1

FUNDAMENTALS OF PROTEIN FOLDING

VLADIMIR N. UVERSKY

1.1 FOLDING–MISFOLDING–NONFOLDING CROSSROADS

Data accumulated to date suggest that a protein can exist in at least three global forms: functional and folded, functional and intrinsically disordered (nonfolded), and nonfunctional and misfolded or aggregated. Therefore, in *in vitro* experiments or in the cellular environment, a polypeptide chain faces a choice among three potential routes: nonfolding, folding, and misfolding, with the last two representing competitive routes to higher structural order (see Figure 1.1). For a single-chain protein, folding, nonfolding, and misfolding pathways represent a choice of each individual molecule, whereas unproductive protein aggregation, which frequently follows protein misfolding and is often associated with the pathogenesis of several diseases, is a fate of the ensemble of molecules.

Multiple factors, originating from the peculiarities of the protein amino acid sequence and/or features of the protein environment, might affect the choice among nonfolding, folding, and misfolding. The primary selection between low structural complexity (nonfolding) and increased complexity (folding) is

Protein Aggregation in Bacteria: Functional and Structural Properties of Inclusion Bodies in Bacterial Cells, First Edition. Edited by Silvia Maria Doglia and Marina Lotti.

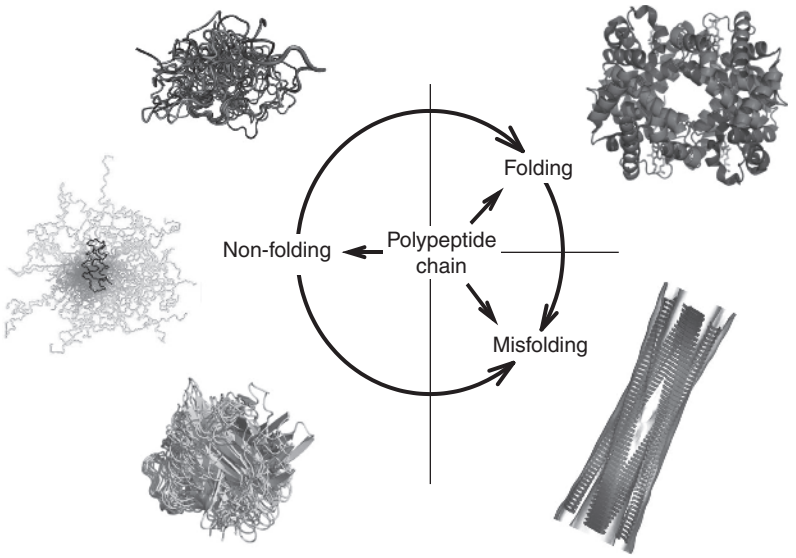

FIGURE 1.1 Fate of a newly synthesized polypeptide chain in a cell. (*See insert for color representation of the figure*.)

determined only by the amino acid composition. An abnormally highly charged polypeptide with low overall hydrophobicity will not fold, giving rise to a natively unfolded protein, whereas usually, under identical conditions a polypeptide chain will choose the folding path. On the other hand, some changes in the amino acid sequence (point mutations) may favor the misfolding pathway for both natively unfolded and natively folded proteins.

Importantly, for a given polypeptide chain a chosen fate is not a final one and a choice may be modulated further by environmental pressure (Figure 1.1) (Uversky, 2003b). For example, intrinsically disordered proteins (IDPs) may be forced to fold or misfold via the modification of their environment (i.e., addition of natural binding partners, changes in properties of solvent, etc.), whereas a destabilizing environment may push a natively folded protein to the misfolding route. Alternatively, the presence of chaperones may reverse the misfolding route and effectively dissolve small aggregates (Ben-Zvi and Goloubinoff, 2001). Another important point is that to some extent, the misfolding of natively unfolded proteins resembles the process of protein folding [i.e., it represents a pathway from a simple, flexible, and disordered conformation (e.g., an unfolded polypeptide chain), via a somehow more ordered partially folded intermediate, to a complex and rigid structure (e.g., an amyloid fibril)]. On the other hand, misfolding 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.

1.2 PROTEIN FOLDING

The ability of proteins to adopt their functional highly structured states in the intracellular environment during or after biosynthesis on the ribosome is one of the most remarkable evolutionary achievements in biology. In recent years our understanding of the mechanisms of the protein self-organization process has increased dramatically. This understanding has been achieved as a result of the application of a variety of novel experimental and theoretical approaches to this complex task (Bryngelson et al., 1995; Dill and Chan, 1997; Dobson et al., 1998; Dobson and Karplus, 1999; Dinner et al., 2000; Fersht and Daggett, 2002; Uversky, 2003b; Turoverov et al., 2010). In this chapter we cover several protein folding–related problems, such as a protein-folding code; major models of protein folding; polymer aspects underlying protein folding and structural peculiarities of the unfolded state and folding intermediates; and peculiarities of protein amino acid compositions favoring formation of equilibrium partially folded intermediate(s).

1.2.1 Protein-Folding Code

In protein biosynthesis, the information encoded in the DNA/mRNA nucleotide sequence is read step by step, and the corresponding amino acids are gathered into the polypeptide chain one after another. Therefore, the one-dimensional information encoded in the DNA nucleotide sequence is sequentially transformed into the one-dimensional information included in the protein amino acid sequence. As the interactions between remote amino acid residues play a crucial role in protein folding, this process obviously deviates from the linear information transduction. Only some amino acid residues are crucial for protein folding. Therefore, proteins with very low sequence homology can have similar structures, whereas a single amino acid replacement can significantly affect the rate of protein folding or in extreme cases can halt correct protein folding completely (Turoverov et al., 2010).

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 on conformational freedom. On the other hand, 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 ϕ and ψ angles. Therefore, the possibility of a given polypeptide chain folding into a compact state is determined by its ability to form numerous intramolecular contacts that differ in their physical nature, to compensate for the free-energy increase due to the decrease in the entropy component (Finkelstein and Ptitsyn, 2002). Although a native state of a globular protein has a clearly defined and unique three-dimensional structure,

the folding and ordering degree can vary greatly for the various parts of a given protein. In x-ray data this is seen from the *B*-factor values, which characterize the mobility of separate atoms (Berman et al., 2000). For example, the atoms in the active center of an enzyme typically have high *B*-factors. Furthermore, some globular proteins are shown to have disordered, highly dynamic fragments (i.e., termini, loops, etc.) which could not be detected by x-ray analysis, thus corresponding to regions of missing electron density.

The first direct evidence that all the information necessary for a given polypeptide chain to fold into a unique tertiary structure is encoded in its amino acid sequence was obtained by Anfinsen's group (Anfinsen, 1973). The reduced and urea-denatured ribonuclease A was shown to restore its native structure and functional state completely after 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 essence, protein folding can be regarded as a second part of the genetic code, as the protein amino acid sequence contains information about its functional three-dimensional structure.

The folding of a typical globular protein occurs on a millisecond-to-second time scale. However, for a small protein consisting of 100 amino acid residues, a simple search for a native state (which meets the requirements of the free-energy minimum) among all the alternative conformations would take a billion years. This contradiction represents the essence of Levinthal's paradox (Levinthal, 1968), which is resolved by the fact that amino acid sequences bear information related not only to their native structures but also to the pathways of their formation.

Some globular proteins fold into a unique globular structure only after ligand binding. A typical example of such proteins is a globular actin, which denatures after the removal of ligands (Ca^{2+} and ATP) (Kuznetsova et al., 1999; Altschuler and Willison, 2008). It is very likely that such proteins could not attain ordered structured spontaneously and should therefore be classified as IDPs.

1.2.2 Protein-Folding Models

Several models of protein folding have been developed (Turoverov et al., 2010). The *nucleation and growth model* is based on the assumption that protein folding is similar to the crystallization process and that the limiting step in the folding process is nucleus formation (Radford, 2000). This model describes the folding of small single-domain proteins that follow the "all-or-none" principle.

The *sequential protein-folding model*, also known as the *framework* or *hierarchic model*, was proposed in 1973 by O.B. Ptitsyn (1973). It suggests that folding starts with the backbone forming secondary structure elements, which then interact to form a more advanced folding intermediate; the specific packing of the side chains concludes the process. Each stage of the folding process stabilizes the major structural elements formed in the preceding state,

suggesting the existence of several folding intermediates. Therefore, long before the appearance of convincing experimental data, Ptitsyn put forward the idea of a partially folded conformation that serves as a universal folding intermediate. Eight years later, such a folding intermediate was found in a test tube (Dolgikh et al., 1981) and termed the *molten globule state* (Ohgushi and Wada, 1983). Other partially folded intermediates (e.g., pre-molten globule and highly ordered molten globule) were found later (Uversky, 2003b).

According to the current view, protein folding is realized via different pathways that are determined by the protein's energy landscape (Radford, 2000; Jahn and Radford, 2005). This landscape describes the dependence of the free energy on all the coordinates determining the protein conformation. The number of conformational states accessible by a polypeptide chain is reduced while approaching the native state. Therefore, this energetic surface is often called the *energy funnel* (Figure 1.2). The unfolded polypeptide

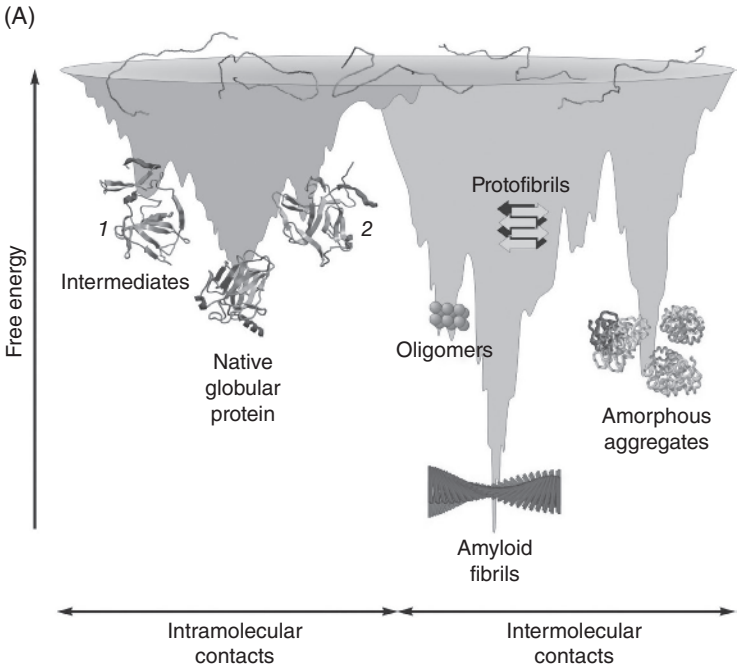

FIGURE 1.2 The energy landscape model, illustrating the formation of native globular and intrinsically disordered proteins, supramolecular complexes, amorphous aggregates, and amyloid fibrils. (A) Globular proteins. In the globular protein folding, the increase in the free energy associated with the folding-induced entropy decrease is compensated by the formation of specific intramolecular contacts. Local free-energy minima at the energy landscape correspond to the formation of partially folded intermediates. Intermolecular contacts of partially folded protein molecules can result in the formation of oligomers, amorphous aggregates, or amyloid fibrils. (*See insert for color representation of the figure*.)

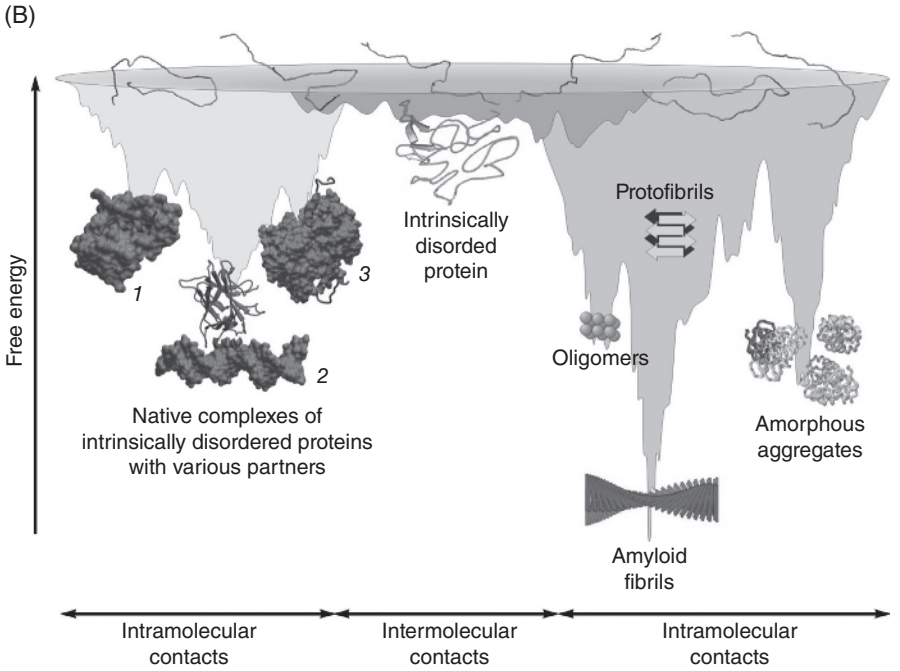

FIGURE 1.2 (*Continued*) (B) Intrinsically disordered proteins. Many native proteins with distinctive biological functions lack compact globular structure in aqueous solutions. Disordered segments of these proteins can gain ordered structure at the interaction with specific binding partners in a case if the free energy of such complexes is lower than the free energies of the intrinsically disordered protein and its partner. The propensity of native completely or partially disordered proteins to interact with various partners determines their biological functions in recognition of various binding partners (ligands, nucleic acids, and other proteins), in regulation of almost all cellular processes, and in signal transduction. In contrast to the folded globular proteins, which have to unfold to become amyloidogenic, disordered proteins seem always to be ready for such intermolecular interactions. 1, 2, and 3 represent native complexes of intrinsically disordered proteins with various partners. This figure is based on the energy funnel model developed for globular proteins (Schultz, 2000; Jahn and Radford, 2005). [Adapted from Turoverov et al. (2010).] (*See insert for color representation of the figure*.)

chain's free energy represents a large "hilly plateau" describing the dynamic ensemble of a large number of conformations. Hills on the plateau correspond to the forbidden conformations, and the plateau is separated from the entrance to the folding funnel by high energetic barrier(s), corresponding to the transitional state(s) (Finkelstein and Ptitsyn, 2002). This barrier is of great importance for 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.

In principle, the transition from the unfolded state to the uniquely folded native state can be realized via different pathways. Frequently, it seems that the experimental data prove the validity of this hypothesis, but in fact the rates of the various folding–unfolding stages might change dramatically, depending on the nature of the denaturing conditions (or the concentration of a given denaturant). As a result, not all of the folding or unfolding steps can be registered experimentally, suggesting that the character of the unfolding pathway depends on the choice of the denaturant (or denaturant concentration). At the same time, our experimental data on several proteins, including actin, showed that the number and the order of appearance of intermediate states is not dictated by denaturing agents (Kuznetsova et al., 2005; Povarova et al., 2007).

The energy landscape model not only elucidates the mechanisms of the globular protein folding, but also explains the nature of the IDPs, describes the formation of their supramolecular complexes, and delineates the formation of potentially pathogenic oligomers, amorphous aggregates, and amyloid-like fibrils (Turoverov et al., 2010).

1.2.3 Polymer Aspects of Protein Folding

It is generally accepted that the protein molecule has a unique primary sequence, which governs its three-dimensional structure and ensures proper biological activity. In this respect, each protein represents a unique case. That is why an understanding of the effect of sequence variations on biological performance represents a difficult challenge. It is also thought that natural polypeptides have originated as random copolymers of amino acids and were evolutionarily adjusted or "edited" (based on the principle of natural selection) to acquire and refine their various unique three-dimensional structures and functional properties (Ptitsyn and Volkenstein, 1986; Lau and Dill, 1990; Ptitsyn, 1995). Since protein molecules are remarkably unique, a serious question arises as to the existence of any general features in protein self-organization. To understand the common physicochemical principles underlying the protein-folding process, it is important to delineate the common polymer roots and their impact on the protein structures. The traditional way of performing such an analysis is a determination of the correlation between different physical characteristics of a polymer (e.g., its molecular density) and its length.

The molecular dimensions of 180 proteins in a variety of conformational states have been analyzed to establish a potential correlation between

the hydrodynamic dimensions and the length of the polypeptide chain (Tcherkasskaya and Uversky, 2001; Uversky, 2002a, 2002c, 2003b; Tcherkasskaya et al., 2003; Turoverov et al., 2010). The protein categories analyzed included native globular proteins with nearly spherical shapes, equilibrium molten globules (MGs), compact denatured [or pre-molten globules (pre-MGs)], denaturant-unfolded proteins without cross-links in the presence of strong denaturants (8 M urea or 6 M GdmHCl), and extended intrinsically disordered proteins (native coils and native pre-MGs). Figure 1.3 represents the results of this analysis and shows that 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 is observed. This gave rise to a set of standard equations $R_S = K_h M^{\varepsilon}$ for a number of conformational states of a polypeptide chain (Uversky, 2002a, 2002c, 2003b; Tcherkasskaya et al., 2003):

$$\log R_S^{\text{globular}} = -(0.204 \pm 0.023) + (0.357 \pm 0.005) \cdot \log(M) \quad (1)$$

$$\log R_S^{\text{MG}} = -(0.053 \pm 0.094) + (0.334 \pm 0.021) \cdot \log(M) \quad (2)$$

$$\log R_S^{\text{pre MG}} = -(0.21 \pm 0.18) + (0.392 \pm 0.041) \cdot \log(M) \quad (3)$$

$$\log R_S^{\text{native-coil}} = -(0.551 \pm 0.032) + (0.493 \pm 0.008) \cdot \log(M) \quad (5)$$

$$\log R_S^{\text{native pre-MG}} = -(0.239 \pm 0.055) + (0.403 \pm 0.012) \cdot \log(M) \quad (6)$$

$$\log R_S^{8\text{ M urea}} = -(0.649 \pm 0.016) + (0.521 \pm 0.004) \cdot \log(M) \quad (7)$$

$$\log R_S^{6\text{ M GdmHCl}} = -(0.723 \pm 0.033) + (0.543 \pm 0.007) \cdot \log(M) \quad (8)$$

This analysis showed that for a given conformational state, parameters K_h and ε were invariable over a wide range of chain lengths. Furthermore, statistical analysis has revealed that the relative errors of the recovered approximations exhibited random distribution over a wide range of chain lengths and did not generally exceed 10% (Uversky, 2002a,c, 2003b; Tcherkasskaya et al., 2003). This meant that the effective protein dimensions in a variety of conformational states can be predicted based on the chain length, with an accuracy of 10%. Thus, regardless of the differences in the amino acid sequences and biological functions, protein molecules behave as polymer homologs in a number of conformational states.

