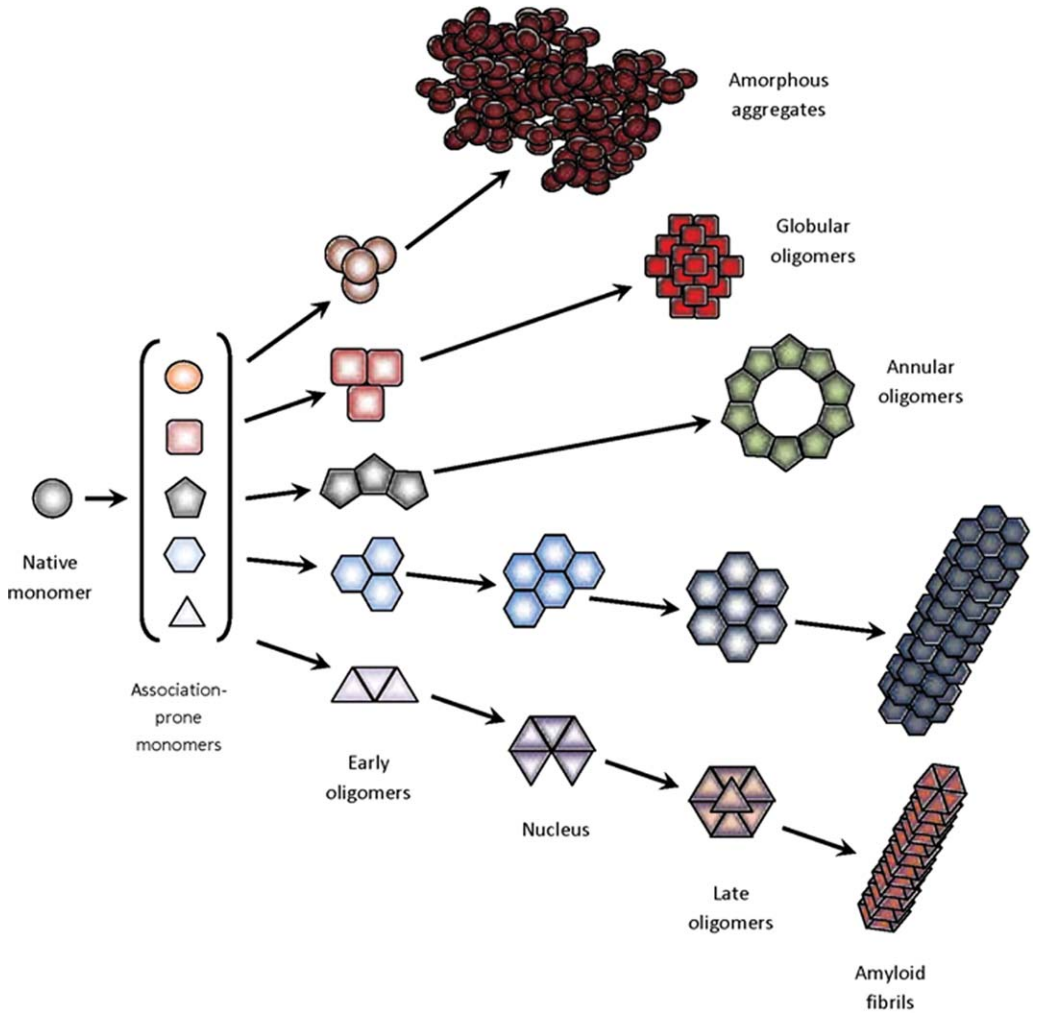

FIGURE 5 An oversimplified schematic representation of protein self-association process. Formation of multiple association-prone monomeric forms generates multiple aggregation pathways. There are three major products of the aggregation reaction–amorphous aggregates (bottom pathway), morphologically different soluble oligomers (second and third from the top pathways), and morphologically different amyloid fibrils (two bottom pathways). Two types of soluble oligomers (spheroidal and annular) and two morphologically different amyloid fibrils are shown. Changes in color reflect potential structural changes within a monomer taking place at each elementary step. In reality, the picture is much more complex and much more species can be observed. Interconversions between various species at different pathways are also possible. Figure is adopted, with permission, from Ref. 246.

interactions. For example, the major reason for the appearance of globular conformation (in our particular case, we are talking about the correctly folded form of a "normal" globular protein) in a poor solvent (water) is that this conformation effectively excludes a portion of segments from the unfavorable contacts with the solvent and forms the shielding interface between the polymer interior and solvent. In turn, the stability of globular conformation also depends on the peculiarities of interactions between protein globule and solvent. Obviously, many factors may affect the efficiency of coil-globule transition (i.e., the efficiency and direction of the process of protein folding), as well as change the efficiency of the shield (interface between the polymer and solvent) and, thus, may modulate stability of a native protein molecule. Basically, point amino acid substitutions, changes in pH, temperature, and numerous other environmental circumstances, may considerably affect the mode of polymer-solvent interactions. Thus, protein misfolding (aggregation) may originate from the changes in relative quality of solvent, which appear either due to the specific

changes in protein amino acid composition or because of the solvent composition modifications.

