Biologia Serbica, 2017, 39(1):105-111

DOI 10.5281/zenodo.827151

Minireview

The protein folding problem

Brankica G. Janković[1] and Natalija Đ. Polović[1*]

[1]*Department of Biochemistry, Faculty of Chemistry – University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Republic of Serbia*

Summary. The protein folding problem is the most important unsolved problem in structural biochemistry. The problem consists of three related puzzles: i) what is the physical folding code? ii) what is the folding mechanism? and iii) can we predict the 3D structure from the amino acid sequences of proteins? Bearing in mind the importance of protein folding, misfolding, aggregation and assembly in many different disciplines, from biophysics to biomedicine, finding solutions that would be generally applicable is of the utmost importance in biosciences.

Keywords: folding code, funnel energy landscapes, folding kinetics, protein aggregation, protein misfolding, structure prediction.

Introduction

Native proteins perform many different functions in a living organism. A protein's biological function is determined by its three-dimensional (3D) native structure, which is encoded by its amino acid sequence (Berg et al. 2002). Protein folding is the physical process by which the inactive nascent polypeptide chain obtains its native 3D structure, a conformation that is active and functional (Branden and Tooze 1999).

This year marks the 55th anniversary of the 1962 Nobel Prize in Chemistry awarded to Max Perutz and John Kendrew for their work in determining the 3D structure of globular proteins (Kendrew et al. 1958, 1960; Perutz et al. 1960). Kendrew described myoglobin structure in which polar amino acid residues are found at the surface of the molecule while the vast majority of hydrophobic amino acid residues are buried inside the protein (Fig. 1). Such an arrangement implied that water presence is essential for protein folding and raised the question of how protein native structure could be explained by physical principles. Moreover, Kendrew et al. (1958) elaborated: "perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be totally lacking in the kind of regularities which one instinctively anticipates, and it is more complicated than it has been predicated by any theory of protein structure". The basic principles underlying the protein folding phenomenon had been summarized in the pioneering works of Perutz (1960) and Kendrew (1960), and yet they have remained unsolved. According to the journal *Science* (the 125th anniversary special issue), the protein folding problem is listed as one of 125 most important unsolved scientific problems (Kennedy and Norman 2005). Nowadays, the protein folding problem has come to be associated with three separate but related questions: 1) the physical folding code: how is the native structure of a protein determined by the physicochemical properties encoded in the protein sequence? 2) the folding mechanism: how can a protein fold so fast and assume just one native conformation when a polypeptide chain has an enormous number of possible conformations? and 3) can we predict the 3D structure that an unknown protein will adopt? (Dill et al. 2008).

The physical folding code

Anfinsen's dogma – thermodynamic hypothesis

A major breakthrough in the understanding of the protein folding phenomenon was the thermodynamic hypothesis postulated by Christian Anfinsen. From his notorious experiment on the renaturation of ribonuclease (Nobel Prize 1972), Anfin-

*Corresponding author, e-mail: polovicn@chem.bg.ac.rs

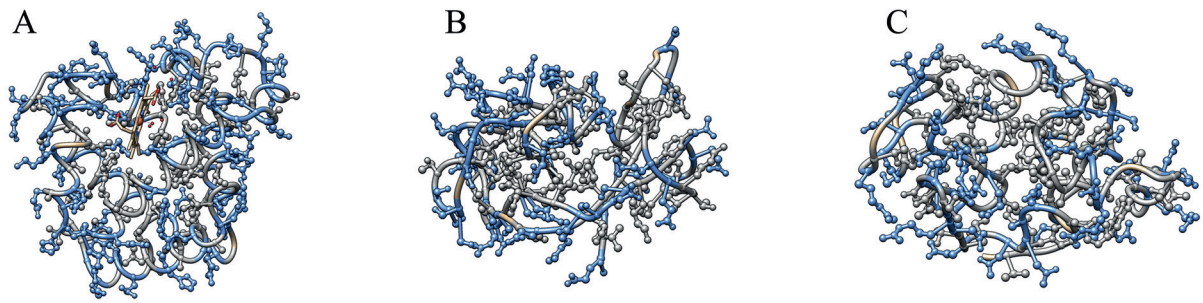

Fig. 1. Hydrophobic core of globular proteins. The protein backbone is shown as a ribbon, hydrophilic side chains are shown in blue and hydrophobic side chains are in grey. **A** – Sperm whale myoglobin; **B** – human ribonuclease; **C** – chicken lysozyme. Molecular graphics and analyses were performed with the UCSF Chimera package (Pettersen et al. 2004) using crystal structures deposed in the PDB data base.