It is important to remember that the most unambiguous characteristic of the conformation of a polymer molecule remains the molecular density. For example, the density of a globule is expected to be independent of chain length, whereas the density of a partially collapsed or swelled macromolecules depends on both the chain length (and therefore on its molecular weight M)

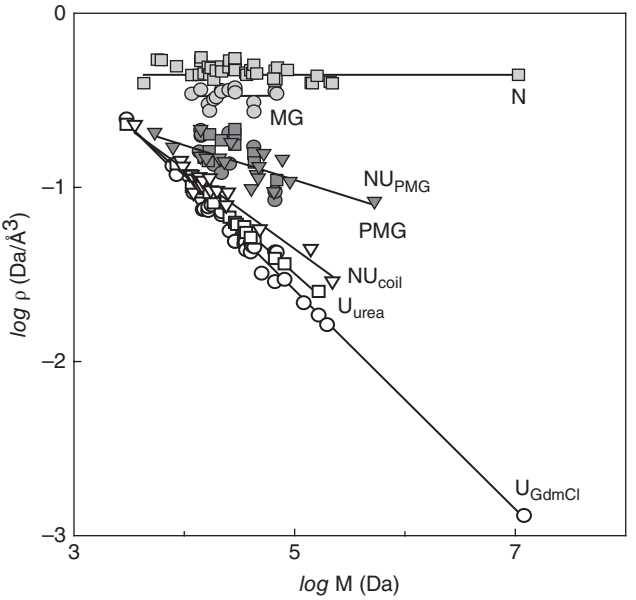

FIGURE 1.3 Variation of the density of protein molecules, ρ, with protein molecular weight, M, for a number of thermodynamically stable conformational states: native, molten globule; pre-molten globules (intermediates accumulated during the unfolding by urea or GdmCl are shown by circles; proteins with intact disulfide bridges in 8 M urea or 6 M GdmHCl are shown as squares; native pre-molten globules are shown as reversed triangles); native coils; proteins without cross-links or with reduced cross-links unfolded in 8 M urea; proteins without cross-links or with reduced cross-links unfolded in 6 M GdmHCl. The proteins used for this analysis are from Tcherkasskaya et al. (2003), Uversky (2002a,c), and Uversky (2003b). The solid lines represent the best fit of the data to the standard function $R_S = K_h M^{\varepsilon}$.

and on the nonspecific interactions of the monomer units with the solvent (Grossberg and Khohlov, 1989).

1.2.4 Different Conformations Seen in Protein Folding

Unfolded State: General Information on the Unfolded State of Proteins. Obviously, to solve the problem of protein folding it is necessary first to characterize the unfolded state, which represents the starting point of the folding reaction. It is known that the unfolded state represents an ensemble of rapidly interchanging conformations, some of which are extended and some of which are 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 (Baldwin, 1986). In this respect, it is important to mention that theoretical studies have shown that small preferences for native-like interactions in the unfolded state will substantially increase the probability of reaching the native state.

Returning to the polymer roots, under conditions known as *ideal* or *Θ-conditions* (i.e., 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 $R_S \sim lN^{0.5}$, with l being a statistical chain length (Tanford, 1961, 1968; Grossberg and Khohlov, 1989). Here the polymer is assumed to be in a random coil conformation, and its conformational behavior can be described using Gaussian statistics (Grossberg and Khohlov, 1989). Further, in a good solvent, the macromolecular coil is expanded due to the prevalence of the repulsive interactions between polymer segments, and the molecular dimensions change more significantly with increasing chain length, $R_S \sim (l^2B)^{0.2}N^{0.6}$.

Residual Structure in Proteins Unfolded by Strong Denaturants

Note that the "fully" unfolded states induced by GdmHCl or urea provide $\varepsilon = 0.54$ and 0.52, respectively (Tanford, 1961, 1968; Grossberg and Khohlov, 1989). Given that these ε values are less than 0.6, it appears that under these conditions the unfolded polypeptide chains exhibit features of macromolecular coils in *Θ*-solvents. This observation is in good agreement with the results of earlier studies (Tanford, 1968; Smith et al., 1996), where, in particular, it has been shown that the size of unfolded proteins in 6 M GdmCl is in reasonable agreement with the random coil model (Tanford, 1968). Recently, this conclusion was supported further by an 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 (Kohn et al., 2004). 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, suggesting that the mean dimensions of chemically denatured proteins are effectively indistinguishable from the mean dimensions of a random coil ensemble (Kohn et al., 2004).

However, it has been also 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 (Fitzkee and Rose, 2004). In fact, analysis of model systems where 33 proteins of known structure were used to generate disordered conformers by varying backbone torsion angles at random for approximately 8% of the residues, with the remaining approximately 92%

of the residues being remained fixed in their native conformation, revealed that despite this extreme degree of imposed internal structure, the ensembles analyzed were shown to have end-to-end distances and mean radii of gyration that agree well with random coil expectations (Fitzkee and Rose, 2004). Furthermore, the values of the hydrodynamic dimensions that Tanford measured for the unfolded proteins (Tanford, 1968) correspond better to a model where 20% of the residues are located in the collapsed structures (Miller and Goebel, 1968). In agreement with these observations, more recent Monte Carlo analysis of the effects of two types of local structures, α-helix and polyproline II (PPII) helix, on the dimensions of random coil polyalanine chains viewed as a model of highly denatured proteins revealed that although Flory's power-law scaling, long regarded as a signature of random coil behavior, holds for chains containing up to 90% α-helix or PPII, the absolute magnitude of the chain dimensions is sensitive to helix content (Wang et al., 2007). Furthermore, this study showed that the increase in α-helix content was accompanied by detectable chain contraction, reaching a minimum radius at about 70% α-helix, after which the chain dimensions expand rapidly. The presence of about 20% α-helix generated an unfolded state with the experimentally observed radii of gyration. Importantly, at an α-helix content of about 87%, polypeptide chains were also characterized by R_g values similar to those measured in experiments. On the other hand, R_g increased monotonically with increasing PPII content, always being more expanded than the dimensions observed experimentally. These results suggest that PPII is unlikely to be the sole dominant preferred conformation for unfolded proteins (Wang et al., 2007).

A great many protein-folding studies have revealed the presence of an assured residual structure even under the most severe denaturing conditions, such as high concentrations of strong denaturants. For example, considerable residual structure involving both α-helical and β-structural elements has been detected in the staphylococcal nuclease (SNase) (James et al., 1992; Alexandrescu et al., 1994; Wang and Shortle, 1995, 1996, 1997; Gillespie and Shortle, 1997a, b; Shortle and Ackerman, 2001). A stretch of continuous nonpolar residues ranging between Ile 95 and Tyr 103, forming β-strand 3, was shown to be involved in the formation of a compact cluster in the unfolded state of the tryptophan synthetase α-subunit (Saab-Rincon et al., 1993, 1996). Similarly, a contiguous stretch of nonpolar residues comprising Val 54, Val 56, Trp 58, and Leu 59 was found to form a cluster in a urea-unfolded fragment of the protein 434 (Neri et al., 1992a, b, c). Analysis of the urea-unfolded human fibroblast growth factor 1 (hFGF-1), a protein that consists of 12 antiparallel β-strands arranged into a β-trefoil architecture, by direct measurement of the hydrogen/deuterium (H/D)-exchange rates revealed that 38 residues whose heteronuclear single quantum coherence cross-peaks can

be observed after exchange show higher protection level 2 than those predicted for the same residues in a random coil conformation, suggesting the existence of residual structure(s) (Wang and Yu, 2011).

Investigations of the SH3 have identified turn-like structure in regions that fold into β-strands (Zhang and Forman-Kay, 1995, 1997). The existence of defined residual structure has been observed in unfolded states of barstar (Nolting et al., 1997) and barnase (Arcus et al., 1995). The WW-domain retains a native-like core in high concentrations of GdmCl and urea (Koepf et al., 1999). The unfolded states of BPTI have been reported to be relatively compact in 6 M GdmCl under reducing conditions (Amir et al., 1992; Gottfried and Haas, 1992). Human carbonic anhydrase II, HCA II, has also been shown to possess a compact unfolded state (Martensson et al., 1993; Svensson et al., 1995; Hammarstrom and Carlsson, 2000). Unfolded apomyoglobin (8 M urea at pH 2) displays distinct regions with dramatically different backbone mobility, assuming the existence of residual structure (Schwarzinger et al., 2002). Extensive clusters of hydrophobic structure exist in the unfolded state of lysozyme even under strongly denaturing conditions (Klein-Seetharaman et al., 2002). The residual structure in fully unfolded photoactive yellow protein (PYP) was shown to be affected by isomerization of its *p*-coumaric acid (pCA) chromophore, suggesting that the residual structure in the fully unfolded state of PYP can be controlled directly by pCA photoexcitation (Lee et al., 2010). The *Escherichia coli* outer membrane protein X (OmpX) contains two polypeptide segments that present nonrandom residual structure in 8 M aqueous urea, whereas the remainder of the protein is in a flexibly disordered conformation (Tafer et al., 2004). Residual structure was also found in the chemically unfolded state of the N-terminal domain of a multidomain protein enzyme I from *Streptomyces coelicolor* (Romero-Beviar et al., 2010).

NMR analysis of urea-induced unfolding of the molten globule state of α-lactalbumin revealed a remarkably stable part of the protein in the core of the helical domain, which comprises interactions in both the N- and C-terminal parts of the protein (Schulman et al., 1997). More recent comparison of the unfolded states of three homologous proteins with very similar folds—the reduced unfolded states of hen lysozyme and the calcium-binding proteins bovine and human α-lactalbumins—by heteronuclear NMR spectroscopy showed the presence of noticeable residual structure manifested as significant deviations from random coil predictions (Wirmer et al., 2006). Furthermore, although these homologous proteins possessed very similar structures in their native states, their unfolded states differed noticeably from each other. For example, the residual structure was found in the α- and the β-domains in

bovine α-lactalbumin and lysozyme, and only in the α-domain of human α-lactalbumin (Wirmer et al., 2006).

Using a combination of NMR experiments and molecular dynamics simulations it was shown that although the denatured state of chymotrypsin inhibitor 2 is highly unfolded, it contains some residual native helical structure along with hydrophobic clustering in the center of the chain (Kazmirski et al., 2001). Residual dipolar couplings (RDCs) for HN–N and HA–CA bond vectors measured by NMR spectroscopy for the 70-residue protein eglin C revealed the existence of a native-like global structure in the urea unfolded state (Ohnishi et al., 2004). Based on ^{19}F NMR analysis it has been concluded that a residual structure in unfolded intestinal fatty acid–binding protein with incorporated fluorinated aromatic amino acids consists of amino acids that are neighbors in the native state (Ropson et al., 2006). In yeast alcohol dehydrogenase (YADH) with the cysteine residues covalently modified by *N*-(1-pyrenyl) maleimide (PM) residual structure was detected even in the presence of 5 M GdnHCl using the excimer fluorescence of PM-YADH (Santra et al., 2006). By a combination of circular dichroism (CD) and small-angle x-ray scattering (SAXS), some residual structure was found in the unfolded state of the HIV-1 protease (Kogo et al., 2009).

In the unfolded state of the ultrafast folder "Trp-cage" miniprotein TC5b, residual structure due to hydrophobic collapse, with strong interresidue contacts between side chains that are relatively distant from one another in the native state, was detected in photochemically induced dynamic nuclear polarization NMR pulse-labeling experiments (Mok et al., 2007). When a general photochemical modification with methylene carbene was used to evaluate the accessible surface area (ASA) of the polypeptide chain in different conformational states of *Bacillus licheniformis* β-lactamase (BL-betaL), the unfolded state in 7 M urea was labeled just 60% more than the native state (Ureta et al., 2007). Since the retrieved value was well below the increment of ASA expected from theoretical estimates, it has been concluded that the unfolded BL-betaL possessed noticeable residual organization (Ureta et al., 2007).

Comparative analysis of the unfolded states of the ribosomal protein S16 isolated from a mesophilic ($S16_{meso}$) and a hyperthermophilic ($S16_{thermo}$) bacterium by a variety of biophysical methods revealed that the denatured-state ensemble of $S16_{thermo}$ is more compact relative to $S16_{meso}$, suggesting that the extreme temperature tolerance of a hyperthermophilic protein is coupled to residual structure in its unfolded state (Wallgren et al., 2008). Similarly, detailed comparison of unfolded states of ribonucleases H from organisms that grow at different temperatures revealed that a tolerance to higher temperatures correlates well with the residual structure in the unfolded state of the thermophilic proteins (Ratcliff and Marqusee, 2010).

Using a combination of residual dipolar coupling (RDC) and paramagnetic relaxation enhancement (PRE) data as constraints in ensemble structure calculations, it was shown that only a relatively small number of conformers is necessary to fully reproduce the experimental RDCs, PREs, and average radius of gyration of the urea-unfolded ubiquitin (Huang and Grzesiek, 2010). This conformational ensemble was characterized by a characteristic residual structure corresponding to an intact native first β-hairpin and α-helix combined with the nonnative α-helical conformations in the C-terminal half of the protein (Huang and Grzesiek, 2010). Therefore, by combining modern RDC, PRE, and computational methods, statistically significant detection of subconformations in the unfolded ensemble at population levels of a few percent is achievable (Huang and Grzesiek, 2010).

All these facts show unambiguously that the existence of profound residual structure might be a general characteristic of an unfolded polypeptide chain under aggressively denaturing conditions (Dill and Shortle, 1991; Shortle, 1996a, b; Baldwin and Zimm, 2000; Plaxco and Gross, 2001). This residual structure may be important in the folding process. 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 high affinity for each other in a folded protein (Hammarstrom and Carlsson, 2000). In addition, some secondary structure within unfolded protein could be expected, due to the preferential distribution of ϕ and ψ angles (Ramachandran and Sasisekharan, 1968; Smith et al., 1996; Shortle, 2002); and some residual hydrophobic interactions can also be present (Hammarstrom and Carlsson, 2000; Shimizu and Chan, 2002). All this restricts considerably the conformational space of the unfolded polypeptide chain. Thus, it seems most likely that the polypeptide chains under "strong denaturing conditions" are still below the critical point (bad solvent conditions) and can easily be transformed to the compact state. For example, a small fluctuation of temperature from 25 to 30°C was shown to encourage the cooperative collapse of the fully unfolded proteins in 6 M GdmCl (Ptitsyn, 1995). Thus, globular proteins are never random coils without positional correlations, and biological polypeptide chains represent the macromolecular coils below a critical point even under harsh denaturing conditions.

Nonglobular Pre-molten Globule State Now let us consider a situation when the thermodynamic quality of the solvent worsens. In this case the binary interactions between monomers become mainly attractive (Tcherkasskaya et al., 2003). As a result, the polymer chain collapses partially, leading to an increase in the molecular density and bringing

many-body interactions into the scene. It has been found that many proteins under the appropriate conditions can form a specific compact partially folded conformation, a pre-molten globule state (Palleros et al., 1993; Uversky and Ptitsyn, 1994; Ptitsyn et al., 1995; Uversky and Ptitsyn, 1996a; Zhang et al., 1996; Karnoup and Uversky, 1997; Uversky et al., 1998a; Bushmarina et al., 2001; Tcherkasskaya and Uversky, 2001; Kuznetsova et al., 2002; Georlette et al., 2003a, b; Tcherkasskaya et al., 2003). Major structural characteristics of this intermediate are summarized below. This conformation is characterized by considerable secondary structure, although much less than that of the molten globule. The pre-molten globule state is considerably less compact than the molten globule state, but it is still more compact than the corresponding random coil. Pre-molten globules can interact with the hydrophobic fluorescent probe ANS, although more weakly than in the molten globule state. This means that at least some hydrophobic clusters are already formed in the pre-molten globule state, although there is no globular structure.

Finally, the pre-molten globule and the molten globule are separated by an all-or-none phase transition, reflecting the fact that these partially folded intermediates represent discrete phase states (Uversky et al., 1992; Uversky and Ptitsyn, 1996b). Importantly, several structural elements of pre-molten globules, these squeezed coils, may occupy native-like positions (Tcherkasskaya and Uversky, 2001). We assume that this compact denatured state might represent a general intermediate in protein folding. Obviously, 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 this state to a molten globule state or native state would not require significant energy changes and could occur quite easily. It may well be the case that the protein folding *in vivo* involves mostly the molten globule and the denatured compact states. On this occasion it would require only slight changes to make the coil "condense onto itself" and form a globule.

Analysis of hydrodynamic data reveals that the molecular dimensions of pre-molten globules follow the chain length as $R_S = 0.6M^{0.40}$ (see above). This indicates the bad solvent conditions. Thus, one can conclude that this conformation exhibits behavior that is typical for squeezed macromolecular coils. Therefore, any small variations in the protein environment (i.e., changes in the thermodynamic quality of the solvent, or changes induced by proton transfer, interactions with a ligand, fluctuations of temperature, etc.) can trigger the transition of the compact protein molecule to the more rigid molten globule or native states (Grossberg and Khohlov, 1989). Importantly, Figure 1.3 shows that stabilization of the pre-molten globule state might be achieved by the incorporation of disulfide bridges into the protein sequence.

In fact, the hydrodynamic behavior of unfolded proteins with intact disulfide bridges was shown to be close to that of pre-molten globules. Finally, the pre-exponential term $K_h = 0.6$ retrieved for the pre-molten globule state is significantly larger than those retrieved for the fully unfolded species, which probably indicates developing multiple body interactions.

Globular Molten Globule and Ordered States The theory of *coil–globule transition* predicts that the overall dimension of a polymer globule, R_S, is anticipated to change with the chain length, N, as $R \sim (C/B)^{1/3}N^{1/3}$. Here B and C are the second and third virial coefficients, which characterize the pair collisions and three-body interactions of the monomer units of the polymer chain (Grossberg and Khohlov, 1989). The density of the globules is expected to show no change with increasing chain length, owing to $\rho \sim N/R^3 \sim (-B/C)$. These results are in excellent agreement with data obtained for the ordered and molten globules of proteins (Figure 1.3, see also Tcherkasskaya and Uversky, 2001; Uversky, 2002a,c, 2003b; Tcherkasskaya et al., 2003). In particular it has been found that for ordered and molten globule proteins, the parameter K_h has a value of 0.75 and 0.9, respectively, whereas ε equals 0.33 for both conformational states. Given that the overall dimensions of globules are determined by the balance of pair collisions and three-body interactions, the difference in preexponential terms K_h observed for the native and molten globules probably reflects the larger probability of three-body interactions in the molten globule state. This could be because of the compact but flexible nature of the molten globules.