Overall Abundance of IDPs and Hybrid Proteins with Long IDPRs in Human Diseases

The intensive involvement of IDPs in pathogenesis of many human diseases is determined by the crucial place of these proteins in the regulation and control of various biological processes. Besides protein deposition diseases, IDPs/IDPRs are known to be responsible for pathogenesis of various cancers, diabetes, CVD, and several other maladies. The validity of this statement is based not only on a multitude of individual examples of IDPs playing various pathological roles, but also on the results of focused computational/bioinformatics studies specifically designed to estimate the abundance of IDPs in various pathological conditions.

The first approach is based on the assembly of specific datasets of proteins associated with a given disease and the computational analysis of these datasets using a number of disorder predictors.[177,215,216,247–249] This approach represents an extension of the analysis of individual proteins to a set of independent proteins. Such analysis revealed that that 79% of cancer-associated and 66% of cell-signaling proteins contain predicted regions of disorder of 30 residues or longer.[177] Similar analysis revealed that the percentage of proteins with 30 or more consecutive disordered residues was 61% for proteins associated with CVD.[248] Many CVD-related proteins were predicted to be wholly disordered, with 101 proteins from the CVD dataset predicted to have a total of almost 200 specific disorder-based binding motifs (thus about 2 binding sites per protein).[248] Finally, the dataset analysis revealed that in addition to being abundant in cancer- and CVD-related proteins, intrinsic disorder is commonly found in such maladies as neurodegenerative diseases and diabetes.[193,215]

A second approach used diseasome, a network of genetic diseases where the related proteins are interlinked within one disease and between different diseases.[250] Here, the abundance of intrinsic disorder was analyzed in the human diseasome,[250] which is a complex network that systematically links the human disease phenome with the human disease genome.[251]

These analyses showed that many human genetic diseases are caused by alteration of IDPs, that different disease classes varied in the disorder contents of their associated proteins, and that many IDPs involved in some diseases were enriched on disorder-based protein interaction sites.[250]

Finally, a third approach is based on the evaluation of the association between a particular protein function (including the disease-specific functional keywords) with the level of intrinsic disorder in a set of proteins known to carry out this function.[179,180,252] This analysis revealed that many diseases were strongly correlated with proteins predicted to be disordered.[179,180,252] Contrary to this, no disease-associated proteins were found to be strongly correlated with absence of disorder.[252]

CONCLUDING REMARKS

This review emphasizes the unique roles that conformational ensembles play in protein's life. These ensembles, which are either transiently populated (as in protein folding) or represent stable entities (as in IDPs), define peculiarities of protein folding, represent functional states of IDPs/IDPRs, and mark pathogenic traps originating from protein misfolding and leading to the pathogenesis of the realm of human diseases. Predisposition of a given protein for folding, nonfolding, and misfolding is determined by the peculiarities of its amino acid sequence and by the specific features of protein's environment. Furthermore, although the choice between nonfolding, folding, and misfolding is encoded in a given amino acid sequence, transitions between various types of conformational ensembles are also possible and are controlled by multiple factors, starting from the peculiarities of protein amino acid sequence and ending with specific features of protein environment. For example, IDPs may be forced to fold or misfold via the posttranslational modifications, addition of natural binding partners, or modification of their environment (e.g., changes in properties of solvent, etc.). A destabilizing environment may push an ordered protein to the misfolding route or can awake its dormant disorder for function, whereas the presence of chaperones may reverse the misfolding route and effectively dissolve small aggregates.[29]

REFERENCES

1. Anfinsen, C. B. Science 1973, 181, 223–230.
2. Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H., Jr. Proc Natl Acad Sci U S A 1961, 47, 1309–1314.
3. Fischer, E. Ber Dt Chem Ges 1894, 27, 2985–2993.
4. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. Nucleic Acids Res 2000, 28, 235–242.
5. Dobson, C. M. Trends Biochem Sci 1999, 24, 329–332.
6. Dobson, C. M. Biochem Soc Symp 2001, 1–26.
7. Dobson, C. M. Philos Trans R Soc Lond B Biol Sci 2001, 356, 133–145.
8. Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. Biochemistry 2002, 41, 6573–6582.
9. Dunker, A. K.; Brown, C. J.; Obradovic, Z. Adv Protein Chem 2002, 62, 25–49.
10. Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. FEBS J 2005, 272, 5129–5148.
11. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.;

Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. J Mol Graph Model 2001, 19, 26–59.
12. Dyson, H. J.; Wright, P. E. Curr Opin Struct Biol 2002, 12, 54–60.
13. Dyson, H. J.; Wright, P. E. Nat Rev Mol Cell Biol 2005, 6, 197–208.
14. Tompa, P. Trends Biochem Sci 2002, 27, 527–533.
15. Tompa, P. Curr Opin Struct Biol 2011, 21, 419–425.
16. Tompa, P.; Fuxreiter, M. Trends Biochem Sci 2008, 33, 2–8.
17. Uversky, V. N. Protein Sci 2002, 11, 739–756.
18. Uversky, V. N. Eur J Biochem 2002, 269, 2–12.
19. Uversky, V. N. Cell Mol Life Sci 2003, 60, 1852–1871.
20. Uversky, V. N. Int J Biochem Cell Biol 2011, 43, 1090–1103.
21. Uversky, V. N. Chem Soc Rev 2011, 40, 1623–1634.
22. Uversky, V. N. Curr Pharm Des 2013, 42, 4191–4213.
23. Uversky, V. N.; Dunker, A. K. Biochim Biophys Acta 2010, 1804, 1231–1264.
24. Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Proteins 2000, 41, 415–427.
25. Wright, P. E.; Dyson, H. J. J Mol Biol 1999, 293, 321–331.
26. Wright, P. E.; Dyson, H. J. Curr Opin Struct Biol 2009, 19, 31–38.
27. Turoverov, K. K.; Kuznetsova, I. M.; Uversky, V. N. Prog Biophys Mol Biol 2010, 102, 73–84.
28. Dyson, H. J. Q Rev Biophys 2011, 44, 467–518.
29. Ben-Zvi, A. P.; Goloubinoff, P. J Struct Biol 2001, 135, 84–93.
30. Finkelstein, A. V.; Ptitsyn, O. B. Protein Physics: A Course of Lectures; Academic Press: London, UK; San Diego, CA, 2002.
31. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Proteins 1995, 21, 167–195.
32. Dill, K. A.; Chan, H. S. Nat Struct Biol 1997, 4, 10–19.
33. Dobson, C. M.; Sali, A.; Karplus, M. Angew Chem-Int Ed 1998, 37, 868–893.
34. Dobson, C. M.; Karplus, M. Curr Opin Struct Biol 1999, 9, 92–101.
35. Dinner, A. R.; Sali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M. Trends Biochem Sci 2000, 25, 331–339.
36. Fersht, A. R.; Daggett, V. Cell 2002, 108, 573–582.
37. Levinthal, C. J Chem Phys 1968, 65, 44–45.
38. Ptitsyn, O. B. Dokl Akad Nauk SSSR 1973, 210, 1213–1215.
39. Ohgushi, M.; Wada, A. FEBS Lett 1983, 164, 21–24.
40. Dolgikh, D. A.; Gilmanshin, R. I.; Brazhnikov, E. V.; Bychkova, V. E.; Semisotnov, G. V.; Venyaminov, S.; Ptitsyn, O. B. FEBS Lett 1981, 136, 311–315.
41. Radford, S. E. Trends Biochem Sci 2000, 25, 611–618.
42. Jahn, T. R.; Radford, S. E. FEBS J 2005, 272, 5962–5970.
43. Kim, P. S.; Baldwin, R. L. Annu Rev Biochem 1982, 51, 459–489.
44. Ikeguchi, M.; Kuwajima, K.; Mitani, M.; Sugai, S. Biochemistry 1986, 25, 6965–6972.
45. Hughson, F. M.; Wright, P. E.; Baldwin, R. L. Science 1990, 249, 1544–1548.
46. Ptitsyn, O. B.; Pain, R. H.; Semisotnov, G. V.; Zerovnik, E.; Razgulyaev, O. I. FEBS Lett 1990, 262, 20–24.
47. Goldberg, M. E.; Semisotnov, G. V.; Friguet, B.; Kuwajima, K.; Ptitsyn, O. B.; Sugai, S. FEBS Lett 1990, 263, 51–56.