sen concluded that the native protein structure in its physiological milieu is the one with the lowest Gibbs free energy; it is adopted by trial and error; it is exclusively dictated by the amino acid sequence and solution conditions and not by the folding pathway (Anfinsen 1973; Kresge et al. 2006). It became widely recognized that the native 3D structure does not thermodynamically depend on whether a protein is synthesized and folded in a cellular environment (in the presence of chaperones, etc.), or whether it is simply *in vitro* refolded as an isolated polypeptide in a test tube.

Native structure stability

Native proteins are only 20-65 kJ/mol more stable than unfolded chains, stabilized (and destabilized) by both enthalpic and entropic factors (Pace et al. 1996). Although there were many attempts to describe the effects of weak intramolecular interactions (hydrogen bonds, electrostatic interactions, van der Waals interactions) on protein folding and to estimate their impact on native state stability, it appears that intramolecular interactions do not predominantly stabilize the native state and cannot be a driving force of the protein folding process (Baldwin 2005; Dill et al. 2008).

Conformational entropy leads to an unordered coil, since an unordered chain has more degrees of freedom than the native structure (Doig and Sternberg 1995). There is considerable evidence that the water-driven process, the hydrophobic effect, is a dominant factor of the folding code (Dill et al. 2008). The main similarity in the native structure of different proteins, e.g. myoglobin, ribonuclease and lysozyme (Fig.1), is that they have hydrophobic cores, which implies that nonpolar amino acid residues are driven inside the native structure to lower the surface exposed to water. The hydrophobic effect is widely present in biological systems and is a major driving force of not just protein folding, but membrane assembly, exclusion of lipid drops, etc. (Spolar et al. 1989). The hydrophobic effect is presented in Fig. 2. Water molecules have a tendency to form ordered cage-like structures around nonpolar molecules, which dramatically decreases the entropy of water. During folding, nonpolar amino acid residues are grouped together into a globular hydrophobic core releasing an enormous number of water molecules. The rapid increase in water entropy provides sufficient energy to overcome unfavorable conformational entropy and to ensure stabilization of the native state (Lins and Brasseur 1995; Camilloni et al. 2016). There is plenty of experimental evidence for the importance of the hydrophobic effect in protein folding and stabilization. Model compound studies show that the transfer of just one hydrophobic amino acid residue from water to hydrophobic media corresponds to an energy release of 4-20 kJ/mol (Wolfenden 2007). Novel proteins designed by binary patterning of polar and nonpolar amino acid regions fold as their parent homologs (Kamtekar et al. 1993; Roy et al. 1997).

The folding mechanism

Levinthal's paradox

Anfinsen's experiments showed that all the information a protein needs in order to fold is already included in its amino acid sequence. But what is the first step in protein folding? How can proteins fold so fast in just one native conformation while a polypeptide chain has a vast number of possible conformations? Theoretically, in an unordered polypeptide, each amino acid residue can adopt any of 3 allowed angles to bind the neighboring residue, thus a small protein of 100 amino acids can have 3^{99} different conformations. On the other hand, native protein amino acids have just one unique combination of φ and ψ angles. In a trial-and-error folding scenario (the case of a polypeptide passing through all possible conformations), the folding process would last 10^{27} years; however, *in vitro* proteins fold within milliseconds (Zwanzig et al. 1992). Cyrus Levinthal (1968) pointed out this paradox and proposed that a polypeptide follows funnel-like pathways during folding, passing through several intermediate states, thereby reducing the range of possible conformations (Levinthal 1968; Bai 2003; Cruzeiro and Degrève 2017).