Amino Acid Composition Favoring Equilibrium Partially Folded Intermediate(s) A variety of different physicochemical forces play a role in stabilizing the unique three-dimensional structure of a protein. Both the strength and specificity of many of these forces are strongly dependent on environmental conditions in such a way that changes in the environment can reduce or even eliminate some of the conformational interactions, while the remaining interactions are unchanged or even intensified. Under some environmental conditions, the native protein structure can be transformed into new conformations with properties intermediate between those of the native and completely unfolded states. Thus, the ability of a protein to adopt different stable partially folded conformations should be considered as an intrinsic property of a polypeptide chain. Since all the necessary and sufficient information to fold into the native, biologically active conformation is thought to be present in a protein amino acid sequence (Anfinsen et al., 1961), the capability of a given protein to adopt equilibrium partially folded conformation(s) may also be encoded in specific features of its amino acid sequence.

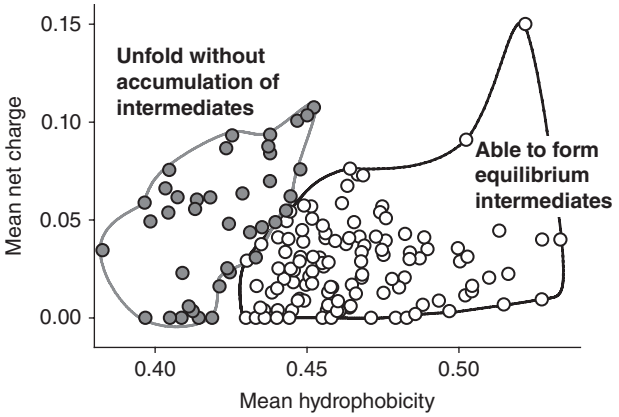

FIGURE 1.4 Comparison of mean net charge vs. mean hydrophobicity for the set of 115 proteins able to form equilibrium intermediates (open symbols and black lines) and the set of 39 proteins shown to unfold without accumulation of partially folded conformations (gray symbols and lines). [Modified from data presented by Uversky (2002b).]

Interestingly, it has been shown that not all proteins (even homologous ones) have an identical response to changes in their environment. For example, hen egg white lysozyme represents a textbook illustration for the two-state model of denaturant-induced unfolding (Tanford, 1968), whereas accumulation of classical molten globules under different experimental conditions was described for its homolog, α-lactalbumin (Dolgikh et al., 1981, 1985; Kuwajima, 1996). Analysis of literature data on equilibrium unfolding of globular proteins induced by changes in pH, temperature, or strong denaturants (urea or guanidinium chloride) revealed that unfolding in 115 proteins is accompanied by accumulation of equilibrium intermediate states of one sort or another. Another set comprises 39 proteins, which were shown to unfold according to a simple two-state model; that is, no equilibrium intermediate of any kind was formed during their unfolding (Uversky, 2002b).

In an attempt to understand which factors may be responsible for such tremendous differences in the formation of equilibrium partially folded intermediates, the general sequence features of proteins from both groups have been analyzed using a simple method comparing global sequence charge and hydrophobicity (Uversky, 2002b). Figure 1.4 represents the results of this analysis as a plot of mean hydrophobicity versus mean net charge (i.e., as a distribution of groups within the charge-hydrophobicity phase space). Figure 1.4 shows that the reliable separation of both groups of proteins takes place in such coordinates. These data imply that the competency of a protein to form equilibrium intermediate(s) may be predetermined by the bulk content of hydrophobic and charged amino acid residues (Uversky, 2002b).

In other words, this competency may be encoded in the charge/hydrophobicity ratio of its polypeptide chain, not its sequence. This may mean that partially folded conformations are stabilized mostly by nonspecific side-chain/side-chain interactions of hydrophobic amino acid residues (Uversky, 2002b). Interestingly, proteins that do not have equilibrium intermediates are less hydrophobic and have, in general, a larger net charge than those competent to form discrete intermediate states. This may indicate that such proteins are less strengthened by hydrophobic interactions and more disturbed by electrostatic repulsion. Thus, smaller environmental changes may be required to overcome the marginal stabilization energies, leading to immediate and complete unfolding of the protein.

Concluding Remarks on Protein Folding Regardless of the differences in primary amino acid sequences, protein molecules behave as polymer homologs in a number of conformational states, allowing speculation as to the volume interactions being a driving force in a formation of equilibrium structures. For example, both native and molten globules exhibit key features of polymer globules, where fluctuations of the molecular density are expected to be much less than the molecular density itself. Protein molecules in the pre-molten globule state possess properties of squeezed coils. Furthermore, even high concentrations of strong denaturants are more likely to constitute bad solvents for protein chains. Thus, globular proteins are never random coils without positional correlations, and biological polypeptide chains represent the macromolecular coils below a critical point, even under harsh denaturing conditions.

1.3 NONFOLDING

1.3.1 Intrinsically Disordered Proteins and Their Abundance

It is recognized now that a considerable number of biologically active proteins are not completely rigid but possess some amount of disorder under physiological conditions. These proteins are known as *intrinsically disordered proteins* (IDPs), among other names. Structurally, IDPs are highly diverse, and some compact IDPs possess noticeable secondary structure and behave as native molten globules, whereas other IDPs are extended and possess little residual structure (i.e., these IDPs behave as native coils or native pre-molten globules) (Uversky, 2011a; Uversky and Dunker, 2010).

IDPs and intrinsically disordered protein regions (IDPRs) constitute a unique tribe of the protein kingdom (Uversky, 2003b, 2011a, Turoverov et al., 2010;

Tompa, 2011), and the number of IDPs and IDPRs is amplifying rapidly. For example, the first list of extended IDPs (also known as natively unfolded proteins, i.e., proteins that do not possess almost any specific structure under physiological conditions), published in 2000 (Uversky et al., 2000), contained 91 members (the full-length proteins and their domains with a chain length of more than 50 amino acid residues). Now this set has almost doubled.

The commonness of intrinsic disorder was estimated further by predicting disorder for whole genomes, including both known and putative protein sequences. First predictions of such type have been published for 31 genomes that span the three kingdoms: archaea, bacteria, and eukaryotes. The percentage of sequences in each genome with segments predicted to have ≥ 40 consecutive disordered residues was used to gain an overview of proteomic disorder. For so many consecutive predictions of disorder, the false-positive error rate was estimated from ordered proteins to be less than 0.5% of the segments of 40 and less than 6% of the fully ordered proteins (Dunker et al., 2000, 2001). The eukaryotes exhibited more disorder by this measure than either the prokaryotes or the archaea, with *C. elegans*, *A. thaliana*, *S. cerevisiae*, and *D. melanogaster* predicted to have 52 to 67% of their proteins with such long predicted regions of disorder, while bacteria and archaea were predicted to have 16 to 45% and 26 to 51% of their proteins with such long disorder regions, respectively (Dunker et al., 2000; Oldfield et al., 2005a). The increased amount of disorder in the eukaryota may be related to the increase in cellular signaling in the eukaryota (Dunker et al., 2000, 2001; Oldfield et al., 2005a). Recently, this analysis was extended to include entire proteomes of 3484 species from the three main kingdoms of life and in viruses (Xue et al., 2012). This analysis revealed that the process of evolution is characterized by the unique patterns of changes in the protein intrinsic disorder content. In fact, viruses were characterized by the widest spread of the proteome disorder content (the number of disordered residues ranges from 7.3% in human coronavirus NL63 to 77.3% in *avian carcinoma virus*) (Xue et al., 2012). For several organisms, 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. For both prokaryotes and eukaryotes, the disorder content was generally independent of proteome size. However, it showed a sharp increase associated with the transition from prokaryotic to eukaryotic cells (see Figure 1.5). This suggests 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 (Xue et al., 2012).

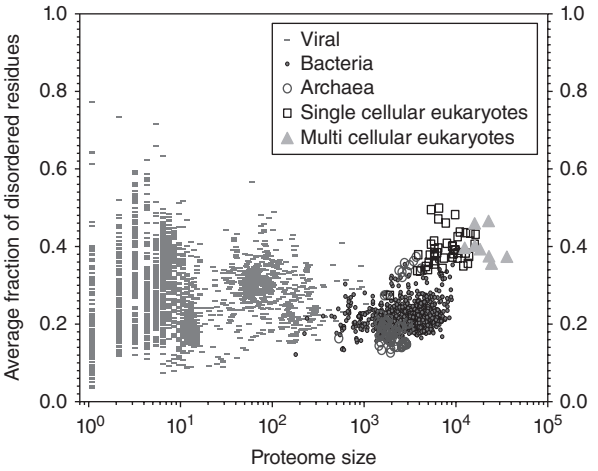

FIGURE 1.5 Correlation between the intrinsic disorder content and proteome size for 3484 species from viruses, archaea, bacteria, and eukaryotes. Each symbol indicates a species. There are a total of five groups of species: viruses (red horizontal marks), bacteria (small green circles), archaea (open blue circles), unicellular eukaryotes (open brown squares), and multicellar eukaryotes (pink triangles). The proteome size is the number of proteins in the proteome of that species and is shown on a logarithmic scale. The average fraction of disordered residues is calculated by averaging the fraction of disordered residues of each sequence over all the sequences of that species. Disorder prediction is evaluated by PONDR-VSL2B. [Adapted from Xue et al. (2012).] (*See insert for color representation of the figure*.)

1.3.2 Some Functional Advantages of IDPs

For several reasons, the intriguing family of IDPs has attracted an excessive attention of researchers over the past decade and a half. First, the existence of IDPs, with their lack of ordered rigid structure under "physiological" conditions *in vitro*, disputes the validity of one of the cornerstones in protein biology, chemistry, and physics (i.e., the structure–function paradigm). Note that this paradigm, formulated more than 100 years ago as a lock-and-key concept for explaining the amazing specificity of the enzymatic hydrolysis of glucosides (Fischer, 1894), claims that a specific function of a protein is determined by its unique and rigid three-dimensional structure. IDPs, being disordered (at least in part), are involved in countless biological activities (see below) and effectively resist those evolutionary pressures that normally favor stable globular folds. For example, α-synuclein (one of the best-characterized IDPs) isolated from different organisms possesses a high degree of sequence conservation; for example, the mouse and rat α-synucleins are identical throughout the first 93 residues, and the human and canary proteins differ from them by only two residues (Clayton and George, 1998). To overcome the conflict in the existence of functionally active unfolded polypeptides with

the structure–function paradigm, it has been suggested that the lack of rigid globular structure under physiological conditions might represent a considerable functional advantage for IDPs, as their large plasticity allows them to interact efficiently with several different targets (Wright and Dyson, 1999; Dunker et al., 2001, 2002a,b; Tompa, 2002, 2011; Uversky, 2002a,c, 2011a; Uversky, 2003b; Dyson and Wright, 2005; Uversky and Dunker, 2010). Furthermore, a disorder-to-order transition induced in IDPs during the binding of specific targets *in vivo* might represent a simple mechanism for the regulation of numerous cellular processes, including the regulation of transcription and translation and cell-cycle control (reviewed by (Wright and Dyson, 1999; Dunker et al., 2001, 2002a,b; Tompa, 2002; 2011; Uversky, 2002a,c, 2003b, 2011a; Dyson and Wright, 2005; Uversky and Dunker, 2010).

Therefore, in addition to being highly abundant in nature (Dunker et al., 2000; Romero et al., 1998; Uversky et al., 2000; Ward et al., 2004; Xue et al., 2012), IDPs/IDPRs possess numerous intriguing properties (Uversky and Dunker, 2010), are intimately involved in various cellular processes (Wright and Dyson, 1999, 2009; Dunker et al., 2001, 2002a,b, 2005; Iakoucheva et al., 2002; Tompa, 2002; Uversky, 2002a,c; Dyson and Wright, 2005; Uversky et al., 2005; Vucetic et al., 2007; Xie et al., 2007a,b; Kim et al., 2008; Oldfield et al., 2008; Liu et al., 2009), and are commonly found to be related to the pathogenesis of various diseases (Iakoucheva et al., 2002; Cheng et al., 2006; Xie et al., 2007a; Uversky, 2008, 2009; Uversky et al., 2008, 2009; Midic et al., 2009a, b). 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 (Wright and Dyson, 1999, 2009; Uversky et al., 2000; Dunker et al., 2001; Tompa, 2002; Uversky, 2002a,b, 2011a,b; 2013; Dyson and Wright, 2005; Oldfield et al., 2005b; Mohan et al., 2006; Vacic et al., 2007a; Dosztanyi et al., 2009; Meszaros et al., 2009). Furthermore, intrinsic disorder opens a unique capability for one protein to be involved in interaction with several unrelated binding partners and to gain different bound structures (Oldfield et al., 2008; Hsu et al., 2012). Some IDPs/IDPRs can form highly stable complexes; others are involved in signaling interactions, where they undergo constant "bound–unbound" transitions, thus acting as dynamic and sensitive "on–off" switches. These proteins typically return to their intrinsically disordered state after the completion of a particular function. Many of the IDPs/IDPRs can gain different conformations, depending on the environmental peculiarities (Oldfield et al., 2005b; Hsu et al., 2012). All this constitutes an important arsenal of the unique physiological properties of IDPs/IDPRs, which determines their ability to exert different functions in different cellular contests according to a specific conformational state (Uversky and

Dunker, 2010). The folding-at-binding principle is believed to help IDPs or IDPRs to obtain maximal specificity in a protein–protein interaction without very high affinity (Liu et al., 2009). This combination of high specificity with low affinity defines the broad utilization of intrinsic disorder in regulatory interactions, where turning a signal off is as important as turning it on (Dunker et al., 2001). Although some partial folding during the IDP/IDPR-based interactions is a widespread phenomenon, with a significant fraction (about one-third) of the interacting residues in IDPs/IDPRs adopting α-helix and β-strand structures (Mohan et al., 2006; Vacic et al., 2007a), there are still many other IDPs/IDPRs that are involved in the formation of "fuzzy complexes," where an IDP/IDPR keeps a certain amount of disorder in its bound conformations (Tompa and Fuxreiter, 2008; Hazy and Tompa, 2009; Meszaros et al., 2011; Uversky, 2011a,b).

Often, the interacting regions in IDPs are observed as loosely structured fragments in their unbound forms. These disorder-based binding sites are known as *molecular recognition elements* or *features* (MoREs or MoRFs) (Oldfield et al., 2005b; Mohan et al., 2006) preformed structural elements (Fuxreiter et al., 2004), or pre-structured motifs (PreSMos) (Chi et al., 2007). Although the existence of such loosely structured regions suggests that IDPs can adopt their bound structure(s) at a free-energy cost that is not too high, it is important to remember that increasing the stability of the bound conformation does not necessarily enhance the binding affinity (Wright and Dyson, 2009). Another important feature of the disorder-based interactions is their increased speed due to the greater capture radius and the ability to search spatially through interaction space (called the "fly-casting" mechanism) (Shoemaker et al., 2000), or due to the fewer encounter complexes required for the binding event with IDPs/IDPRs (Huang and Liu, 2009). Linking all these considerations with the recent report showing that IDP affinities are tuned mostly by association rates (Prakash, 2011) suggests that the degree of pre-adoption of binding conformations in IDPs has to be limited, but not unfavorable.

1.3.3 Function-Induced Folding of IDPs

An important feature of IDPs is that they are able to undergo a disorder-to-order transition (i.e., partial or complete folding) during or prior to their biological function (Wright and Dyson, 1999, 2009; Dunker et al., 2001, 2002a,b, 2005; Tompa, 2002, 2011; Uversky, 2002a,b, 2003b, 2011a,b; Dyson and Wright, 2005; Uversky et al., 2005; Tompa and Fuxreiter, 2008; Uversky and Dunker, 2010). In other words, IDPs are likely to be stabilized *in vivo* by function-related binding to specific targets and ligands (such as a variety of small molecules, substrates, cofactors, other proteins, nucleic acids,

membranes, etc.). The functional importance of being disordered has been analyzed intensively, and it has been established that increased intrinsic plasticity represents an important prerequisite for effective molecular recognition. The diapason of biological functions of the IDPs/IDPRs is extremely wide, including cell-cycle control, transcriptional and translational regulation, modulation of activity and/or assembly of other proteins, and even regulation of nerve cell function [reviewed in Dunker et al., 2001; 2002a, b; 2005; Dyson and Wright, 2002, 2005; Tompa, 2002; Uversky and Dunker, 2010; Uversky, 2011a; 2013]. It has been suggested that the persistence of IDPs throughout evolution may be due to the definite advantages of disorder-to-order transitions accompanying functional performance of flexible structures in comparison with action of rigid proteins (Dunker et al., 1998, 2001, 2005; Wright and Dyson, 1999; Dyson and Wright, 2002; Uversky et al., 2005; Uversky and Dunker, 2010). Some of these potential advantages of intrinsic lack of structure and function-related folding include:

1. The ability of binding to several different targets, known as one-to-many signaling
2. The possibility of high specificity coupled with low affinity
3. The precise control and simple regulation of the binding thermodynamic
4. The capability to overcome steric restrictions, enabling essentially larger interaction surfaces in the complex than could be obtained for the rigid partners
5. The increased rates of specific macromolecular association
6. The reduced lifetime of the intrinsically disordered proteins in the cell, possibly representing a mechanism of rapid turnover of the important regulatory molecules

Finally, a few words should be added about the extent of function-related folding of IDPs. It has been shown that the range of conformational changes induced in natively unfolded proteins as a result of their interaction with natural partners is very wide (Uversky, 2002a,c, 2011b). In fact, the examples of all possible conformational transitions have been described, including function-induced transitions of coil to pre-molten globule, coil to molten globule, coil to ordered conformation, pre-molten globule to molten globule, pre-molten globule to rigid structure, and molten globule to ordered and rigid form (Uversky, 2002a,c). This means that the structure–function paradigm, which emphasizes that ordered three-dimensional structures represent the indispensable prerequisite to the effective protein functioning, should be redefined to include IDPs (Wright and Dyson, 1999; Dunker et al., 2001; Uversky and Dunker, 2010). According to this redefined paradigm, native proteins

(or their functional regions) can exist in any of the known conformational states, ordered, molten globule, pre-molten globule, and coil. Function can arise from any of these conformations and transitions between them. Thus, not just the ordered state but any of the known polypeptide conformations can be the native state of a protein.

It was suggested that the structure-forming effect of natural partners might be explained by their influence on the hydrophobicity and/or net charge of the natively unfolded polypeptide. Obviously, these parameters could be changed in such a way that they will approach values typical of ordered native proteins forcing a polypetide to fold (Uversky et al., 2000); that is, the formation of protein–ligand complexes may increase a "critical mass" of the folding unit, thus initiating a chain of folding reactions.