48. Jennings, P. A.; Wright, P. E. Science 1993, 262, 892–896.
49. Loh, S. N.; Kay, M. S.; Baldwin, R. L. Proc Natl Acad Sci U S A 1995, 92, 5446–5450.
50. Balbach, J.; Forge, V.; van Nuland, N. A.; Winder, S. L.; Hore, P. J.; Dobson, C. M. Nat Struct Biol 1995, 2, 865–870.
51. Ptitsyn, O. B. Curr Opin Struct Biol 1995, 5, 74–78.
52. Ptitsyn, O. B. Adv Protein Chem 1995, 47, 83–229.
53. Ptitsyn, O. B.; Bychkova, V. E.; Uversky, V. N. Philos Trans R Soc Lond B Biol Sci 1995, 348, 35–41.
54. Fink, A. L. Subcell Biochem 1995, 24, 27–53.
55. Fink, A. L. Annu Rev Biophys Biomol Struct 1995, 24, 495–522.
56. Fink, A. L. Methods Mol Biol 1995, 40, 343–360.
57. Fink, A. L.; Oberg, K. A.; Seshadri, S. Fold Des 1998, 3, 19–25.
58. Uversky, V. N.; Karnoup, A. S.; Segel, D. J.; Seshadri, S.; Doniach, S.; Fink, A. L. J Mol Biol 1998, 278, 879–894.
59. Uversky, V. N.; Ptitsyn, O. B. Biochemistry 1994, 33, 2782–2791.
60. Uversky, V. N.; Ptitsyn, O. B. J Mol Biol 1996, 255, 215–228.
61. Tcherkasskaya, O.; Davidson, E. A.; Uversky, V. N. J Proteome Res 2003, 2, 37–42.
62. Tcherkasskaya, O.; Uversky, V. N. Proteins 2001, 44, 244–254.
63. Tcherkasskaya, O.; Uversky, V. N. Protein Pept Lett 2003, 10, 239–245.
64. Ptitsyn, O. B.; Volkenstein, M. V. J Biomol Struct Dyn 1986, 4, 137–156.
65. Lau, K. F.; Dill, K. A. Proc Natl Acad Sci U S A 1990, 87, 638–642.
66. Anson, M. L.; Mirsky, A. E. J Gen Physiol 1932, 15, 341–350.
67. Mirsky, A. E.; Pauling, L. Proc Natl Acad Sci U S A 1936, 22, 439–447.
68. Neurath, H.; Greenstein, J. P.; Putnam, F. W.; Erickson, J. O. Chem Rew 1944, 34, 157–265.
69. Tanford, C. Adv Protein Chem 1968, 23, 121–282.
70. Uversky, V. N.; Fink, A. L. Biochemistry (Mosc) 1999, 64, 552–555.
71. Baldwin, R. L. Trends Biochem Sci 1986, 11, 6–9.
72. Grossberg, A. Y.; Khohlov, A. R. Statistical Physics of Macromolecules; Nauka: Moscow, 1989.
73. Tanford, C. Physical Chemistry of Macromolecules; Willey: New York, 1961.
74. Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W. Proc Natl Acad Sci U S A 2004, 101, 12491–12496.
75. Miller, W. G.; Goebel, C. V. Biochemistry 1968, 7, 3925–3935.
76. Wang, Z.; Plaxco, K. W.; Makarov, D. E. Biopolymers 2007, 86, 321–328.
77. Fitzkee, N. C.; Rose, G. D. Proc Natl Acad Sci U S A 2004, 101, 12497–12502.
78. James, E.; Wu, P. G.; Stites, W.; Brand, L. Biochemistry 1992, 31, 10217–10225.
79. Alexandrescu, A. T.; Abeygunawardana, C.; Shortle, D. Biochemistry 1994, 33, 1063–1072.
80. Wang, Y.; Shortle, D. Protein Sci 1996, 5, 1898–1906.
81. Wang, Y.; Shortle, D. Biochemistry 1995, 34, 15895–15905.
82. Wang, Y.; Shortle, D. Fold Des 1997, 2, 93–100.
83. Gillespie, J. R.; Shortle, D. J Mol Biol 1997, 268, 170–184.