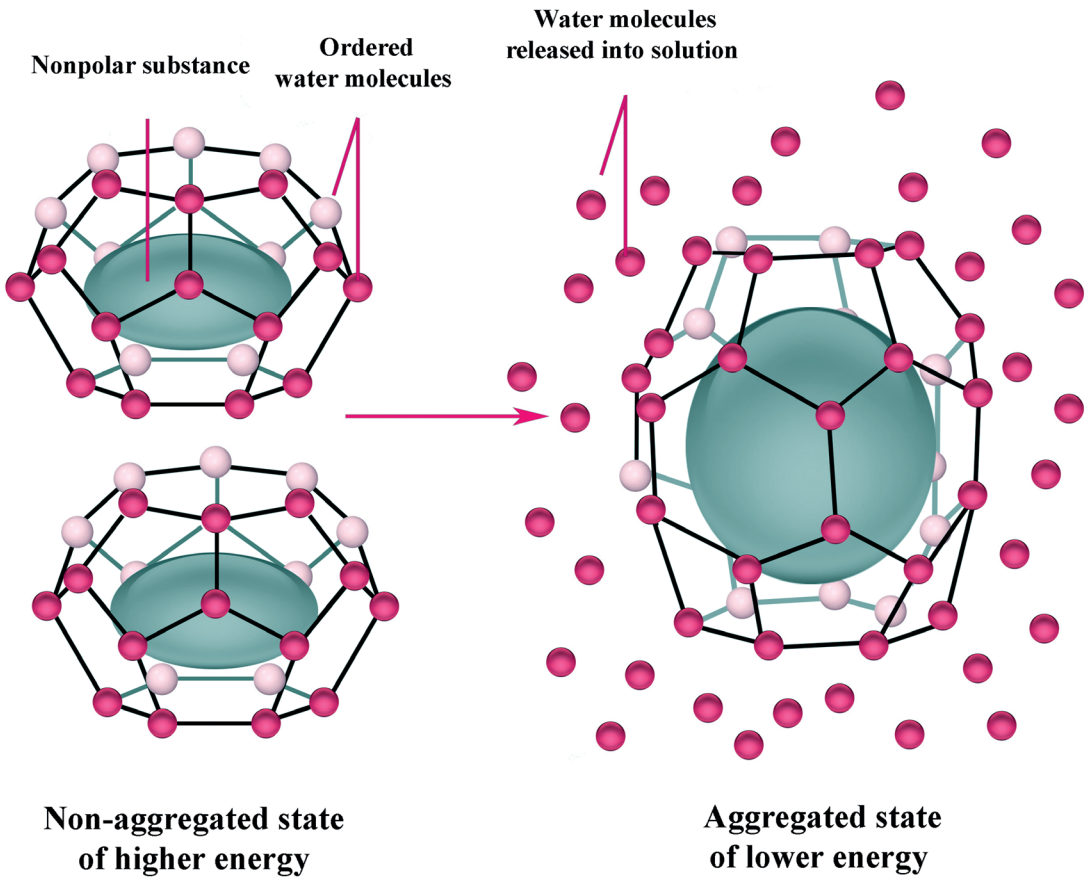

Fig. 2. Schematic representation of the hydrophobic effect. Explanations are given in the text.

Funnel-shaped energy landscape theory of folding

The funnel-shaped energy landscape theory of folding was postulated by Wolynes, Bryngelson and Dill (Service 2008) in the late nineties of the last century and up to now it is supported by experimental data regarding the detected folding intermediates, and by computational efforts aimed by making a series of snapshots of unordered polypeptides, folding intermediates and native structures, and calculations of their free energies (intramolecular interactions plus solvation free energy) (reviewed in Dill and Chan 1997; Dill et al. 2008).

A schematic representation of the folding energy landscape is shown in Fig. 3 (green funnel). The folding funnel shows that there is a great number of open unfolded conformations (funnel is wider), few local minima (folding intermediates), and just one, the most stable native structure. An unfolded polypeptide has the highest energy level (lowest stability). The polypeptide can use different pathways (it can have different folding intermediates) on its way down to a more stable (native) folded state. Local minima show kinetically stabile intermediates that can be kinetically entrapped since there is an activation energy barrier between them and the native state. The energy landscape theory of protein folding provides a universal answer to both protein folding basic kinetic and thermodynamic principles; however, according to Dill, "we are still missing a 'folding mechanism.' By mechanism, we mean a narrative that explains how the time evolution of a protein's folding to its native state derives from its amino acid sequence and solution conditions. A mechanism is more than just the sequences of events followed by any given protein in experiments or in computed trajectories. We do not yet have in hand a general principle that is applicable to a broad range of proteins that would explain differences and similarities of the folding routes and rates of different proteins in advance of the data." (Dill and MacCallum 2012).