There is, however, an alternative view of the problem of existence and the functionality of IDPs. It is possible that these proteins are folded in the cell as "normal" proteins, due to the fact that the concentration of their ligands is always high enough *in vivo* to fold them. Thus, evolution never sees them as any different from normal proteins. In this view, the property of being intrinsically disordered is merely an artifact resulting from the proteins being studied after being purified to homogeneity. Evolutionarily, IDPs persist not because there is an advantage to being intrinsically disordered, but because there is no disadvantage as long as there are enough ligands around *in vivo* to fold the proteins. The stability of the folded, ligand-bound form may only be a few kilocalories per mole, but this would be enough to keep the vast majority of molecules folded and functional (even most normal, folded proteins are only marginally stable).

1.3.4 IDPs and Human Diseases

Because of the fact that ID proteins play crucial roles in numerous biological processes, it was not too surprising to find that many of them are involved in human diseases. For example, a number of human diseases originate from the deposition of stable, ordered, filamentous protein aggregates, commonly referred to as *amyloid fibrils*. In each of these pathological states, a specific protein or protein fragment changes from its natural soluble form into insoluble fibrils, which accumulate in a variety of organs and tissues (Kelly, 1998; Bellotti et al., 1999; Dobson, 1999; Uversky et al., 1999b,c; Rochet and Lansbury, 2000; Uversky and Fink, 2004). Approximately 20 different proteins are known so far to be involved in these diseases. These proteins are unrelated in terms of sequence or starting structure. Several ID proteins are found in this list of 20, being associated with the development of several neurodegenerative diseases (Uversky and Fink, 2004, 2005). An incomplete list of disorders associated with ID proteins includes Alzheimer's disease [deposition of

amyloid-β, tau-protein, α-synuclein fragment NAC (Glenner and Wong, 1984; Masters et al., 1985; Lee et al., 1991; Ueda et al., 1993); Niemann–Pick disease type C, subacute sclerosing panencephalitis, argyrophilic grain disease, myotonic dystrophy, and motor neuron disease with neurofibrillary tangles (accumulation of tau-protein in the form of neurofibrillary tangles (Lee et al., 1991)]; Down's syndrome [nonfilamentous amyloid-β deposits (Wisniewski et al., 1985)]; Parkinson's disease, dementia with Lewy body, diffuse Lewy body disease, Lewy body variant of Alzheimer's disease, multiple-system atrophy and Hallervorden–Spatz disease (deposition of α-synuclein in the form of a Lewy body, or Lewy neuritis (Dev et al., 2003)]; prion diseases [deposition of PrP^{SC} (Prusiner, 2001)]; and a family of polyQ diseases, a group of neurodegenerative disorders caused by expansion of GAC trinucleotide repeats coding for polyQ in the gene products (Zoghbi and Orr, 1999). Furthermore, most mutations in rigid globular proteins associated with accelerated fibrillation and protein deposition diseases have been shown to destabilize the native structure, increasing the steady-state concentration of partially folded (disordered) conformers (Kelly, 1998; Bellotti et al., 1999; Dobson, 1999; Rochet and Lansbury, 2000; Uversky et al., 1999b,c; Uversky and Fink, 2004).

The maladies listed above have been called *conformational diseases*, as they are characterized by the conformational changes, misfolding, and aggregation of an underlying protein. However, there is another side to this coin: protein functionality. In fact, many of the proteins associated with the conformational disorders are also involved in recognition, regulation, and cell signaling. For example, functions ascribed to α-synuclein, a protein involved in several neurodegenerative disorders, include binding fatty acids and metal ions; regulation of certain enzymes, transporters, and neurotransmitter vesicles; and regulation of neuronal survival (reviewed by Dev et al., 2003). Overall, about 50 proteins and ligands have been shown to physically interact and/or co-localize with this protein. Furthermore, α-synuclein has amazing structural plasticity and adopts a series of different monomeric, oligomeric, and insoluble conformations (reviewed by Uversky, 2003a). The choice between these conformations is determined by the peculiarities of the protein environment, assuming that α-synuclein has an exceptional ability to fold in a template-dependent manner. Based on these observations, we hypothesize that the development of the conformational diseases may originate from misidentification, misregulation, and missignaling, accompanied by misfolding. In other words, mutations and/or changes in the environment may result in protein confusion, for which its ID becomes lost, thus reducing its capability to recognize proper binding partners and leading to the formation of nonfunctional and deadly aggregates.

Recent analysis of polyglytamine diseases gives support to this hypothesis (Okazawa, 2003). *Polyglytamine diseases* are a specific group of hereditary neurodegeneration caused by expansion of CAG triplet repeats in an exon of disease genes, which leads to the production of a disease protein containing an expanded polyglutamine (polyQ) stretch. Nine neurodegenerative disorders, including Kennedy's disease, Huntington's diseases, spinocerebellar atrophy-1, -2, -3, -6, -7, -17, and dentatorubral pallidoluysian atrophy are known to belong to this class of diseases (Cummings and Zoghbi, 2000; Gusella and MacDonald, 2000; Fischbeck, 2001; Orr, 2001). In most polyQ diseases, expansion to over 40 repeats leads to the onset (Fischbeck, 2001). It has been emphasized that such molecular processes as unfolded protein response, protein transport, and synaptic transmission and transcription are implicated in the pathology of polyQ diseases (Okazawa, 2003). Importantly, more than 20 transcription-related factors have been reported to interact with pathological polyQ proteins. Furthermore, these interactions were shown to repress the transcription, leading finally to neuronal dysfunction and death (reviewed by Okazawa, 2003). These results suggest that polyQ diseases represent a type of transcriptional disorder (Okazawa, 2003), supporting our misidentification hypothesis for at least some conformational disorders.

Generally, three computational/bioinformatics approaches were elaborated to estimate the abundance of IDPs under various pathological conditions. The first approach is based on the assembly of specific data sets of proteins associated with a given disease and the computational analysis of these data sets using a number of disorder predictors (Iakoucheva et al., 2002; Cheng et al., 2006; Uversky et al., 2006; Mohan et al., 2008; Uversky, 2008, 2009). In essence, this is an analysis of individual proteins extended to a set of independent proteins. A second approach utilized diseasome, a network of genetic diseases in which the related proteins are interlinked within one disease and between different diseases (Midic et al., 2008). A third approach is based on evaluation of the association between a particular protein function (including the disease-specific functional keywords) and the level of intrinsic disorder in a set of proteins known to carry out this function (Vucetic et al., 2007; Xie et al., 2007a, b). These three approaches are described briefly below and the results of their use are presented in a subsequent section.

The data set analysis approach was used for the first time in 2002, when it was found that 79% of cancer-associated and 66% of cell-signaling proteins contain predicted regions of disorder of 30 residues or longer (Iakoucheva et al., 2002). In contrast, only 13% of a set of proteins with well-defined ordered structures contained such long regions of predicted disorder. For this study, cancer-associated proteins were defined as those human proteins in Swiss-Prot containing the keyword *oncogene* (this included anti- and proto-oncogenes) or containing the word *tumor* in the description field. In experimental studies,

the presence of disorder has been observed directly in several cancer-associated proteins, including p53 (Lee et al., 2000), $p57^{kip2}$ (Adkins and Lumb, 2002), Bcl-X_L and Bcl-2 (Chang et al., 1997), c-Fos (Campbell et al., 2000), and most recently, a thyroid cancer–associated protein, TC-1 (Sunde et al., 2004). Following a similar analytical model, a data set of 487 proteins related to cardiovascular disease (CVD) was collected and analyzed (Cheng et al., 2006). On average, CVD-related proteins were found to be highly disordered. The percentage of proteins with 30 or more consecutive disordered residues was 61% for CVD-associated proteins. Many proteins were predicted to be wholly disordered, with 101 proteins from the CVD data set predicted to have a total of almost 200 specific disorder-based binding motifs (thus, about two binding sites per protein), α-MoRFs (Cheng et al., 2006). Finally, data set 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 (Uversky et al., 2008; Uversky, 2009).

The human diseasome systematically links the human disease phenome (which includes 1284 human genetic diseases, 867 of which had at least one link to other diseases, and 516 diseases formed a giant component) with the human disease genome (which contains 1777 disease genes, of which 1377 were shown to be connected to other disease genes, and 903 genes belonged to a giant cluster) (Goh et al., 2007). The abundance of intrinsic disorder in the human diseasome was evaluated using a set of computational tools such as PONDR VSL2, CDF analysis, CH plots, and α-MoRF prediction (Midic et al., 2008). These analyses uncovered an unfoldome associated with human genetic diseases and revealed that intrinsic disorder is common in proteins associated with many human genetic diseases. It was also shown that different disease classes vary in the IDP content of their associated proteins and that α-MoRFs are common in the diseasome, and their abundance correlates with the intrinsic disorder level. Finally, some disease classes were shown to have a significant fraction of genes affected by alternative splicing, and the alternatively spliced regions in the corresponding proteins are predicted to be highly disordered and in some diseases contain a significant number of MoRFs (Midic et al., 2008).

The studies on correlation of ID with various functional key words (Vucetic et al., 2007; Xie et al., 2007a, b) revealed that many diseases were strongly correlated with proteins predicted to be disordered. Contrary to this, no disease-associated proteins were found to be strongly correlated with an absence of disorder (Xie et al., 2007a). Among disease-related Swiss-Prot keywords strongly associated with ID were *oncoproteins, malaria, trypanosomiasis, human immunodeficiency virus* (HIV) and *acquired immunodeficiency syndrome* (AIDS), *deafness, obesity, cardiovascular disease, diabetes mellitus, albinism*, and *prion* (Xie et al., 2007a). In agreement with this bioinformatics

analysis, at least one illustrative, experimentally validated example of functional disorder or order was found for the vast majority of functional keywords related to diseases (Xie et al., 2007a).

Summarizing, 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" (D^2) concept was introduced to summarize work in this area (Uversky et al., 2008), and concepts of the disease-related unfoldome and unfoldomics were developed (Uversky et al., 2009).

1.3.5 How Does an Amino Acid Sequence Encode Intrinsic Disorder?

As has already been pointed out, the correct folding of a protein into its rigid biologically active conformation is determined by its amino acid sequence (Anfinsen et al., 1961). This gave rise to the reasonable assumption that the absence of rigid structure in natively unfolded proteins may also somehow be encoded in the specific features of their amino acid sequences. All the functional and structural peculiarities of IDPs and IDPRs are encoded in their amino acid sequences. In fact, it was recognized long ago that there are significant differences between the ordered proteins/domains and IDPs and IDPRs at the level of their amino acid sequences (Dunker et al., 1998, 2001; Uversky et al., 2000).

In early studies on finding the relationship between sequence and disorder, Dunker and co-authors developed several neuronal network predictors and applied them to large data sets (Dunker et al., 1998, 2000, 2001; Romero et al., 1998, 2001). The results of these analyses were more than impressive, as it was established that the disordered regions shared at least some common sequence features between many proteins and that more than 15,000 proteins in the Swiss-Prot database were identified as having long IDPRs (Romero et al., 1998).

Interestingly, the disordered proteins with the highest scores were shown to have low sequence complexity, assuming that the sequences of IDPs may essentially be degenerated. However, it was later established that distributions of the complexity values for ordered and disordered sequences overlapped, suggesting that low sequence complexity did not represent the only characteristic feature of intrinsically disordered proteins (Romero et al., 2001). Overall, the sequence of the intrinsically disordered proteins is characterized by amino acid compositional bias and the existence of highly predictable flexibility (Dunker et al., 1998; Garner et al., 1998). In fact, in comparison with ordered proteins, IDPs and IDPRs are characterized by noticeable biases in their amino acid compositions (Dunker et al., 1998, 2001; Uversky et al., 2000; Radivojac et al., 2007; Vacic et al., 2007b; Uversky and Dunker, 2010), containing less of

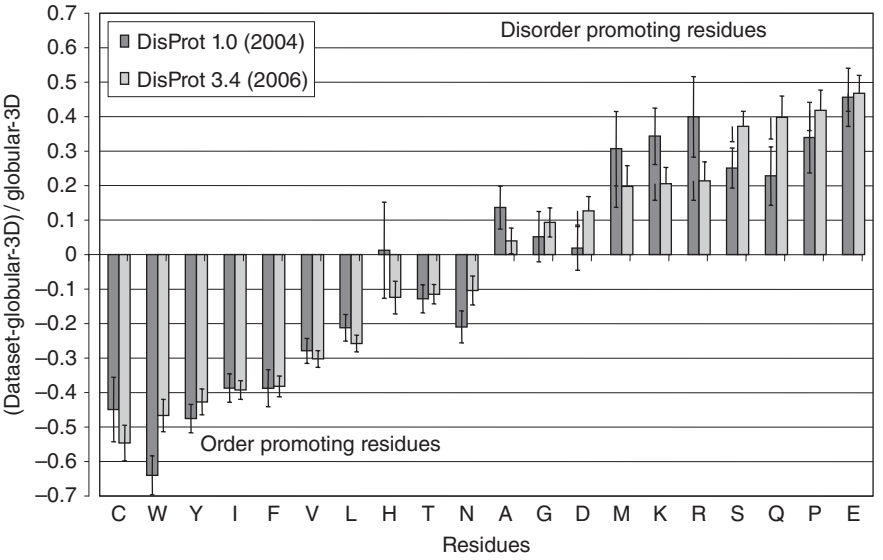

FIGURE 1.6 Peculiarities of amino acid composition of IDPs. An order/disorder composition profile is shown for comparisons of amino acid compositions of ordered protein with each of two databases of disordered protein. The ordinates are (% amino acid in disordered data set – % amino acid in ordered data set)/(% amino acid in ordered data set). Names of each database indicate how the disordered regions were identified. Negative values indicate that the disordered database has less than order; positive indicates more than order.

the "order-promoting" residues (cysteine, tryptophan, isoleucine, tyrosine, phenylalanine, leucine, histidine, valine, asparagines, and methionine, which are mostly hydrophobic residues commonly found within the hydrophobic cores of foldable proteins) and more of the "disorder-promoting" residues (lysine, glutamine, serine, glutamic acid, and proline, which are mostly polar and charged residues typically located at the surface of foldable proteins) (see Figure 1.6).

Also, in one of the early studies on these subjects, comparison of the overall hydrophacy and the net charge of ordered proteins and extended IDPs might be used to predict whether a given amino acid sequence encodes a native (folded) or an intrinsically disordered protein (Uversky et al., 2000). In fact, this survey shows that the natively unfolded proteins are specifically localized within a unique region of the charge–hydrophacy phase space, being completely isolated from the native globular proteins. Obviously, this allows estimation of the "boundary" mean hydrophacy value, below which a polypeptide chain with a given mean net charge will most probably be unfolded (Uversky et al., 2000).

1.3.6 Polymer Aspects of Nonfolding

It is interesting to apply the same formalism of polymer physics as that represented in the first part of the review to the two classes of extended IDPs (native coils and native pre-molten globules). This type of analysis is available because a considerable number of data have been accumulated on the hydrodynamic properties of these proteins. Obviously, the behavior of any polypeptide chain in a solution is determined by interaction of the polymer with the solvent. The fact that natively unfolded proteins, with their depleted hydrophobicity, are noncompact under physiological conditions indicates that "salted water" (typical "physiological" buffer contains 100 to 150 mM NaCl) does not represent for them a poor solvent. In other words, these conditions do not force polymer segments to interact specifically with each other and thus do not force them to be excluded effectively from the solvent. On the other hand, it has already been noted that even high concentrations of strong denaturants do not represent a good solvent for a polypeptide chain encoding for a typical globular protein, and a globular protein was assumed never to be a random coil.

In this view, the data on the hydrodynamic analysis of extended IDPs are extremely important. Such an analysis reveals that the molecular dimensions of extended IDPs follow the chain length as $R_S = 0.28\,M^{0.49}$ or $R_S = 0.6\,M^{0.40}$ for the native coils and native pre-molten globules, respectively (see Figure 1.3). Figure 1.3 clearly reflects the fact that native coils belong to the class of relatively extended unfolded conformations. Importantly, these coils show the highest K_h values and the smallest ε values between different unfolded conformations of a polypeptide chain [cf. equations (5), (7), and (8)]. This means that native coils under physiological conditions are in considerably worsened solvent conditions than those of globular proteins in urea or GdmHCl solutions (the lowest ε value), which gives rise to the increased probability of multiple body interactions (the highest K_h value). Finally, Figure 1.3 shows that native pre-molten globules follow exactly the same dependence as that of pre-molten globules of normal globular proteins. Thus, these proteins may exhibit the structural features of a squeezed polymer coil.

1.4 MISFOLDING

1.4.1 Molecular Mechanisms of Protein Misfolding

As has already been noted, the sequences of proteins have evolved in such a way that their unique native states can be found very efficiently even in the complex environment inside a living cell. However, under some conditions, proteins fail to fold properly or to remain folded correctly; this misfolding

can lead to the development of different pathological conditions. A number of human diseases, including the amyloidoses and several neurodegenerative disorders, such as Alzheimer's disease, Parkinson's disease, and transmissible spongiform encephalopathies, originate from the deposition of the stable, ordered, and filamentous protein aggregates commonly known as *amyloid fibrils*. In each of these pathological states, a specific protein or protein fragment changes from its natural soluble form into insoluble fibrils, which accumulate in a variety of organs and tissues (Kelly, 1998; Bellotti et al., 1999; Dobson and Karplus, 1999; Uversky et al., 1999b,c; Rochet and Lansbury, 2000). Approximately 20 different proteins are known to be involved in the amyloidoses (extracellular deposits), which are unrelated structurally or at the level of primary structure. In addition, a number of diseases also originate from the deposition of fibrillar proteins, but within cells (i.e., intracellular deposits).

Importantly, prior to fibrillation, amyloidogenic polypeptides may be rich in β-sheet, α-helix, β-helix, or contain both α-helices and β-sheets. They may be well folded or IDPs. Despite these differences, the fibrils from different pathologies display many common properties, including a core cross-β-sheet structure in which continuous β-sheets are formed with β-strands running perpendicular to the long axis of the fibrils (Sunde et al., 1997). Amyloid fibrils have been shown to form *in vitro* from disease-associated as well as disease-unrelated proteins and peptides. Furthermore, it is clear that the ability to form fibrils is a generic property of the polypeptide chain; that is, many proteins, perhaps all, are able to form amyloid fibrils under appropriate conditions (Dobson and Karplus, 1999). If so, this would dramatically extend the structural diversity of polypeptide chains able to fibrillate.

Since all 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. Such changes cannot happen in a native protein, due to its stable and rigid tertiary structure. Thus, protein destabilization favoring partial unfolding and culminating in the formation of a partially unfolded conformation is required. Presumably, such a partially unfolded conformation favors reciprocal and specific intermolecular interactions, including electrostatic attraction, hydrogen bonding, and hydrophobic contacts, which are necessary for oligomerization and fibrillation (Fink, 1998; Kelly, 1998; Bellotti et al., 1999; Dobson and Karplus, 1999; Lansbury, 1999; Uversky et al., 1999b,c; Rochet and Lansbury, 2000; Dobson, 2001a, b; Zerovnik, 2002).