84. Gillespie, J. R.; Shortle, D. J Mol Biol 1997, 268, 158–169.
85. Shortle, D.; Ackerman, M. S. Science 2001, 293, 487–489.
86. Saab-Rincon, G.; Froebe, C. L.; Matthews, C. R. Biochemistry 1993, 32, 13981–13990.
87. Saab-Rincon, G.; Gualfetti, P. J.; Matthews, C. R. Biochemistry 1996, 35, 1988–1994.
88. Neri, D.; Billeter, M.; Wider, G.; Wuthrich, K. Science 1992, 257, 1559–1563.
89. Neri, D.; Wider, G.; Wuthrich, K. Proc Natl Acad Sci U S A 1992, 89, 4397–4401.
90. Neri, D.; Wider, G.; Wuthrich, K. FEBS Lett 1992, 303, 129–135.
91. Wang, H. M.; Yu, C. Biophys J 2011, 100, 154–164.
92. Zhang, O.; Forman-Kay, J. D. Biochemistry 1995, 34, 6784–6794.
93. Zhang, O.; Forman-Kay, J. D. Biochemistry 1997, 36, 3959–3970.
94. Nolting, B.; Golbik, R.; Soler-Gonzalez, A. S.; Fersht, A. R. Biochemistry 1997, 36, 9899–9905.
95. Arcus, V. L.; Vuilleumier, S.; Freund, S. M.; Bycroft, M.; Fersht, A. R. J Mol Biol 1995, 254, 305–321.
96. Koepf, E. K.; Petrassi, H. M.; Sudol, M.; Kelly, J. W. Protein Sci 1999, 8, 841–853.
97. Gottfried, D. S.; Haas, E. Biochemistry 1992, 31, 12353–12362.
98. Amir, D.; Krausz, S.; Haas, E. Proteins 1992, 13, 162–173.
99. Kazmirski, S. L.; Wong, K. B.; Freund, S. M.; Tan, Y. J.; Fersht, A. R.; Daggett, V. Proc Natl Acad Sci U S A 2001, 98, 4349–4354.
100. Svensson, M.; Jonasson, P.; Freskgard, P. O.; Jonsson, B. H.; Lindgren, M.; Martensson, L. G.; Gentile, M.; Boren, K.; Carlsson, U. Biochemistry 1995, 34, 8606–8620.
101. Martensson, L. G.; Jonsson, B. H.; Freskgard, P. O.; Kihlgren, A.; Svensson, M.; Carlsson, U. Biochemistry 1993, 32, 224–231.
102. Hammarstrom, P.; Carlsson, U. Biochem Biophys Res Commun 2000, 276, 393–398.
103. Schwarzinger, S.; Wright, P. E.; Dyson, H. J. Biochemistry 2002, 41, 12681–12686.
104. Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. Science 2002, 295, 1719–1722.
105. Lee, B. C.; Kumauchi, M.; Hoff, W. D. J Biol Chem 2010, 285, 12579–12586.
106. Tafer, H.; Hiller, S.; Hilty, C.; Fernandez, C.; Wuthrich, K. Biochemistry 2004, 43, 860–869.
107. Romero-Beviar, M.; Martinez-Rodriguez, S.; Prieto, J.; Goormaghtigh, E.; Ariz, U.; Martinez-Chantar Mde, L.; Gomez, J.; Neira, J. L. Protein Eng Des Sel 2010, 23, 729–742.
108. Wirmer, J.; Berk, H.; Ugolini, R.; Redfield, C.; Schwalbe, H. Protein Sci 2006, 15, 1397–1407.
109. Ohnishi, S.; Lee, A. L.; Edgell, M. H.; Shortle, D. Biochemistry 2004, 43, 4064–4070.
110. Ropson, I. J.; Boyer, J. A.; Dalessio, P. M. Biochemistry 2006, 45, 2608–2617.
111. Santra, M. K.; Dasgupta, D.; Panda, D. Photochem Photobiol 2006, 82, 480–486.
112. Kogo, H.; Takeuchi, K.; Inoue, H.; Kihara, H.; Kojima, M.; Takahashi, K. Biochim Biophys Acta 2009, 1794, 70–74.
113. Mok, K. H.; Kuhn, L. T.; Goez, M.; Day, I. J.; Lin, J. C.; Andersen, N. H.; Hore, P. J. Nature 2007, 447, 106–109.