Obviously, this model does take into account a class of IDPs, as they are devoid of rigid tertiary structure in their native state. The primary step in the fibrillation of these proteins represents partial folding, that is, stabilization of a partially folded conformation (Schweers et al., 1994; Teplow, 1998; Kayed et al.,

1999; Uversky et al., 2001a,b; Pavlov et al., 2002). Thus, the general hypothesis of fibrillogenesis states: Structural transformation of a polypeptide chain into a partially folded conformation represents an important prerequisite for successful protein fibrillation.

A question then arises as to the nature of amyloidogenic intermediate(s). It has already been mentioned that the conformational space of globular proteins involves four general conformations: the native state, unfolded states, and two classes of partially folded intermediates. The latter are compact, relatively well-folded intermediates with substantial native-like secondary structure but little tertiary structure, often referred to as molten globules; and significantly less compact, relatively unfolded intermediates, with substantially less secondary structure, often known as pre-molten globules. Potentially, either of these latter conformations may play a role as the crucial amyloidogenic species. Detailed structural analysis of the early fibrillation events in several proteins has demonstrated that the amyloidogenic conformation is only slightly folded and shares many structural properties with the pre-molten globule state.

1.4.2 Fibrillogenesis of Globular Proteins: Requirement for Partial Unfolding

Data have been reported indicating that the first critical step in protein fibrillogenesis is partial unfolding of the protein. Due to structural fluctuations (conformational breathing) the structure of a globular protein under physiological conditions represents a mixture of tightly folded and multiple partially unfolded conformations, with a great prevalence of the former. Most mutations associated with accelerated fibrillation and protein deposition diseases have been shown to destabilize the native structure, increasing the steady-state concentration of partially folded conformers (Colon et al., 1996; Kelly, 1996, 1998; Wetzel, 1997; Bellotti et al., 1999; Canet et al., 1999; Dobson and Karplus, 1999; Lashuel et al., 1999; Uversky et al., 1999b,c; Rochet and Lansbury, 2000; Heegaard et al., 2001; Saraiva, 2001). Conversely, it has been shown that the amyloidogenicity of a protein can be reduced significantly by stabilization of the native structure, for example, via specific binding of ligands (Miroy et al., 1996; Chiti et al., 2001; Nielsen et al., 2001).

The fibrillation-provoking destabilization of a rigid protein is achieved *in vitro* at low or high pH, high temperatures, and low to moderate concentrations of strong denaturants, organic solvents, and so on. This fact is well illustrated in recent studies on the fibrillogenesis of transthyretin (TTR), also known as prealbumin, which is a homotetramer composed of 127 amino acid subunits characterized by 2,2,2 molecular symmetry. TTR is found in human plasma (0.1 to 0.4 mg/mL) and cerebral spinal fluid (0.017 mg/mL), with the plasma

form being the amyloidogenic precursor. Wild-type TTR amyloidogenesis may cause senile systemic amyloidosis, characterized by deposition and pathology in the heart after age 60. Early-onset amyloid formation (as early as the second decade) by one of more than 80 single-site TTR variants provokes a number of diseases collectively termed *familial amyloid polyneuropathy* (Saraiva et al., 1984). TTR can be converted into amyloid *in vitro* by acid-mediated dissociation of the homotetramer into monomers. The pH required for disassembly also results in tertiary structural changes within the monomeric subunits, finally leading to the enhanced fibrillation (Lai et al., 1996). Recently, in a quest to understand the relationship between the tertiary structural changes and amyloidogenicity, a monomeric mutant with native-like structure and stability has been designed, which was nonamyloidogenic unless partially unfolded (Jiang et al., 2001).

Light chain, or AL, amyloidosis is a pathological condition arising from systemic extracellular deposition of monoclonal immunoglobulin light chain variable domains in the form of insoluble amyloid fibrils, especially in the kidneys (Buxbaum, 1992). Structural and fibrillation properties of one of the amyloidogenic light chain variable domains, SMA, have been analyzed under a variety of conditions (Khurana et al., 2001). The results of biophysical analysis revealed that a decrease in pH resulted in the accumulation of two partially folded intermediates. A relatively native-like intermediate, I_N, was observed between pH 4 and 6, and was characterized by little loss of secondary structure, combined with significant changes in tertiary structure and enhanced ANS binding. At a pH below 3, a relatively unfolded but compact intermediate, I_U, with decreased tertiary and secondary structure, was observed. The I_U intermediate readily forms amyloid fibrils, whereas I_N preferentially leads to amorphous aggregates (Khurana et al., 2001). Comparable data have recently been reported for another light chain variable domain, LEN (Souillac et al., 2002a, b).

α-Lactalbumin, or α-LA, is a small acidic protein with a single Ca^{2+}-binding site. It is very attractive for studies of partially folded conformations, since it adopts the classic molten globule conformation at acidic pH, moderate guanidinium–chloride concentrations, or elevated temperatures (apo form) (Permyakov and Berliner, 2000). α-LA is comprised of a large α-helix domain and a small β-sheet domain connected by a calcium-binding loop and four disulfide bridges (Permyakov and Berliner, 2000). α-LA forms amyloid fibrils at low pH. *S*-Carboxymethyl-α-lactalbumin, a disordered form of the protein with three of the disulfide bridges reduced, was even more susceptible to fibrillation. *S*-Carboxymethyl-α-lactalbumin exhibits the properties of a pre-molten globule, and its fibrillation is orders of magnitude faster than when starting with the molten globule conformation (Goers et al., 2002). Other partially folded conformations induced in α-LA at neutral pH, either by

removal of Ca^{2+} or by binding of Zn^{2+} to the Ca^{2+}-protein, did not fibrillate, although Zn^{2+}-loaded α-lactalbumin precipitated out of solution as amorphous aggregates. Based on these data it was concluded that the transformation from native state to a substantially unfolded conformation is required for a successful fibril formation, whereas less unfolded species may form amorphous aggregates (Goers et al., 2002).

Finally, fibrillation of bovine β-lactoglobulin represents another illustrative example of the importance of being unfolded (Hamada and Dobson, 2002). It has been shown that this protein will form fibrils in urea solutions; the process is denaturant concentration-dependent, showing the highest efficiency in the vicinity of 5 M urea, which corresponds roughly to the C_m value. Importantly, it was shown that molten globule-like intermediate was not accumulated during the urea-induced unfolding of β-lactoglobulin (Hamada and Dobson, 2002). Thus, it has been concluded that amyloid fibril formation by bovine β-lactoglobulin is promoted under conditions where significant accumulation of unfolded protein occurs, but is inhibited under conditions where higher denaturant concentrations destabilize intermolecular interactions (Hamada and Dobson, 2002). Generally speaking, it has been concluded that amyloid formation *in vitro* can be achieved by destabilizing the native state of the protein under conditions in which noncovalent interactions still remain favorable (Ramirez-Alvarado et al., 2000).

1.4.3 Fibrillogenesis of IDPs: Requirement for Partial Folding

We now consider details of the fibrillogenesis of intrinsically disordered proteins, which constitute a significant portion of known amyloidogenic proteins (Uversky, 2008, 2009). It seems to be reasonable to assume that such proteins are well suited for amyloidogenesis, as they lack significant secondary and tertiary structure as well as many specific intrachain interactions. In the absence of such conformational constraints, they would be expected to be substantially more conformationally flexible and thus able to polymerize more readily than tightly packed globular proteins. Substantial evidence suggests that the earliest stage of fibrillation of these proteins is their partial refolding.

α-*Synuclein* α-Synuclein is a small (14 kDa), soluble, intracellular, highly conserved protein that is abundant in various regions of the brain. This protein has been estimated to account for as much as 1% of the total protein in soluble cytosolic brain fractions. Structurally, purified α-synuclein is a typical extended IDP (Uversky et al., 2001b), which is unstructured under conditions of neutral pH but does not represent a random coil. It has some residual structure [at least a region with a preference for helical conformation (Eliezer et al., 2001)], leading to partial compaction (Uversky et al., 2001b; Uversky, 2003a).

Either a decrease in pH, or an increase in temperature, transformed α-synuclein into a partially folded conformation. This partially folded conformation resembles the pre-molten globule state, an intermediate, preceding the molten globule in the refolding of globular proteins (Uversky and Ptitsyn, 1994, 1996a; Uversky, 2003b).

Misfolding, dysfunction, and aggregation of α-synuclein are associated with several diseases, known collectively as *synucleinopathies*: a group of neurodegenerative disorders characterized by fibrillary aggregates of α-synuclein protein in the cytoplasm of selective populations of neurons and glia (Goedert, 1999; Spillantini and Goedert, 2000; Galvin et al., 2001; Trojanowski and Lee, 2003). Clinically, synucleinopathies are characterized by a chronic and progressive decline in motor, cognitive, behavioral, and autonomic functions, depending on the distribution of the lesions. Because of clinical overlap, differential diagnosis is sometimes very difficult (Marti et al., 2003). The neuropathological spectrum of synucleinopathies has been discussed intensively (Goedert, 1999, 2001; Duda et al., 2000; Spillantini and Goedert, 2000; Dickson, 2001; Galvin et al., 2001; Dev et al., 2003; Jellinger, 2003; Marti et al., 2003; Mitra et al., 2003; Trojanowski and Lee, 2003; Norris et al., 2004), and the potential mechanisms linking the α-synuclein aggregation with the development of several of these diseases are the major focus of numerous studies.

Several observations implicate α-synuclein in the pathogenesis of Parkinson's disease (PD). For example, a direct role for α-synuclein in the neurodegenerative processes in PD is demonstrated by genetic evidence and autosomal dominant early-onset PD is associated with three different missense mutations in the α-synuclein gene, corresponding to A30P, E46K, and A53T substitutions in α-synuclein (Polymeropoulos et al., 1997; Kruger et al., 1998; Zarranz et al., 2004) or with the hyperexpression of the wild-type α-synuclein protein, due to gene triplication (Farrer et al., 2004; Singleton et al., 2003, 2004). Antibodies to α-synuclein detect this protein in LBs and LNs, the hallmark lesions of PD. A substantial portion of fibrillar material in these specific inclusions was shown to be comprised of α-synuclein, and the insoluble α-synuclein filaments were recovered from purified LBs (Spillantini et al., 1997, 1998).

The fibrillogenesis of this protein is probably the most precisely studied and the best understood among other amyloidogenic members of the family of natively unfolded proteins. In particular, accumulated data strongly suggest that the formation of a partially folded intermediate (possessing the major characteristics of the pre-molten globule) represents the critical first step of α-synuclein fibrillogenesis. This partially folded intermediate can be stabilized by numerous factors, including high temperatures, low pH (Uversky et al., 2001b), the presence of low concentrations of organic solvents (Munishkina et al., 2003) and TMAO (Uversky et al., 2001c), the presence of

different metal ions (Uversky et al., 2001e), various salts (Munishkina et al., 2004), several common pesticides and herbicides (Uversky et al., 2001d, 2002a; Manning-Bog et al., 2002), heparin and other glycosoaminoglycans (Cohlberg et al., 2002), some polycations (Goers et al., 2003), or as a result of a spontaneous oligomerization both *in vitro* and *in vivo* (Uversky et al., 2001a). Furthermore, the addition of various alcohols was shown to increase the content of ordered secondary structure in α-synuclein (Munishkina et al., 2003). Interestingly, the structural transformations induced by high solvent concentrations were dependent on the type of alcohol, with simple alcohols inducing a β-sheet-enriched conformation and fluorinated alcohols promoting α-helix-rich species (Munishkina et al., 2003). Interestingly, both α-helical and β-structural species were shown to be initially monomeric but underwent association over longer times, and β-sheet-rich conformations were strongly prone to form amorphous aggregates (Munishkina et al., 2003). Oligomeric α-helical globular species potentially possessing rigid tertiary structure were induced in α-synuclein by high concentrations of TMAO (Uversky et al., 2001c).

Importantly, under conditions stabilizing the pre-molten globule-like conformation, α-synuclein was shown to undergo significantly enhanced fibrillation. In contrast, fibril formation was considerably slowed or inhibited under conditions favoring the formation of more folded conformations, or by stabilization of the fully unfolded form (e.g., by oxidation of its methionines) (Uversky et al., 2002b).

Amylin In addition to insulin, pancreatic islet β-cells produce a polypeptide called *amylin* or *islet amyloid polypeptide*, IAAP (Cooper et al., 1987). Amylin has several functions associated with the normal regulation of energy metabolism. Dysfunction of amylin due to mutation and/or amyloid fibril formation has been associated with the development of non-insulin-dependent diabetes mellitus (NIDDM), also known as type 2 diabetes (Higham et al., 2000; Jaikaran and Clark, 2001; Jaikaran et al., 2001). Type 2 diabetes is characterized by chronic insulin resistance and progressive decline in pancreatic β-cell function. One of the most common pathological features of type 2 diabetes is the deposition of amyloid fibrils in the islets of Langerhans of the pancreas (Schneider et al., 1980; Clark et al., 1988; Rocken et al., 1992; Kahn et al., 1999). Human IAPP or amylin, is the major protein component of these amyloid deposits (Cooper et al., 1987; Westermark et al., 1987). Amylin is an unstructured peptide hormone of 37 amino acid residues. The natively unfolded nature of this peptide was established using far-ultraviolet (UV) circular dichroism (CD) spectroscopy (Goldsbury et al., 2000; Jaikaran et al., 2001; Yoon and Welsh, 2005) and electron paramagnetic resonance spectroscopy (Jayasinghe and Langen, 2004). For example, eight spin-labeled derivatives of IAPP were analyzed using electron paramagnetic resonance spectroscopy.

In solution, all eight derivatives gave rise to electron paramagnetic resonance spectra with sharp lines indicative of rapid motion on a sub-nanosecond time scale, which are consistent with a rapidly tumbling and highly dynamic peptide (Jayasinghe and Langen, 2004). Human amylin and its 8–37 fragment were shown to form fibrils under physiological conditions. The process of polymerization is relatively fast (lag times were 100 and 50 min for full-length amylin and its 8–37 fragment, respectively) and results in the appearance of typical amyloid fibrils (Goldsbury et al., 2000). Interestingly, both peptides showed the formation of a partially folded (pre-molten globule-like) intermediate early in the fibrillation process. It takes about 90 min for full-length amylin to form such an intermediate, whereas this period was almost half as long for the truncated peptide, showing excellent agreement with the fibrillation lag times (Goldsbury et al., 2000).

Amyloid β-Protein (Aβ) Alzheimer's disease (AD) is the most prevalent age-dependent dementia. AD is characterized pathologically by the accumulation of extracellular amyloid deposits in the cerebral neuropil and vasculature and of intracellular neurofibrillary tangles. Amyloid deposits contain the amyloid β-protein (Aβ), which is a 40- to 42-residue peptide that is produced by endo-proteolytic cleavage of the amyloid β-protein precursor (APP). Many lines of evidence support the crucial role of Aβ in AD. Fibrillation of Aβ is associated with development of the cascade of neuropathogenetic events, ending with the appearance of cognitive and behavioral features typical of AD. Aβ appears to be unfolded at the beginning of the fibrillation under physiological conditions. NMR studies have shown that monomers of Aβ(1–40) or Aβ(1–42) possess no α-helical or β-sheet structure (Zagorski et al., 1999); they exist predominately as "random" extended chains. Partial refolding to the pre-molten globule-like conformation has been detected at the earliest stages of Aβ amyloidosis (Kirkitadze et al., 2001).

Tau-Protein Tau, a microtubule assembly protein isolated from brain microtubules, represents a family of isoforms which migrate as close bands of 55 to 62 kDa in SDS gel electrophoresis. Heterogeneity is explained in part by alternative mRNA splicing, leading to the appearance of one, two, three, or four repeats in the C-terminal region (Himmler, 1989; Himmler et al., 1989). Post-translational phosphorylation of tau is an additional source of microheterogeneity (Kenessey and Yen, 1993). *In vitro*, tau binds to microtubules, promotes microtubule assembly, and affects the dynamic instability of individual microtubules (Cleveland et al., 1977a, b; Drechsel et al., 1992; Brandt and Lee, 1993a, b). In situ, tau is highly enriched in the axons (Binder et al., 1985). In living cells and brain tissue, tau protein has been estimated as comprising 0.025 to 0.25% of total protein (Drubin et al., 1985;

Khatoon et al., 1992). On the basis of its *in vitro* activity and its distribution, it is believed that tau regulates the organization of neuronal microtubules. Interest in tau increased dramatically with the discovery of its aggregation in neuronal cells in the progress of Alzheimer's disease and various other neurodegenerative disorders, especially frontotemporal dementia (Delacourte and Buee, 1997; Crowther and Goedert, 2000). In these cases, specific tau-containing neurofibrillary tangles (paired helical filaments) are formed (Delacourte and Buee, 1997). Filaments isolated from end-stage AD are particularly well characterized and consist of all six full-length tau isoforms extensively phosphorylated and organized into twisted paired helical filaments (PHFs) and nontwisted straight filaments (SFs) (Lee et al., 1991). Hyperphosphorylation was shown to be a common characteristic of pathological tau (Vulliet et al., 1992). Hyperphosphorylated tau isolated from patients with AD was shown to be unable to bind to microtubules and promote microtubule assembly. However, both of these activities were restored after enzymatic dephosphorylation of tau protein (Lu and Wood, 1993; Alonso et al., 1994, 1996; Iqbal et al., 1994).

During brain development, tau is phosphorylated at many residues, including sites phosphorylated with GSK-3β, cdk 5, and MAPK (Watanabe et al., 1993). *In vitro*, tau can be phosphorylated on multiple sites by several kinases (for a review, see Billingsley and Kincaid, 1997). Most of the *in vitro* phosphorylation sites of tau are located within the microtubule interacting region (repeat domain) and sequences flanking the repeat domain. Many of these sites are also phosphorylated in PHF-tau (Morishima-Kawashima et al., 1995a,b). In fact, 10 major phosphorylation sites have been identified in tau isolated from PHFs from patients with AD (Morishima-Kawashima et al., 1995a,b). All of these sites are located in regions flanking tau's repeat domain and constitute recognition sites for several AD diagnostic antibodies, which may point to an important role for these phosphorylation sites for AD pathogenesis. Hyperphosphorylation was shown to be accompanied by transformation from the unfolded state of tau into a partially folded conformation (Hagestedt et al., 1989; Uversky et al., 1998b), accelerating dramatically the self-assembly of this protein into paired helical filaments *in vitro* (Alonso et al., 1996). To analyze the potential role of tau hyperphosphorylation in tauopathies, mutated tau proteins have been produced in which all 10 serine/threonine residues known to be highly phosphorylated in PHF-tau were substituted for negatively charged residues, thus producing a model for a defined and permanent hyperphosphorylation-like state of tau protein (Eidenmuller et al., 2000). It has been demonstrated that, like hyperphosphorylation, glutamate substitutions induce compact structure elements and SDS-resistant conformational domains in tau protein, as well as lead to the dramatic acceleration of its fibrillation (Eidenmuller et al., 2000).