114. Ureta, D. B.; Craig, P. O.; Gomez, G. E.; Delfino, J. M. Biochemistry 2007, 46, 14567–14577.
115. Wallgren, M.; Aden, J.; Pylypenko, O.; Mikaelsson, T.; Johansson, L. B.; Rak, A.; Wolf-Watz, M. J Mol Biol 2008, 379, 845–858.
116. Ratcliff, K.; Marqusee, S. Biochemistry 2010, 49, 5167–5175.
117. Huang, J. R.; Grzesiek, S. J Am Chem Soc 2010, 132, 694–705.
118. Dill, K. A.; Shortle, D. Annu Rev Biochem 1991, 60, 795–825.
119. Shortle, D. R. Curr Opin Struct Biol 1996, 6, 24–30.
120. Shortle, D. FASEB J 1996, 10, 27–34.
121. Baldwin, R. L.; Zimm, B. H. Proc Natl Acad Sci U S A 2000, 97, 12391–12392.
122. Plaxco, K. W.; Gross, M. Nat Struct Biol 2001, 8, 659–660.
123. Shimizu, S.; Chan, H. S. Proteins 2002, 49, 560–566.
124. Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M. Fold Des 1996, 1, R95–106.
125. Ramachandran, G. N.; Sasisekharan, V. Adv Protein Chem 1968, 23, 283–438.
126. Shortle, D. Adv Protein Chem 2002, 62, 1–23.
127. Palleros, D. R.; Shi, L.; Reid, K. L.; Fink, A. L. Biochemistry 1993, 32, 4314–4321.
128. Zhang, Y. L.; Zhou, J. M.; Tsou, C. L. Biochim Biophys Acta 1996, 1295, 239–244.
129. Karnoup, A. S.; Uversky, V. N. Macromolecules 1997, 30, 7427–7434.
130. Bushmarina, N. A.; Kuznetsova, I. M.; Biktashev, A. G.; Turoverov, K. K.; Uversky, V. N. ChemBioChem 2001, 2, 813–821.
131. Kuznetsova, I. M.; Stepanenko, O. V.; Turoverov, K. K.; Zhu, L.; Zhou, J. M.; Fink, A. L.; Uversky, V. N. Biochim Biophys Acta 2002, 1596, 138–155.
132. Georlette, D.; Blaise, V.; Dohmen, C.; Bouillenne, F.; Damien, B.; Depiereux, E.; Gerday, C.; Uversky, V. N.; Feller, G. J Biol Chem 2003, 278, 49945–49953.
133. Georlette, D.; Damien, B.; Blaise, V.; Depiereux, E.; Uversky, V. N.; Gerday, C.; Feller, G. J Biol Chem 2003, 278, 37015–37023.
134. Eliezer, D.; Chiba, K.; Tsuruta, H.; Doniach, S.; Hodgson, K. O.; Kihara, H. Biophys J 1993, 65, 912–917.
135. Kataoka, M.; Hagihara, Y.; Mihara, K.; Goto, Y. J Mol Biol 1993, 229, 591–596.
136. Kataoka, M.; Kuwajima, K.; Tokunaga, F.; Goto, Y. Protein Sci 1997, 6, 422–430.
137. Semisotnov, G. V.; Kihara, H.; Kotova, N. V.; Kimura, K.; Amemiya, Y.; Wakabayashi, K.; Serdyuk, I. N.; Timchenko, A. A.; Chiba, K.; Nikaido, K.; Ikura, T.; Kuwajima, K. J Mol Biol 1996, 262, 559–574.
138. Baum, J.; Dobson, C. M.; Evans, P. A.; Hanley, C. Biochemistry 1989, 28, 7–13.
139. Bushnell, G. W.; Louie, G. V.; Brayer, G. D. J Mol Biol 1990, 214, 585–595.
140. Chyan, C. L.; Wormald, C.; Dobson, C. M.; Evans, P. A.; Baum, J. Biochemistry 1993, 32, 5681–5691.
141. Jeng, M. F.; Englander, S. W.; Elove, G. A.; Wand, A. J.; Roder, H. Biochemistry 1990, 29, 10433–10437.
142. Bose, H. S.; Whittal, R. M.; Baldwin, M. A.; Miller, W. L. Proc Natl Acad Sci U S A 1999, 96, 7250–7255.
143. Bracken, C. J Mol Graph Model 2001, 19, 3–12.

144. Eliezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. Nat Struct Biol 1998, 5, 148–155.
145. Wu, L. C.; Laub, P. B.; Elove, G. A.; Carey, J.; Roder, H. Biochemistry 1993, 32, 10271–10276.
146. Merrill, A. R.; Cohen, F. S.; Cramer, W. A. Biochemistry 1990, 29, 5829–5836.
147. Fontana, A.; Polverino de Laureto, P.; De Philipps, V. In Protein Stability and Stabilization; van den Tweel, W.; Harder, A.; Buitelear, M., Eds.; Elsevier: Amsterdam, 1993; p 101–110.
148. Semisotnov, G. V.; Rodionova, N. A.; Razgulyaev, O. I.; Uversky, V. N.; Gripas, A. F.; Gilmanshin, R. I. Biopolymers 1991, 31, 119–128.
149. Uversky, V. N.; Winter, S.; Lober, G. Biophys Chem 1996, 60, 79–88.
150. Uversky, V. N. Biochemistry 1993, 32, 13288–13298.
151. Ptitsyn, O. B. J Prot Chem 1987, 6, 273–293.
152. Kuwajima, K. Proteins 1989, 6, 87–103.
153. Christensen, H.; Pain, R. H. Eur Biophys J 1991, 19, 221–229.
154. Rodionova, N. A.; Semisotnov, G. V.; Kutyshenko, V. P.; Uverskii, V. N.; Bolotina, I. A. Mol Biol (Mosk) 1989, 23, 683–692.
155. Ptitsyn, O. B.; Uversky, V. N. FEBS Lett 1994, 341, 15–18.
156. Uversky, V. N.; Ptitsyn, O. B. Fold Des 1996, 1, 117–122.
157. Hill, T. L. Thermodynamics of the Small Systems; Wiley: New York, 1963-1964.
158. Grosberg, A. Y.; Khokhlov, A. R. Statistical Physics of Macromolecules; Nauka: Moscow, 1989.
159. Uversky, V. N.; Semisotnov, G. V.; Pain, R. H.; Ptitsyn, O. B. FEBS Lett 1992, 314, 89–92.
160. Uversky, V. N. Biochim Biophys Acta 2012.
161. Uversky, V. N. Biochim Biophys Acta 2013, 1834, 932–951.
162. Lindberg, M. O.; Oliveberg, M. Curr Opin Struct Biol 2007, 17, 21–29.
163. Maity, H.; Maity, M.; Krishna, M. M.; Mayne, L.; Englander, S. W. Proc Natl Acad Sci U S A 2005, 102, 4741–4746.
164. Englander, S. W.; Mayne, L.; Krishna, M. M. Q Rev Biophys 2007, 40, 287–326.
165. Bedard, S.; Krishna, M. M.; Mayne, L.; Englander, S. W. Proc Natl Acad Sci U S A 2008, 105, 7182–7187.
166. Krishna, M. M.; Englander, S. W. Protein Sci 2007, 16, 449–464.
167. Dunker, A. K.; Garner, E.; Guilliot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; Villafranca, J. E. Pac Symp Biocomput 1998, 473–484.
168. Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. Proteins 2001, 42, 38–48.
169. Garner, E.; Cannon, P.; Romero, P.; Obradovic, Z.; Dunker, A. K. Genome Inform Ser Workshop Genome Inform 1998, 9, 201–213.
170. Radivojac, P.; Iakoucheva, L. M.; Oldfield, C. J.; Obradovic, Z.; Uversky, V. N.; Dunker, A. K. Biophys J 2007, 92, 1439–1456.
171. Vacic, V.; Uversky, V. N.; Dunker, A. K.; Lonardi, S. BMC Bioinformatics 2007, 8, 211.
172. Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Romero, P.; Uversky, V. N.; Dunker, A. K. Biochemistry 2005, 44, 12454–12470.
173. Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. Genome Inform Ser Workshop Genome Inform 2000, 11, 161–171.
174. Romero, P.; Obradovic, Z.; Kissinger, C. R.; Villafranca, J. E.; Garner, E.; Guilliot, S.; Dunker, A. K. Pac Symp Biocomput 1998, 437–448.
175. Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. J Mol Biol 2004, 337, 635–645.
176. Xue, B.; Dunker, A. K.; Uversky, V. N. J Biomol Struct Dyn 2012, 30, 137–149.
177. Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. J Mol Biol 2002, 323, 573–584.
178. Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. J Mol Recognit 2005, 18, 343–384.
179. Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N.; Obradovic, Z. J Proteome Res 2007, 6, 1882–1898.
180. Vucetic, S.; Xie, H.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Obradovic, Z.; Uversky, V. N. J Proteome Res 2007, 6, 1899–1916.
181. Liu, J.; Faeder, J. R.; Camacho, C. J. Proc Natl Acad Sci U S A 2009, 106, 19819–19823.
182. Kim, P. M.; Sboner, A.; Xia, Y.; Gerstein, M. Mol Syst Biol 2008, 4, 179.
183. Oldfield, C. J.; Meng, J.; Yang, J. Y.; Yang, M. Q.; Uversky, V. N.; Dunker, A. K. BMC Genomics 2008, 9 Suppl 1, S1.
184. Mohan, A.; Oldfield, C. J.; Radivojac, P.; Vacic, V.; Cortese, M. S.; Dunker, A. K.; Uversky, V. N. J Mol Biol 2006, 362, 1043–1059.
185. Vacic, V.; Oldfield, C. J.; Mohan, A.; Radivojac, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. J Proteome Res 2007, 6, 2351–2366.
186. Meszaros, B.; Simon, I.; Dosztanyi, Z. PLoS Comput Biol 2009, 5, e1000376.
187. Dosztanyi, Z.; Meszaros, B.; Simon, I. Bioinformatics 2009, 25, 2745–2746.
188. Hsu, W. L.; Oldfield, C.; Meng, J.; Huang, F.; Xue, B.; Uversky, V. N.; Romero, P.; Dunker, A. K. Pac Symp Biocomput 2012, 116–127.
189. Meszaros, B.; Simon, I.; Dosztanyi, Z. Phys Biol 2011, 8, 035003.
190. Hazy, E.; Tompa, P. ChemPhysChem 2009, 10, 1415–1419.
191. Gsponer, J.; Futschik, M. E.; Teichmann, S. A.; Babu, M. M. Science 2008, 322, 1365–1368.
192. Uversky, V. N.; Dunker, A. K. Science 2008, 322, 1340–1341.
193. Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Annu Rev Biophys 2008, 37, 215–246.
194. Uversky, V. N.; Oldfield, C. J.; Midic, U.; Xie, H.; Xue, B.; Vucetic, S.; Iakoucheva, L. M.; Obradovic, Z.; Dunker, A. K. BMC Genomics 2009, 10 Suppl 1, S7.
195. Kelly, J. W. Curr Opin Struct Biol 1998, 8, 101–106.
196. Uversky, V. N.; Talapatra, A.; Gillespie, J. R.; Fink, A. L. Med Sci Monitor 1999, 5, 1238–1254.
197. Uversky, V. N.; Talapatra, A.; Gillespie, J. R.; Fink, A. L. Med Sci Monitor 1999, 5, 1001–1012.
198. Bellotti, V.; Mangione, P.; Stoppini, M. Cell Mol Life Sci 1999, 55, 977–991.
199. Rochet, J. C.; Lansbury, P. T., Jr. Curr Opin Struct Biol 2000, 10, 60–68.
200. Sunde, M.; Serpell, L. C.; Bartlam, M.; Fraser, P. E.; Pepys, M. B.; Blake, C. C. J Mol Biol 1997, 273, 729–739.