Prior to aggregation, tau protein was shown to be in a mostly random coil structure. This conclusion followed from the conformational analysis of this protein by circular dichroism, Fourier transform infrared, x-ray scattering, and biochemical assays (von Bergen et al., 2006). Analysis of the primary structure reveals a very low content of hydrophobic amino acids and a high content of charged residues, which was sufficient to explain the lack of folding (von Bergen et al., 2006). Analysis of the hydrodynamic radii confirms a mostly random coil structure of various tau isoforms and tau domains. However, the protein was further unfolded by high concentrations of strong denaturant GdmCl, indicating the presence of some residual structure in this protein. This conclusion was supported by a FRET-based approach where the distances between different domains of tau were determined. The combined data show that tau is mostly disordered and flexible but tends to assume a hairpin-like overall fold which may be important in the transition to a pathological aggregate (von Bergen et al., 2006).

Intriguingly, purified recombinant tau isoforms do not detectably aggregate over days of incubation under physiological conditions. However, aggregation and fibrillization of tau protein can be greatly accelerated under nearly-physiological conditions *in vitro* by the addition of anionic surfactants (Chirita et al., 2003). Based on the detailed analysis of tau fibrillation in the presence of anionic inducers using a set of spectroscopic techniques (circular dichroism spectroscopy and reactivity with thioflavin S and 8-anilino-1-naphthalene-sulfonic acid fluorescent probes), it has been established that the inducer stabilized a monomeric partially folded species with the structural characteristics of a pre-molten globule state (Chirita et al., 2005). The stabilization of this intermediate was sufficient to trigger the fibrilliation of full-length tau protein (Chirita et al., 2005).

Prothymosin α Prothymosin α is a very acidic protein, containing about 50% aspartic and glutamic acid, no aromatic or cysteine residues, and very few large hydrophobic aliphatic amino acids (Gast et al., 1995). Because of these features, prothymosin α adopts a random coil-like conformation with no regular secondary structure at neutral pH (Gast et al., 1995; Uversky et al., 1999a). However, at acidic pH, prothymosin α folds into a partially folded pre-molten globule-like conformation (Uversky et al., 1999a). Interestingly, it has recently been shown that at low pH [below pH 3, i.e., under conditions favoring the formation of the partially folded conformation (Uversky et al., 1999a)], prothymosin α is capable of relatively fast formation (lag time of about 100 min) of regular elongated fibrils with a flat ribbon-like structure 4_5 nm in height and 12 to 13 nm in width (Pavlov et al., 2002).

1.4.4 Conformational Prerequisites for Amyloidogenesis

Based on a comparison of far-UV CD spectra measured for "pure" amyloidogenic conformations of 11 proteins (SH3-domain: cytocrome c_{552}, monellin, methionine aminopeptidase, SMA, α-lactalbumin, phosphoglycerate kinase, PGK, amylin, prothymosin α, Aβ, and α-synuclein) with those retrieved for the four basic protein conformations—native, molten globule, pre-molten globule, and unfolded states—it has been concluded that all amyloidogenic conformations can be grouped with the set of data for the pre-molten globule proteins (Uversky, 2003a; Uversky and Fink, 2004). This is an extremely important observation which demonstrates the importance of this conformation for protein fibrillogenesis.

The central model of protein misfolding is illustrated by Figure 1.7, which shows that the process of fibrillation (which is used as an example of a more general phenomenon—aggregation) can be divided into three major steps: (1) structural transformation of a native soluble protein (rigid globular or flexible IDP) into the "sticky" amyloidogenic precursor intermediate, (2) nucleation, and (3) fibril elongation/growth. Thus, transformation into the partially folded conformation is a critical early stage of fibrillogenesis and precedes the appearance of any aggregated material (Uversky, 2003a).

1.4.5 Multiple Pathways of Protein Misfolding

Although the data presented above were mostly devoted to a consideration of protein fibrillation, the process of amyloid fibril formation does not represent the only misfolding route. In fact, contrary to the process of productive protein folding leading to the appearance of a rigid conformation with the specific function, the end products of misfolding may have a different appearance. 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. Potentially, the reason for such a morphological difference is connected with the diversity of the partially folded intermediates favoring protein self-association. In fact, multiple environmental factors, such as point mutations, a decrease in pH, an increase in temperature, or the presence of small organic molecules or metal ions or 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 partially folded intermediates, the formation of morphologically different aggregates is expected. This idea is illustrated by Figure 1.6, which represents a model of α-synuclein misfolding. It has been shown that

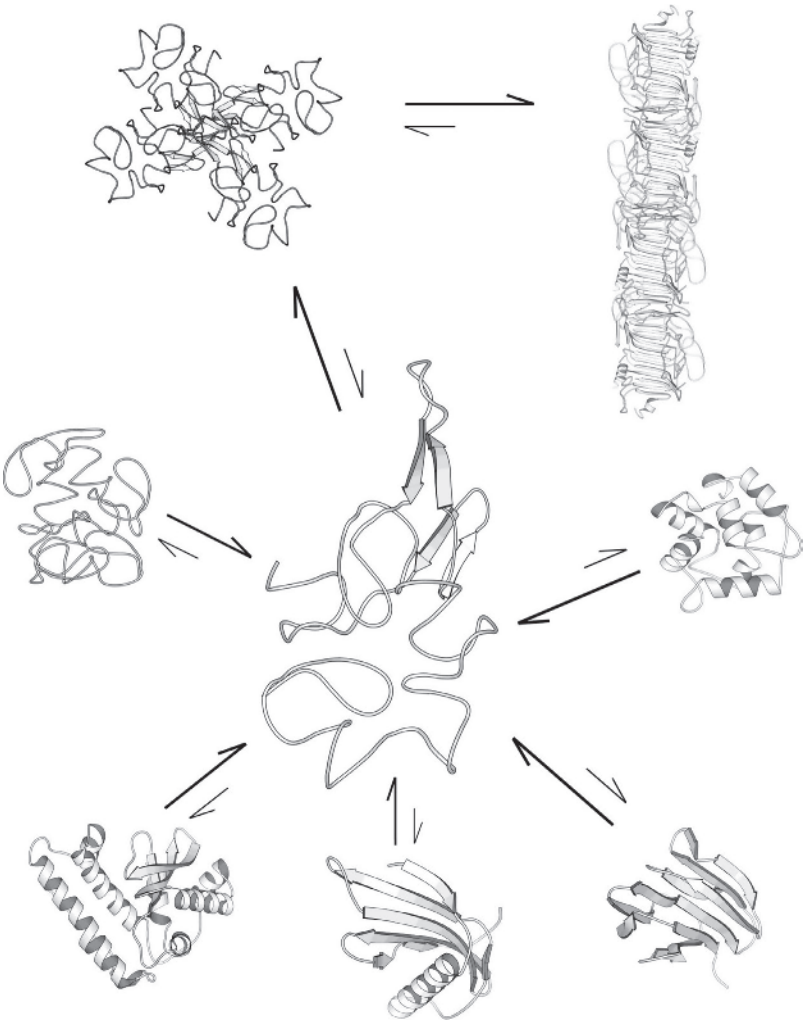

FIGURE 1.7 General model for protein misfolding (fibril formation is considered as an example). Three general stages of the process include the structural transformation of a native soluble protein (rigid or natively unfolded) into the "sticky" amyloidogenic precursor intermediate, nucleation, and fibril elongation/growth. Additional conformational changes may occur between the aggregation-competent intermediate and the fibrils. Structures at the top represent natively folded proteins with different structures. A common amyloidogenic intermediate for different rigid and natively unfolded proteins is shown for convenience only. Experimental data support an idea that amyloidogenic intermediates probably fall in the class of pre-molten globules. However, this still leaves a lot of room for structural diversity. The oligomeric intermediate (representing the nucleus or a soluble aggregate) is shown as a tetramer for convenience only, and could be much larger.

aggregation of this protein depends dramatically on the experimental conditions and might lead to the appearance of one of the three misfolded forms discussed above (soluble oligomers, amorphous aggregates, or amyloid fibrils) or mixture thereof.

1.4.6 Polymer Aspects of Protein Misfolding

The behavior of a given polymer in a given solution is determined by the peculiarities of polymer segment–solvent 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 the stability of a native protein molecule. Basically, point amino acid substitutions, and 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 the relative quality of solvent which appear either due to specific changes in protein amino acid composition or because of solvent composition modifications.

REFERENCES

Adkins JN, Lumb KJ (2002) Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2. Proteins 46:1–7.

Alexandrescu AT, Abeygunawardana C, Shortle D (1994) Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: a heteronuclear NMR study. Biochemistry 33:1063–1072.

Alonso AC, Zaidi T, Grundke-Iqbal I, Iqbal K (1994) Role of abnormally phosphorylated tau in the breakdown of microtubules in Alzheimer disease. Proc Natl Acad Sci U S A 91:5562–5566.

Alonso AC, Grundke-Iqbal I, Iqbal K (1996) Alzheimer's disease hyperphosphorylated tau sequesters normal tau into tangles of filaments and disassembles microtubules. Nat Med 2:783–787.

Altschuler GM, Willison KR (2008) Development of free-energy-based models for chaperonin containing TCP-1 mediated folding of actin. J R Soc Interface 5:1391–1408.

Amir D, Krausz S, Haas E (1992) Detection of local structures in reduced unfolded bovine pancreatic trypsin inhibitor. Proteins 13:162–173.

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230.

Anfinsen CB, Haber E, Sela M, White FH, Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A47:1309–1314.

Arcus VL, Vuilleumier S, Freund SM, Bycroft M, Fersht AR (1995) A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. J Mol Biol 254:305–321.

Baldwin RL (1986) Seeding protein folding. Trends Biochem Sci 11:6–9.

Baldwin RL, Zimm BH (2000) Are denatured proteins ever random coils? Proc Natl Acad Sci U S A 97:12391–12392.

Bellotti V, Mangione P, Stoppini M (1999) Biological activity and pathological implications of misfolded proteins. Cell Mol Life Sci 55:977–991.

Ben-Zvi AP, Goloubinoff P (2001) Review: mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. J Struct Biol 135:84–93.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242.

Billingsley ML, Kincaid RL (1997) Regulated phosphorylation and dephosphorylation of tau protein: effects on microtubule interaction, intracellular trafficking and neurodegeneration. Biochem J 323 (Pt 3):577–591.

Binder LI, Frankfurter A, Rebhun LI (1985) The distribution of tau in the mammalian central nervous system. J Cell Biol 101:1371–1378.

Brandt R, Lee G (1993a) The balance between tau protein's microtubule growth and nucleation activities: implications for the formation of axonal microtubules. J Neurochem 61:997–1005.

Brandt R, Lee G (1993b) Functional organization of microtubule-associated protein tau: identification of regions which affect microtubule growth, nucleation, and bundle formation in vitro. J Biol Chem 268:3414–3419.

Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 21:167–195.

Bushmarina NA, Kuznetsova IM, Biktashev AG, Turoverov KK, Uversky VN (2001) Partially folded conformations in the folding pathway of bovine carbonic anhydrase II: a fluorescence spectroscopic analysis. ChemBioChem 2:813–821.

Buxbaum J (1992) Mechanisms of disease: monoclonal immunoglobulin deposition—amyloidosis, light chain deposition disease, and light and heavy chain deposition disease. Hematol Oncol Clin North Am 6:323–346.

Campbell KM, Terrell AR, Laybourn PJ, Lumb KJ (2000) Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos. Biochemistry 39:2708–2713.

Canet D, Sunde M, Last AM, Miranker A, Spencer A, Robinson CV, Dobson CM (1999) Mechanistic studies of the folding of human lysozyme and the origin of amyloidogenic behavior in its disease-related variants. Biochemistry 38:6419–6427.

Chang BS, Minn AJ, Muchmore SW, Fesik SW, Thompson CB (1997) Identification of a novel regulatory domain in Bcl-X(L) and Bcl-2. EMBO J 16:968–977.

Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN (2006) Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry 45:10448–10460.

Chi SW, Kim DH, Lee SH, Chang I, Han KH (2007) Pre-structured motifs in the natively unstructured preS1 surface antigen of hepatitis B virus. Protein Sci 16:2108–2117.

Chirita CN, Necula M, Kuret J (2003) Anionic micelles and vesicles induce tau fibrillization in vitro. J Biol Chem 278:25644–25650.

Chirita CN, Congdon EE, Yin H, Kuret J (2005) Triggers of full-length tau aggregation: a role for partially folded intermediates. Biochemistry 44:5862–5872.

Chiti F, Taddei N, Stefani M, Dobson CM, Ramponi G (2001) Reduction of the amyloidogenicity of a protein by specific binding of ligands to the native conformation. Protein Sci 10:879–886.

Clark A, Wells CA, Buley ID, Cruickshank JK, Vanhegan RI, Matthews DR, Cooper GJ, Holman RR, Turner RC (1988) Islet amyloid, increased A-cells, reduced B-cells and exocrine fibrosis: quantitative changes in the pancreas in type 2 diabetes. Diabetes Res 9:151–159.

Clayton DF, George JM (1998) The synucleins: a family of proteins involved in synaptic function, plasticity, neurodegeneration and disease. Trends Neurosci 21:249–254.

Cleveland DW, Hwo SY, Kirschner MW (1977a) Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. J Mol Biol 116:227–247.

Cleveland DW, Hwo SY, Kirschner MW (1977b) Purification of tau, a microtubule-associated protein that induces assembly of microtubules from purified tubulin. J Mol Biol 116:207–225.

Cohlberg JA, Li J, Uversky VN, Fink AL (2002) Heparin and other glycosaminoglycans stimulate the formation of amyloid fibrils from alpha-synuclein in vitro. Biochemistry 41:1502–1511.

Colon W, Lai Z, McCutchen SL, Miroy GJ, Strang C, Kelly JW (1996) FAP mutations destabilize transthyretin facilitating conformational changes required for amyloid formation. Ciba Found Symp 199:228–238; discussion 39–42.

Cooper GJ, Willis AC, Clark A, Turner RC, Sim RB, Reid KB (1987) Purification and characterization of a peptide from amyloid-rich pancreases of type 2 diabetic patients. Proc Natl Acad Sci U S A 84:8628–8632.

Crowther RA, Goedert M (2000) Abnormal tau-containing filaments in neurodegenerative diseases. J Struct Biol 130:271–279.

Cummings CJ, Zoghbi HY (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. Hum Mol Genet 9:909–916.

Delacourte A, Buee L (1997) Normal and pathological tau proteins as factors for microtubule assembly. Int Rev Cytol 171:167–224.

Dev KK, Hofele K, Barbieri S, Buchman VL, van der Putten H (2003) Part II: Alpha-synuclein and its molecular pathophysiological role in neurodegenerative disease. Neuropharmacology 45:14–44.

Dickson DW (2001) Alpha-synuclein and the Lewy body disorders. Curr Opin Neurol 14:423–432.

Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. Nat Struct Biol 4:10–19.

Dill KA, Shortle D (1991) Denatured states of proteins. Annu Rev Biochem 60:795–825.

Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M (2000) Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci 25:331–339.

Dobson CM (1999) Protein misfolding, evolution and disease. Trends Biochem Sci 24:329–332.

Dobson CM (2001a) Protein folding and its links with human disease. Biochem Soc Symp: 1–26.

Dobson CM (2001b) The structural basis of protein folding and its links with human disease. Philos Trans R Soc Lond B 356:133–145.

Dobson CM, Karplus M (1999) The fundamentals of protein folding: bringing together theory and experiment. Curr Opin Struct Biol 9:92–101.

Dobson CM, Sali A, Karplus M (1998) Protein folding: a perspective from theory and experiment. Angew Chem Int Ed 37:868–893.

Dolgikh DA, Gilmanshin RI, Brazhnikov EV, Bychkova VE, Semisotnov GV, Venyaminov S, Ptitsyn OB (1981) Alpha-Lactalbumin: compact state with fluctuating tertiary structure? FEBS Lett 136:311–315.

Dolgikh DA, Abaturov LV, Bolotina IA, Brazhnikov EV, Bychkova VE, Gilmanshin RI, Lebedev Yu O, Semisotnov GV, Tiktopulo EI, Ptitsyn OB, et al. (1985) Compact state of a protein molecule with pronounced small-scale mobility: bovine alpha-lactalbumin. Eur Biophys J 13:109–121.

Dosztanyi Z, Meszaros B, Simon I (2009) ANCHOR: Web server for predicting protein binding regions in disordered proteins. Bioinformatics 25:2745–2746.

Drechsel DN, Hyman AA, Cobb MH, Kirschner MW (1992) Modulation of the dynamic instability of tubulin assembly by the microtubule-associated protein tau. Mol Biol Cell 3:1141–1154.

Drubin DG, Feinstein SC, Shooter EM, Kirschner MW (1985) Nerve growth factor–induced neurite outgrowth in PC12 cells involves the coordinate

induction of microtubule assembly and assembly-promoting factors. J Cell Biol 101:1799–1807.

Duda JE, Lee VM, Trojanowski JQ (2000) Neuropathology of synuclein aggregates. J Neurosci Res 61:121–127.

Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput: 473–484.

Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ (2000) Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform 11:161–171.

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59.

Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002a) Intrinsic disorder and protein function. Biochemistry 41:6573–6582.

Dunker AK, Brown CJ, Obradovic Z (2002b) Identification and functions of usefully disordered proteins. Adv Protein Chem 62:25–49.

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets: the roles of intrinsic disorder in protein interaction networks. FEBS J 272:5129–5148.

Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 12:54–60.

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6:197–208.

Eidenmuller J, Fath T, Hellwig A, Reed J, Sontag E, Brandt R (2000) Structural and functional implications of tau hyperphosphorylation: information from phosphorylation-mimicking mutated tau proteins. Biochemistry 39:13166–13175.

Eliezer D, Kutluay E, Bussell R, Jr., Browne G (2001) Conformational properties of alpha-synuclein in its free and lipid-associated states. J Mol Biol 307: 1061–1073.

Farrer M, Kachergus J, Forno L, Lincoln S, Wang DS, Hulihan M, Maraganore D, Gwinn-Hardy K, Wszolek Z, Dickson D, Langston JW (2004) Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. Ann Neurol 55:174–179.

Fersht AR, Daggett V (2002) Protein folding and unfolding at atomic resolution. Cell 108:573–582.

Fink AL (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des 3:R9–R23.

Finkelstein AV, Ptitsyn OB (2002) Protein Physics: A Course of Lectures. San Diego, CA: Academic Press.

Fischbeck KH (2001) Polyglutamine expansion neurodegenerative disease. Brain Res Bull 56:161–163.

Fischer E (1894) Einfluss der Configuration auf die Wirkung der Enzyme. Ber Dt Chem Ges 27:2993.