201. Uversky, V. N.; Fink, A. L. Biochim Biophys Acta 2004, 1698, 131–153.
202. Lansbury, P. T., Jr. Proc Natl Acad Sci U S A 1999, 96, 3342–3344.
203. Fink, A. L. Fold Des 1998, 3, R9–23.
204. Zerovnik, E. Eur J Biochem 2002, 269, 3362–3371.
205. Kelly, J. W. Curr Opin Struct Biol 1996, 6, 11–17.
206. Wetzel, R. Adv Protein Chem 1997, 50, 183–242.
207. Canet, D.; Sunde, M.; Last, A. M.; Miranker, A.; Spencer, A.; Robinson, C. V.; Dobson, C. M. Biochemistry 1999, 38, 6419–6427.
208. Lashuel, H. A.; Wurth, C.; Woo, L.; Kelly, J. W. Biochemistry 1999, 38, 13560–13573.
209. Saraiva, M. J. FEBS Lett 2001, 498, 201–203.
210. Heegaard, N. H.; Sen, J. W.; Kaarsholm, N. C.; Nissen, M. H. J Biol Chem 2001, 276, 32657–32662.
211. Colon, W.; Lai, Z.; McCutchen, S. L.; Miroy, G. J.; Strang, C.; Kelly, J. W. Ciba Found Symp 1996, 199, 228–238; discussion 239-242.
212. Chiti, F.; Taddei, N.; Stefani, M.; Dobson, C. M.; Ramponi, G. Protein Sci 2001, 10, 879–886.
213. Miroy, G. J.; Lai, Z.; Lashuel, H. A.; Peterson, S. A.; Strang, C.; Kelly, J. W. Proc Natl Acad Sci U S A 1996, 93, 15051–15056.
214. Nielsen, L.; Khurana, R.; Coats, A.; Frokjaer, S.; Brange, J.; Vyas, S.; Uversky, V. N.; Fink, A. L. Biochemistry 2001, 40, 6036–6046.
215. Uversky, V. N. Front Biosci 2009, 14, 5188–5238.
216. Uversky, V. N. Curr Alzheimer Res 2008, 5, 260–287.
217. Uversky, V. N.; Li, J.; Fink, A. L. FEBS Lett 2001, 509, 31–35.
218. Pavlov, N. A.; Cherny, D. I.; Heim, G.; Jovin, T. M.; Subramaniam, V. FEBS Lett 2002, 517, 37–40.
219. Teplow, D. B. Amyloid 1998, 5, 121–142.
220. Kayed, R.; Bernhagen, J.; Greenfield, N.; Sweimeh, K.; Brunner, H.; Voelter, W.; Kapurniotu, A. J Mol Biol 1999, 287, 781–796.
221. Schweers, O.; Schonbrunn-Hanebeck, E.; Marx, A.; Mandelkow, E. J Biol Chem 1994, 269, 24290–24297.
222. Eliezer, D.; Kutluay, E.; Bussell, R., Jr.; Browne, G. J Mol Biol 2001, 307, 1061–1073.
223. Uversky, V. N. J Biomol Struct Dyn 2003, 21, 211–234.
224. Norris, E. H.; Giasson, B. I.; Lee, V. M. Curr Top Dev Biol 2004, 60, 17–54.
225. Jellinger, K. A. Mov Disord 2003, 18 Suppl 6, S2–12.
226. Dev, K. K.; Hofele, K.; Barbieri, S.; Buchman, V. L.; van der Putten, H. Neuropharmacology 2003, 45, 14–44.
227. Duda, J. E.; Lee, V. M.; Trojanowski, J. Q. J Neurosci Res 2000, 61, 121–127.
228. Mitra, K.; Gangopadhaya, P. K.; Das, S. K. Neurol India 2003, 51, 183–188.
229. Marti, M. J.; Tolosa, E.; Campdelacreu, J. Mov Disord 2003, 18 Suppl 6, S21–27.
230. Trojanowski, J. Q.; Lee, V. M. Ann N Y Acad Sci 2003, 991, 107–110.
231. Dickson, D. W. Curr Opin Neurol 2001, 14, 423–432.
232. Goedert, M. Clin Chem Lab Med 2001, 39, 308–312.
233. Spillantini, M. G.; Goedert, M. Ann N Y Acad Sci 2000, 920, 16–27.
234. Galvin, J. E.; Lee, V. M.; Trojanowski, J. Q. Arch Neurol 2001, 58, 186–190.
235. Goedert, M. Philos Trans R Soc Lond B Biol Sci 1999, 354, 1101–1118.
236. Munishkina, L. A.; Phelan, C.; Uversky, V. N.; Fink, A. L. Biochemistry 2003, 42, 2720–2730.
237. Uversky, V. N.; Li, J.; Fink, A. L. J Biol Chem 2001, 276, 44284–44296.
238. Munishkina, L. A.; Henriques, J.; Uversky, V. N.; Fink, A. L. Biochemistry 2004, 43, 3289–3300.
239. Uversky, V. N.; Li, J.; Fink, A. L. FEBS Lett 2001, 500, 105–108.
240. Uversky, V. N.; Li, J.; Bower, K.; Fink, A. L. Neurotoxicology 2002, 23, 527–536.
241. Manning-Bog, A. B.; McCormack, A. L.; Li, J.; Uversky, V. N.; Fink, A. L.; Di Monte, D. A. J Biol Chem 2002, 277, 1641–1644.
242. Cohlberg, J. A.; Li, J.; Uversky, V. N.; Fink, A. L. Biochemistry 2002, 41, 1502–1511.
243. Goers, J.; Uversky, V. N.; Fink, A. L. Protein Sci 2003, 12, 702–707.
244. Uversky, V. N.; Lee, H. J.; Li, J.; Fink, A. L.; Lee, S. J. J Biol Chem 2001, 276, 43495–43498.
245. Uversky, V. N.; Yamin, G.; Souillac, P. O.; Goers, J.; Glaser, C. B.; Fink, A. L. FEBS Lett 2002, 517, 239–244.
246. Uversky, V. N. FEBS J 2010, 277, 2940–2953.
247. Uversky, V. N.; Roman, A.; Oldfield, C. J.; Dunker, A. K. J Proteome Res 2006, 5, 1829–1842.
248. Cheng, Y.; LeGall, T.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N. Biochemistry 2006, 45, 10448–10460.
249. Mohan, A.; Sullivan, W. J., Jr.; Radivojac, P.; Dunker, A. K.; Uversky, V. N. Mol Biosyst 2008, 4, 328–340.
250. Midic, U.; Oldfield, C. J.; Dunker, A. K.; Obradovic, Z.; Uversky, V. N. BMC Genomics 2009, 10 (suppl. 1), S12.
251. Goh, K. I.; Cusick, M. E.; Valle, D.; Childs, B.; Vidal, M.; Barabasi, A. L. Proc Natl Acad Sci U S A 2007, 104, 8685–8690.
252. Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Obradovic, Z.; Uversky, V. N. J Proteome Res 2007, 6, 1917–1932.
253. Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R. Proc Natl Acad Sci U S A 2008, 105, 5762–5767.
254. Mittag, T.; Orlicky, S.; Choy, W. Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L. E.; Tyers, M.; Forman-Kay, J. D. Proc Natl Acad Sci U S A 2008, 105, 17772–17777.
255. Mittag, T.; Kay, L. E.; Forman-Kay, J. D. J Mol Recognit 2010, 23, 105–116.
256. McDowell, C.; Chen, J. J Mol Biol 2013, 425, 999–1010.

Reviewing Editor: Jean S. Baum