Fitzkee NC, Rose GD (2004) Reassessing random-coil statistics in unfolded proteins. Proc Natl Acad Sci U S A 101:12497–12502.

Fuxreiter M, Simon I, Friedrich P, Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol 338:1015–1026.

Galvin JE, Lee VM, Trojanowski JQ (2001) Synucleinopathies: clinical and pathological implications. Arch Neurol 58:186–190.

Garner E, Cannon P, Romero P, Obradovic Z, Dunker AK (1998) Predicting disordered regions from amino acid sequence: common themes despite differing structural characterization. Genome Inform Ser Workshop Genome Inform 9:201–213.

Gast K, Damaschun H, Eckert K, Schulze-Forster K, Maurer HR, Muller-Frohne M, Zirwer D, Czarnecki J, Damaschun G (1995) Prothymosin alpha: a biologically active protein with random coil conformation. Biochemistry 34:13211–13218.

Georlette D, Blaise V, Dohmen C, Bouillenne F, Damien B, Depiereux E, Gerday C, Uversky VN, Feller G (2003a) Cofactor binding modulates the conformational stabilities and unfolding patterns of NAD(+)-dependent DNA ligases from *Escherichia coli* and *Thermus scotoductus*. J Biol Chem 278:49945–49953.

Georlette D, Damien B, Blaise V, Depiereux E, Uversky VN, Gerday C, Feller G (2003b) Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. J Biol Chem 278: 37015–37023.

Gillespie JR, Shortle D (1997a) Characterization of long-range structure in the denatured state of staphylococcal nuclease: II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. J Mol Biol 268: 170–184.

Gillespie JR, Shortle D (1997b) Characterization of long-range structure in the denatured state of staphylococcal nuclease: I. Paramagnetic relaxation enhancement by nitroxide spin labels. J Mol Biol 268:158–169.

Glenner GG, Wong CW (1984) Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. Biochem Biophys Res Commun 122:1131–1135.

Goedert M (1999) Filamentous nerve cell inclusions in neurodegenerative diseases: tauopathies and alpha-synucleinopathies. Philos Trans R Soc Lond B 354: 1101–1118.

Goedert M (2001) Parkinson's disease and other alpha-synucleinopathies. Clin Chem Lab Med 39:308–312.

Goers J, Permyakov SE, Permyakov EA, Uversky VN, Fink AL (2002) Conformational prerequisites for alpha-lactalbumin fibrillation. Biochemistry 41:12546–12551.

Goers J, Uversky VN, Fink AL (2003) Polycation-induced oligomerization and accelerated fibrillation of human alpha-synuclein in vitro. Protein Sci 12: 702–707.

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL (2007) The human disease network. Proc Natl Acad Sci U S A 104:8685–8690.

Goldsbury C, Goldie K, Pellaud J, Seelig J, Frey P, Muller SA, Kistler J, Cooper GJ, Aebi U (2000) Amyloid fibril formation from full-length and fragments of amylin. J Struct Biol 130:352–362.

Gottfried DS, Haas E (1992) Nonlocal interactions stabilize compact folding intermediates in reduced unfolded bovine pancreatic trypsin inhibitor. Biochemistry 31:12353–12362.

Grossberg AY, Khohlov AR (1989) Statistical Physics of Macromolecules. Moscow: Nauka.

Gusella JF, MacDonald ME (2000) Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. Nat Rev Neurosci 1:109–115.

Hagestedt T, Lichtenberg B, Wille H, Mandelkow EM, Mandelkow E (1989) Tau protein becomes long and stiff upon phosphorylation: correlation between paracrystalline structure and degree of phosphorylation. J Cell Biol 109:1643–1651.

Hamada D, Dobson CM (2002) A kinetic study of beta-lactoglobulin amyloid fibril formation promoted by urea. Protein Sci 11:2417–2426.

Hammarstrom P, Carlsson U (2000) Is the unfolded state the Rosetta Stone of the protein folding problem? Biochem Biophys Res Commun 276:393–398.

Hazy E, Tompa P (2009) Limitations of induced folding in molecular recognition by intrinsically disordered proteins. ChemPhysChem 10:1415–1419.

Heegaard NH, Sen JW, Kaarsholm NC, Nissen MH (2001) Conformational intermediate of the amyloidogenic protein beta 2-microglobulin at neutral pH. J Biol Chem 276:32657–32662.

Higham CE, Jaikaran ET, Fraser PE, Gross M, Clark A (2000) Preparation of synthetic human islet amyloid polypeptide (IAPP) in a stable conformation to enable study of conversion to amyloid-like fibrils. FEBS Lett 470:55–60.

Himmler A (1989) Structure of the bovine tau gene: alternatively spliced transcripts generate a protein family. Mol Cell Biol 9:1389–1396.

Himmler A, Drechsel D, Kirschner MW, Martin DW, Jr. (1989) Tau consists of a set of proteins with repeated C-terminal microtubule-binding domains and variable N-terminal domains. Mol Cell Biol 9:1381–1388.

Hsu WL, Oldfield C, Meng J, Huang F, Xue B, Uversky VN, Romero P, Dunker AK (2012) Intrinsic protein disorder and protein–protein interactions. Pac Symp Biocomput:116–127.

Huang JR, Grzesiek S (2010) Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin. J Am Chem Soc 132:694–705.

Huang Y, Liu Z (2009) Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. J Mol Biol 393:1143–1159.

Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323: 573–584.

Iqbal K, Zaidi T, Bancher C, Grundke-Iqbal I (1994) Alzheimer paired helical filaments: restoration of the biological activity by dephosphorylation. FEBS Lett 349:104–108.

Jahn TR, Radford SE (2005) The yin and yang of protein folding. FEBS J 272: 5962–5970.

Jaikaran ET, Clark A (2001) Islet amyloid and type 2 diabetes: from molecular misfolding to islet pathophysiology. Biochim Biophys Acta 1537:179–203.

Jaikaran ET, Higham CE, Serpell LC, Zurdo J, Gross M, Clark A, Fraser PE (2001) Identification of a novel human islet amyloid polypeptide beta-sheet domain and factors influencing fibrillogenesis. J Mol Biol 308:515–525.

James E, Wu PG, Stites W, Brand L (1992) Compact denatured state of a staphylococcal nuclease mutant by guanidinium as determined by resonance energy transfer. Biochemistry 31:10217–10225.

Jayasinghe SA, Langen R (2004) Identifying structural features of fibrillar islet amyloid polypeptide using site-directed spin labeling. J Biol Chem 279: 48420–48425.

Jellinger KA (2003) Neuropathological spectrum of synucleinopathies. Mov Disord 18 Suppl 6:S2–S12.

Jiang X, Smith CS, Petrassi HM, Hammarstrom P, White JT, Sacchettini JC, Kelly JW (2001) An engineered transthyretin monomer that is nonamyloidogenic, unless it is partially denatured. Biochemistry 40:11442–11452.

Kahn SE, Andrikopoulos S, Verchere CB (1999) Islet amyloid: a long-recognized but underappreciated pathological feature of type 2 diabetes. Diabetes 48:241–253.

Karnoup AS, Uversky VN (1997) Sequential compaction of a random copolymer of hydrophilic and hydrophobic amino acid residues. Macromolecules 30: 7427–434.

Kayed R, Bernhagen J, Greenfield N, Sweimeh K, Brunner H, Voelter W, Kapurniotu A (1999) Conformational transitions of islet amyloid polypeptide (IAPP) in amyloid formation in vitro. J Mol Biol 287:781–796.

Kazmirski SL, Wong KB, Freund SM, Tan YJ, Fersht AR, Daggett V (2001) Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. Proc Natl Acad Sci U S A 98: 4349–4354.

Kelly JW (1996) Alternative conformations of amyloidogenic proteins govern their behavior. Curr Opin Struct Biol 6:11–17.

Kelly JW (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Curr Opin Struct Biol 8:101–106.

Kenessey A, Yen SH (1993) The extent of phosphorylation of fetal tau is comparable to that of PHF-tau from Alzheimer paired helical filaments. Brain Res 629: 40–46.

Khatoon S, Grundke-Iqbal I, Iqbal K, (1992) Brain levels of microtubule-associated protein tau are elevated in Alzheimer's disease: a radioimmuno-slot-blot assay for nanograms of the protein. J Neurochem 59:750–753.

Khurana R, Gillespie JR, Talapatra A, Minert LJ, Ionescu-Zanetti C, Millett I, Fink AL (2001) Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates. Biochemistry 40:3525–3535.

Kim PM, Sboner A, Xia Y, Gerstein M (2008) The role of disorder in interaction networks: a structural analysis. Mol Syst Biol 4:179.

Kirkitadze MD, Condron MM, Teplow DB (2001) Identification and characterization of key kinetic intermediates in amyloid beta-protein fibrillogenesis. J Mol Biol 312:1103–1119.

Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. Science 295:1719–1722.

Koepf EK, Petrassi HM, Sudol M, Kelly JW (1999) WW: An isolated three-stranded antiparallel beta-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. Protein Sci 8:841–853.

Kogo H, Takeuchi K, Inoue H, Kihara H, Kojima M, Takahashi K (2009) Urea-dependent unfolding of HIV-1 protease studied by circular dichroism and small-angle x-ray scattering. Biochim Biophys Acta 1794:70–74.

Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci U S A 101:12491–12496.

Kruger R, Kuhn W, Muller T, Woitalla D, Graeber M, Kosel S, Przuntek H, Epplen JT, Schols L, Riess O (1998) Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat Genet 18:106–108.

Kuwajima K (1996) The molten globule state of alpha-lactalbumin. FASEB J 10: 102–109.

Kuznetsova IM, Biktashev AG, Khaitlina SY, Vassilenko KS, Turoverov KK, Uversky VN (1999) Effect of self-association on the structural organization of partially folded proteins: inactivated actin. Biophys J 77:2788–2800.

Kuznetsova IM, Stepanenko OV, Turoverov KK, Zhu L, Zhou JM, Fink AL, Uversky VN (2002) Unraveling multistate unfolding of rabbit muscle creatine kinase. Biochim Biophys Acta 1596:138–155.

Kuznetsova IM, Forge V, Turoverov KK (2005) [Structural dynamics, stability, and folding of proteins.] Tsitologiia 47:943–952.

Lai Z, Colon W, Kelly JW (1996) The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. Biochemistry 35:6470–6482.

Lansbury PT, Jr. (1999) Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease. Proc Natl Acad Sci U S A 96:3342–3344.

Lashuel HA, Wurth C, Woo L, Kelly JW (1999) The most pathogenic transthyretin variant, L55P, forms amyloid fibrils under acidic conditions and protofilaments under physiological conditions. Biochemistry 38:13560–13573.

Lau KF, Dill KA (1990) Theory for protein mutability and biogenesis. Proc Natl Acad Sci U S A 87:638–642.

Lee BC, Kumauchi M, Hoff WD (2010) Modulating native-like residual structure in the fully denatured state of photoactive yellow protein affects its refolding. J Biol Chem 285:12579–12586.

Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, Chang J, Sung YC, Choi KY, Han KH (2000) Local structural elements in the mostly unstructured transcriptional activation domain of human p53. J Biol Chem 275:29426–29432.

Lee VM, Balin BJ, Otvos L, Jr., Trojanowski JQ (1991) A68: a major subunit of paired helical filaments and derivatized forms of normal tau. Science 251:675–678.

Levinthal C (1968) Are there pathways for protein folding? Chim Phys 65:44–45.

Liu J, Faeder JR, Camacho CJ (2009) Toward a quantitative theory of intrinsically disordered proteins and their function. Proc Natl Acad Sci U S A 106:19819–19823.

Lu Q, Wood JG (1993) Functional studies of Alzheimer's disease tau protein. J Neurosci 13:508–515.

Manning-Bog AB, McCormack AL, Li J, Uversky VN, Fink AL, Di Monte DA (2002) The herbicide paraquat causes up-regulation and aggregation of alpha-synuclein in mice: paraquat and alpha-synuclein. J Biol Chem 277:1641–1644.

Martensson LG, Jonsson BH, Freskgard PO, Kihlgren A, Svensson M, Carlsson U (1993) Characterization of folding intermediates of human carbonic anhydrase: II. Probing substructure by chemical labeling of SH groups introduced by site-directed mutagenesis. Biochemistry 32:224–231.

Marti MJ, Tolosa E, Campdelacreu J (2003) Clinical overview of the synucleinopathies. Mov Disord 18 Suppl 6:S21–S17.

Masters CL, Multhaup G, Simms G, Pottgiesser J, Martins RN, Beyreuther K (1985) Neuronal origin of a cerebral amyloid: neurofibrillary tangles of Alzheimer's disease contain the same protein as the amyloid of plaque cores and blood vessels. EMBO J 4:2757–2763.

Meszaros B, Simon I, Dosztanyi Z (2009) Prediction of protein binding regions in disordered proteins. PLoS Comput Biol 5:e1000376.

Meszaros B, Simon I, Dosztanyi Z (2011) The expanding view of protein–protein interactions: complexes involving intrinsically disordered proteins. Phys Biol 8: 035003.

Midic U, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN (2008) Protein disorder in the human diseasome: unfoldomics of human genetic diseases. PLoS Computational Biology 10:S12.

Midic U, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN (2009a) Unfoldomics of human genetic diseases: illustrative examples of ordered and intrinsically disordered members of the human diseasome. Protein Pept Lett 16:1533–1547.

Midic U, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN (2009b) Protein disorder in the human diseasome: unfoldomics of human genetic diseases. BMC Genomics 10 Suppl 1:S12.

Miller WG, Goebel CV (1968) Dimensions of protein random coils. Biochemistry 7:3925–935.

Miroy GJ, Lai Z, Lashuel HA, Peterson SA, Strang C, Kelly JW (1996) Inhibiting transthyretin amyloid fibril formation via protein stabilization. Proc Natl Acad Sci U S A 93:15051–15056.

Mitra K, Gangopadhaya PK, Das SK (2003) Parkinsonism plus syndrome: a review. Neurol India 51:183–188.

Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN (2006) Analysis of molecular recognition features (MoRFs). J Mol Biol 362: 1043–1059.

Mohan A, Sullivan WJ, Jr., Radivojac P, Dunker AK, Uversky VN (2008) Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. Mol Biosyst 4:328–340.

Mok KH, Kuhn LT, Goez M, Day IJ, Lin JC, Andersen NH, Hore PJ (2007) A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. Nature 447:106–109.

Morishima-Kawashima M, Hasegawa M, Takio K, Suzuki M, Yoshida H, Titani K, Ihara Y (1995a) Proline-directed and non-proline-directed phosphorylation of PHF-tau. J Biol Chem 270:823–829.

Morishima-Kawashima M, Hasegawa M, Takio K, Suzuki M, Yoshida H, Watanabe A, Titani K, Ihara Y (1995b) Hyperphosphorylation of tau in PHF. Neurobiol Aging 16:365–371; discussion 71–80.

Munishkina LA, Phelan C, Uversky VN, Fink AL (2003) Conformational behavior and aggregation of alpha-synuclein in organic solvents: modeling the effects of membranes. Biochemistry 42:2720–2730.

Munishkina LA, Henriques J, Uversky VN, Fink AL (2004) Role of protein–water interactions and electrostatics in alpha-synuclein fibril formation. Biochemistry 43:3289–3300.

Neri D, Billeter M, Wider G, Wuthrich K (1992a) NMR determination of residual structure in a urea-denatured protein, the 434-repressor. Science 257:1559–1563.

Neri D, Wider G, Wuthrich K (1992b) Complete ^{15}N and ^{1}H NMR assignments for the amino-terminal domain of the phage 434 repressor in the urea-unfolded form. Proc Natl Acad Sci U S A 89:4397–4401.

Neri D, Wider G, Wuthrich K (1992c) ^{1}H, ^{15}N and ^{13}C NMR assignments of the 434 repressor fragments 1–63 and 44–63 unfolded in 7 M urea. FEBS Lett 303: 129–135.

Nielsen L, Khurana R, Coats A, Frokjaer S, Brange J, Vyas S, Uversky VN, Fink AL (2001) Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. Biochemistry 40:6036–6046.

Nolting B, Golbik R, Soler-Gonzalez AS, Fersht AR (1997) Circular dichroism of denatured barstar suggests residual structure. Biochemistry 36:9899–9905.

Norris EH, Giasson BI, Lee VM (2004) Alpha-synuclein: normal function and role in neurodegenerative diseases. Curr Top Dev Biol 60:17–54.

Ohgushi M, Wada A (1983) "Molten-globule state": a compact form of globular proteins with mobile side-chains. FEBS Lett 164:21–24.

Ohnishi S, Lee AL, Edgell MH, Shortle D (2004) Direct demonstration of structural similarity between native and denatured eglin C. Biochemistry 43:4064–4070.

Okazawa H (2003) Polyglutamine diseases: a transcription disorder? Cell Mol Life Sci 60:1427–1439.

Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK (2005a) Comparing and combining predictors of mostly disordered proteins. Biochemistry 44:1989–2000.

Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK (2005b) Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 44:12454–12470.

Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics 9 Suppl 1:S1.

Orr HT (2001) Beyond the Qs in the polyglutamine diseases. Genes Dev 15: 925–932.

Palleros DR, Shi L, Reid KL, Fink AL (1993) Three-state denaturation of DnaK induced by guanidine hydrochloride: evidence for an expandable intermediate. Biochemistry 32:4314–4321.

Pavlov NA, Cherny DI, Heim G, Jovin TM, Subramaniam V (2002) Amyloid fibrils from the mammalian protein prothymosin alpha. FEBS Lett 517:37–40.

Permyakov EA, Berliner LJ (2000) alpha-Lactalbumin: structure and function. FEBS Lett 473:269–274.

Plaxco KW, Gross M (2001) Unfolded, yes, but random? Never! Nat Struct Biol 8:659–660.

Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A, Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI, Nussbaum RL (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276:2045–2047.

Povarova OI, Kuznetsova IM, Turoverov KK (2007) Different disturbances—one pathway of protein unfolding: actin folding–unfolding and misfolding. Cell Biol Int 31:405–412.

Prakash MK (2011) Insights on the role of (dis)order from protein–protein interaction linear free-energy relationships. J Am Chem Soc 133:9976–9979.

Prusiner SB (2001) Shattuck lecture: neurodegenerative diseases and prions. N Engl J Med 344:1516–1526.

Ptitsyn OB (1973) [Stages in the mechanism of self-organization of protein molecules]. Dokl Akad Nauk SSSR 210:1213–1215.

Ptitsyn OB (1995) Molten globule and protein folding. Adv Protein Chem 47: 83–229.

Ptitsyn OB, Volkenstein MV (1986) Protein structure and neutral theory of evolution. J Biomol Struct Dyn 4:137–156.

Ptitsyn OB, Bychkova VE, Uversky VN (1995) Kinetic and equilibrium folding intermediates. Philos Trans R Soc Lond B 348:35–41.

Radford SE (2000) Protein folding: progress made and promises ahead. Trends Biochem Sci 25:611–618.

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. Biophys J 92:1439–1456.

Ramachandran GN, Sasisekharan V (1968) Conformation of polypeptides and proteins. Adv Protein Chem 23:283–438.

Ramirez-Alvarado M, Merkel JS, Regan L (2000) A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. Proc Natl Acad Sci U S A 97:8979–8984.

Ratcliff K, Marqusee S (2010) Identification of residual structure in the unfolded state of ribonuclease H1 from the moderately thermophilic *Chlorobium tepidum*: comparison with thermophilic and mesophilic homologues. Biochemistry 49:5167–5175.

Rochet JC, Lansbury PT, Jr. (2000) Amyloid fibrillogenesis: themes and variations. Curr Opin Struct Biol 10:60–68.

Rocken C, Linke RP, Saeger W (1992) Immunohistology of islet amyloid polypeptide in diabetes mellitus: semi-quantitative studies in a post-mortem series. Virchows Arch A Pathol Anat Histopathol 421:339–344.

Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guilliot S, Dunker AK (1998) Thousands of proteins likely to have long disordered regions. Pac Symp Biocomput: 437–448.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. Proteins 42:38–48.

Romero-Beviar M, Martinez-Rodriguez S, Prieto J, Goormaghtigh E, Ariz U, Martinez-Chantar Mde L, Gomez J, Neira JL (2010) The N-terminal domain of the enzyme I is a monomeric well-folded protein with a low conformational stability and residual structure in the unfolded state. Protein Eng Des Sel 23:729–742.

Ropson IJ, Boyer JA, Dalessio PM (2006) A residual structure in unfolded intestinal fatty acid binding protein consists of amino acids that are neighbors in the native state. Biochemistry 45:2608–2617.

Saab-Rincon G, Froebe CL, Matthews CR (1993) Urea-induced unfolding of the alpha subunit of tryptophan synthase: one-dimensional proton NMR evidence for residual structure near histidine-92 at high denaturant concentration. Biochemistry 32:13981–13990.

Saab-Rincon G, Gualfetti PJ, Matthews CR (1996) Mutagenic and thermodynamic analyses of residual structure in the alpha subunit of tryptophan synthase. Biochemistry 35:1988–1994.

Santra MK, Dasgupta D, Panda D (2006) Pyrene excimer fluorescence of yeast alcohol dehydrogenase: a sensitive probe to investigate ligand binding and unfolding pathway of the enzyme. Photochem Photobiol 82:480–486.

Saraiva MJ (2001) Transthyretin amyloidosis: a tale of weak interactions. FEBS Lett 498:201–203.

Saraiva MJ, Birken S, Costa PP, Goodman DS (1984) Amyloid fibril protein in familial amyloidotic polyneuropathy, Portuguese type: definition of molecular abnormality in transthyretin (prealbumin). J Clin Invest 74:104–119.

Schneider HM, Storkel S, Will W (1980) [Amyloid of islets of Langerhans and its relation to diabetes mellitus (author's transl.).] Dtsch Med Wochenschr 105:1143–1147.

Schulman BA, Kim PS, Dobson CM, Redfield C (1997) A residue-specific NMR view of the non-cooperative unfolding of a molten globule. Nat Struct Biol 4:630–634.

Schultz CP (2000) Illuminating folding intermediates. Nat Struct Biol 7:7–10.

Schwarzinger S, Wright PE, Dyson HJ (2002) Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. Biochemistry 41:12681–12686.

Schweers O, Schonbrunn-Hanebeck E, Marx A, Mandelkow E (1994) Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. J Biol Chem 269:24290–24297.

Shimizu S, Chan HS (2002) Origins of protein denatured state compactness and hydrophobic clustering in aqueous urea: inferences from nonpolar potentials of mean force. Proteins 49:560–566.

Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci U S A 97:8868–8873.

Shortle D (1996a) The denatured state (the other half of the folding equation) and its role in protein stability. FASEB J 10:27–34.

Shortle D (1996b) Structural analysis of non-native states of proteins by NMR methods. Curr Opin Struct Biol 6:24–30.

Shortle D (2002) The expanded denatured state: an ensemble of conformations trapped in a locally encoded topological space. Adv Protein Chem 62:1–23.

Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. Science 293:487–489.

Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, Lincoln S, Crawley A, Hanson M, Maraganore D, Adler C, Cookson MR, Muenter M, Baptista M, Miller D, Blancato J, Hardy J, Gwinn-Hardy K (2003) Alpha-Synuclein locus triplication causes Parkinson's disease. Science 302:841.

Singleton A, Gwinn-Hardy K, Sharabi Y, Li ST, Holmes C, Dendi R, Hardy J, Crawley A, Goldstein DS (2004) Association between cardiac denervation and parkinsonism caused by alpha-synuclein gene triplication. Brain 127:768–772.

Smith LJ, Fiebig KM, Schwalbe H, Dobson CM (1996) The concept of a random coil: residual structure in peptides and denatured proteins. Fold Des 1:R95–R106.

Souillac PO, Uversky VN, Millett IS, Khurana R, Doniach S, Fink AL (2002a) Elucidation of the molecular mechanism during the early events in immunoglobulin light chain amyloid fibrillation: evidence for an off-pathway oligomer at acidic pH. J Biol Chem 277:12666–12679.

Souillac PO, Uversky VN, Millett IS, Khurana R, Doniach S, Fink AL (2002b) Effect of association state and conformational stability on the kinetics of immunoglobulin light chain amyloid fibril formation at physiological pH. J Biol Chem 277: 12657–12665.

Spillantini MG, Goedert M (2000) The alpha-synucleinopathies: Parkinson's disease, dementia with Lewy bodies, and multiple system atrophy. Ann N Y Acad Sci 920:16–27.

Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M (1997) Alpha-Synuclein in Lewy bodies. Nature 388:839–840.

Spillantini MG, Crowther RA, Jakes R, Hasegawa M, Goedert M (1998) Alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies. Proc Natl Acad Sci U S A 95:6469–6473.

Sunde M, Serpell LC, Bartlam M, Fraser PE, Pepys MB, Blake CC (1997) Common core structure of amyloid fibrils by synchrotron x-ray diffraction. J Mol Biol 273:729–739.

Sunde M, McGrath KC, Young L, Matthews JM, Chua EL, Mackay JP, Death AK (2004) TC-1 is a novel tumorigenic and natively disordered protein associated with thyroid cancer. Cancer Res 64:2766–2773.

Svensson M, Jonasson P, Freskgard PO, Jonsson BH, Lindgren M, Martensson LG, Gentile M, Boren K, Carlsson U (1995) Mapping the folding intermediate of

human carbonic anhydrase: II. Probing substructure by chemical reactivity and spin and fluorescence labeling of engineered cysteine residues. Biochemistry 34: 8606–8620.

Tafer H, Hiller S, Hilty C, Fernandez C, Wuthrich K (2004) Nonrandom structure in the urea-unfolded *Escherichia coli* outer membrane protein X (OmpX). Biochemistry 43:860–869.

Tanford C (1961) Physical Chemistry of Macromolecules. New York: Wiley.

Tanford C (1968) Protein denaturation. Adv Protein Chem 23:121–282.

Tcherkasskaya O, Uversky VN (2001) Denatured collapsed states in protein folding: example of apomyoglobin. Proteins 44:244–254.

Tcherkasskaya O, Davidson EA, Uversky VN (2003) Biophysical constraints for protein structure prediction. J Proteome Res 2:37–42.

Teplow DB (1998) Structural and kinetic features of amyloid beta-protein fibrillogenesis. Amyloid 5:121–142.

Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27: 527–533.

Tompa P (2011) Unstructural biology coming of age. Curr Opin Struct Biol 21: 419–425.

Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. Trends Biochem Sci 33:2–8.

Trojanowski JQ, Lee VM (2003) Parkinson's disease and related alpha-synucleinopathies are brain amyloidoses. Ann N Y Acad Sci 991:107–110.

Turoverov KK, Kuznetsova IM, Uversky VN (2010) The protein kingdom extended: ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. Prog Biophys Mol Biol 102:73–84.

Ueda K, Fukushima H, Masliah E, Xia Y, Iwai A, Yoshimoto M, Otero DA, Kondo J, Ihara Y, Saitoh T (1993) Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. Proc Natl Acad Sci U S A 90:11282–11286.

Ureta DB, Craig PO, Gomez GE, Delfino JM (2007) Assessing native and non-native conformational states of a protein by methylene carbene labeling: the case of *Bacillus licheniformis* beta-lactamase. Biochemistry 46:14567–14577.

Uversky VN (2002a) What does it mean to be natively unfolded? Eur J Biochem 269:2–12.

Uversky VN (2002b) Cracking the folding code: Why do some proteins adopt partially folded conformations, whereas other don't? FEBS Lett 514:181–183.

Uversky VN (2002c) Natively unfolded proteins: a point where biology waits for physics. Protein Sci 11:739–756.

Uversky VN (2003a) A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders. J Biomol Struct Dyn 21:211–234.

Uversky VN (2003b) Protein folding revisited. A polypeptide chain at the folding–misfolding–nonfolding cross-roads: Which way to go? Cell Mol Life Sci 60:1852–1871.

Uversky VN (2008) Amyloidogenesis of natively unfolded proteins. Curr Alzheimer Res 5:260–287.

Uversky VN (2009) Intrinsic disorder in proteins associated with neurodegenerative diseases. Front Biosci 14:5188–238.

Uversky VN (2011a) Intrinsically disordered proteins from A to Z. Int J Biochem Cell Biol 43:1090–1103.

Uversky VN (2011b) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. Chem Soc Rev 40:1623–1634.

Uversky VN (2013) Intrinsic disorder-based protein interactions and their modulators. Curr Pharm Des 19:4191–4213.

Uversky VN, Dunker AK (2010) Understanding protein non-folding. Biochim Biophys Acta 1804:1231–1264.

Uversky VN, Fink AL (2004) Conformational constraints for amyloid fibrillation: the importance of being unfolded. Biochim Biophys Acta 1698:131–153.

Uversky VN, Fink AL (2005) Pathways to amyloid fibril formation: partially folded intermediates in fibrillation of unfolded proteins. *In*: JD Sipe, ed. Amyloid Proteins: The Beta Pleated Sheet Conformation and Disease. Weinheim, Germany: Wiley-VCH, pp. 247–265.

Uversky VN, Ptitsyn OB (1994) "Partly folded" state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of beta-lactamase at low temperature. Biochemistry 33:2782–2791.

Uversky VN, Ptitsyn OB (1996a) Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. J Mol Biol 255:215–228.

Uversky VN, Ptitsyn OB (1996b) All-or-none solvent-induced transitions between native, molten globule and unfolded states in globular proteins. Fold Des 1: 117–122.

Uversky VN, Semisotnov GV, Pain RH, Ptitsyn OB (1992) "All-or-none" mechanism of the molten globule unfolding. FEBS Lett 314:89–92.

Uversky VN, Karnoup AS, Segel DJ, Seshadri S, Doniach S, Fink AL (1998a) Anion-induced folding of staphylococcal nuclease: characterization of multiple equilibrium partially folded intermediates. J Mol Biol 278:879–894.

Uversky VN, Winter S, Galzitskaya OV, Kittler L, Lober G (1998b) Hyperphosphorylation induces structural modification of tau-protein. FEBS Lett 439:21–25.

Uversky VN, Gillespie JR, Millett IS, Khodyakova AV, Vasiliev AM, Chernovskaya TV, Vasilenko RN, Kozlovskaya GD, Dolgikh DA, Fink AL, Doniach S, Abramov VM

(1999a) Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. Biochemistry 38:15009–15016.

Uversky VN, Talapatra A, Gillespie JR, Fink AL (1999b) Protein deposits as the molecular basis of amyloidosis: II. Localized amyloidosis and neurodegenerative disorders. Med Sci Monitor 5:1238–1254.

Uversky VN, Talapatra A, Gillespie JR, Fink AL (1999c) Protein deposits as the molecular basis of amyloidosis: I. Systemic amyloidoses. Med Sci Monitor 5:1001–1012.

Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41:415–427.

Uversky VN, Lee HJ, Li J, Fink AL, Lee SJ (2001a) Stabilization of partially folded conformation during alpha-synuclein oligomerization in both purified and cytosolic preparations. J Biol Chem 276:43495–43498.

Uversky VN, Li J, Fink AL (2001b) Evidence for a partially folded intermediate in alpha-synuclein fibril formation. J Biol Chem 276:10737–10744.

Uversky VN, Li J, Fink AL (2001c) Trimethylamine-*N*-oxide–induced folding of alpha-synuclein. FEBS Lett 509:31–35.

Uversky VN, Li J, Fink AL (2001d) Pesticides directly accelerate the rate of alpha-synuclein fibril formation: a possible factor in Parkinson's disease. FEBS Lett 500:105–108.

Uversky VN, Li J, Fink AL (2001e) Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein: a possible molecular NK between Parkinson's disease and heavy metal exposure. J Biol Chem 276: 44284–44296.

Uversky VN, Li J, Bower K, Fink AL (2002a) Synergistic effects of pesticides and metals on the fibrillation of alpha-synuclein: implications for Parkinson's disease. Neurotoxicology 23:527–536.

Uversky VN, Yamin G, Souillac PO, Goers J, Glaser CB, Fink AL (2002b) Methionine oxidation inhibits fibrillation of human alpha-synuclein in vitro. FEBS Lett 517: 239–244.

Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit 18:343–384.

Uversky VN, Roman A, Oldfield CJ, Dunker AK (2006) Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. J Proteome Res 5:1829–1842.

Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37:215–246.

Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, Vucetic S, Iakoucheva LM, Obradovic Z, Dunker AK (2009) Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. BMC Genomics 10 Suppl 1:S7.

Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, Uversky VN, Dunker AK (2007a) Characterization of molecular recognition features, MoRFs, and their binding partners. J Proteome Res 6:2351–2366.

Vacic V, Uversky VN, Dunker AK, Lonardi S (2007b) Composition Profiler: a tool for discovery and visualization of amino acid composition differences. BMC Bioinformatics 8:211.

von Bergen M, Barghorn S, Jeganathan S, Mandelkow EM, Mandelkow E (2006) Spectroscopic approaches to the conformation of tau protein in solution and in paired helical filaments. Neurodegener Dis 3:197–206.

Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN (2007) Functional anthology of intrinsic disorder: 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. J Proteome Res 6:1899–1916.

Vulliet R, Halloran SM, Braun RK, Smith AJ, Lee G (1992) Proline-directed phosphorylation of human tau protein. J Biol Chem 267:22570–22574.

Wallgren M, Aden J, Pylypenko O, Mikaelsson T, Johansson LB, Rak A, Wolf-Watz M (2008) Extreme temperature tolerance of a hyperthermophilic protein coupled to residual structure in the unfolded state. J Mol Biol 379:845–858.

Wang HM, Yu C (2011) Investigating the refolding pathway of human acidic fibroblast growth factor (hFGF-1) from the residual structure(s) obtained by denatured-state hydrogen/deuterium exchange. Biophys J 100:154–164.

Wang Y, Shortle D (1995) The equilibrium folding pathway of staphylococcal nuclease: identification of the most stable chain–chain interactions by NMR and CD spectroscopy. Biochemistry 34:15895–15905.

Wang Y, Shortle D (1996) A dynamic bundle of four adjacent hydrophobic segments in the denatured state of staphylococcal nuclease. Protein Sci 5:1898–1906.

Wang Y, Shortle D (1997) Residual helical and turn structure in the denatured state of staphylococcal nuclease: analysis of peptide fragments. Fold Des 2:93–100.

Wang Z, Plaxco KW, Makarov DE (2007) Influence of local and residual structures on the scaling behavior and dimensions of unfolded proteins. Biopolymers 86:321–328.

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337:635–645.

Watanabe A, Hasegawa M, Suzuki M, Takio K, Morishima-Kawashima M, Titani K, Arai T, Kosik KS, Ihara Y (1993) In vivo phosphorylation sites in fetal and adult rat tau. J Biol Chem 268:25712–25717.

Westermark P, Wernstedt C, Wilander E, Hayden DW, O'Brien TD, Johnson KH (1987) Amyloid fibrils in human insulinoma and islets of Langerhans of the diabetic cat are derived from a neuropeptide-like protein also present in normal islet cells. Proc Natl Acad Sci U S A 84:3881–3885.

Wetzel R (1997) Domain stability in immunoglobulin light chain deposition disorders. Adv Protein Chem 50:183–242.

Wirmer J, Berk H, Ugolini R, Redfield C, Schwalbe H (2006) Characterization of the unfolded state of bovine alpha-lactalbumin and comparison with unfolded states of homologous proteins. Protein Sci 15:1397–1407.

Wisniewski KE, Dalton AJ, McLachlan C, Wen GY, Wisniewski HM (1985) Alzheimer's disease in Down's syndrome: clinicopathologic studies. Neurology 35:957–961.

Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. J Mol Biol 293:321–331.

Wright PE, Dyson HJ (2009) Linking folding and binding. Curr Opin Struct Biol 19:31–38.

Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN (2007a) Functional anthology of intrinsic disorder: 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. J Proteome Res 6:1917–1932.

Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z (2007b) Functional anthology of intrinsic disorder: 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res 6:1882–1898.

Xue B, Dunker AK, Uversky VN (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J Biomol Struct Dyn 30:137–149.

Yoon S, Welsh WJ (2005) Rapid assessment of contact-dependent secondary structure propensity: relevance to amyloidogenic sequences. Proteins 60:110–117.

Zagorski MG, Yang J, Shao H, Ma K, Zeng H, Hong A (1999) Methodological and chemical factors affecting amyloid beta peptide amyloidogenicity. Methods Enzymol 309:189–204.

Zarranz JJ, Alegre J, Gomez-Esteban JC, Lezcano E, Ros R, Ampuero I, Vidal L, Hoenicka J, Rodriguez O, Atares B, Llorens V, Gomez Tortosa E, del Ser T, Munoz DG, de Yebenes JG (2004) The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. Ann Neurol 55:164–173.

Zerovnik E (2002) Amyloid-fibril formation. Proposed mechanisms and relevance to conformational disease. Eur J Biochem 269:3362–3371.

Zhang O, Forman-Kay JD (1995) Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer. Biochemistry 34: 6784–6794.

Zhang O, Forman-Kay JD (1997) NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. Biochemistry 36:3959–3970.

Zhang YL, Zhou JM, Tsou CL (1996) Sequential unfolding of adenylate kinase during denaturation by guanidine hydrochloride. Biochim Biophys Acta 1295: 239–244.

Zoghbi HY, Orr HT (1999) Polyglutamine diseases: protein cleavage and aggregation. Curr Opin Neurobiol 9:566–570.