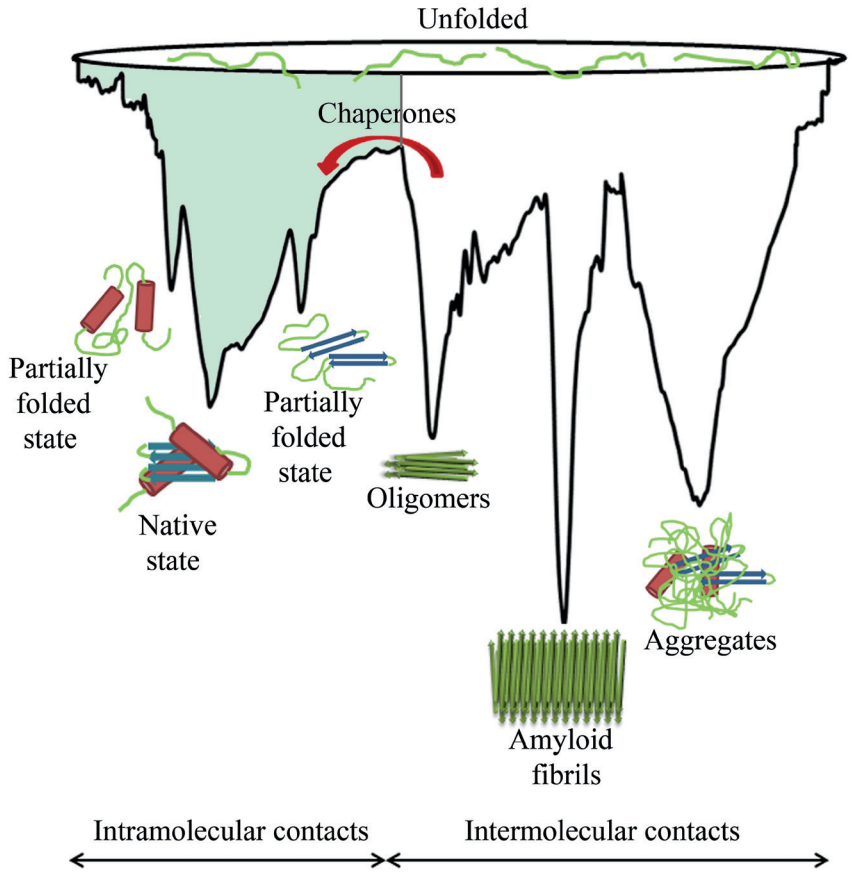

Fig. 3. Schematic representation of the funnel-shaped energy landscape theory of protein folding (green funnel) and aggregation (white funnels). Explanations are given in the text.

The energy landscape theory has been very successful in rationalizing the folding behavior of not only globular proteins, as this representation provides intuitive information on the number of states involved in the folding process, their populations and pathways of interconversion, but of intrinsically disordered proteins as well. The energy landscape of the Aβ40 peptide *per se* has inverted features with respect to those typical for folded proteins; however, the presence of binding partners can modify the energy landscape that provides binding-induced folding, and allows proteins to perform signaling and regulatory tasks (Granata et al. 2015). A number of different helicoidal structures with similar free energies have been described in the case of the p53 upregulated modulator of apoptosis (PUMA) peptide as well, and can be involved in the creation of early contacts with the binding partner and subsequent folding of a protein by induced fit mechanism (Chebaro et al. 2015).

Protein denaturation, misfolding and non-native aggregation

Protein denaturation includes any perturbation of a protein's native 3D structure due to changes in solution conditions. A protein's native structure can be easily disturbed by several possible stress conditions, such as elevated or lowered temperature, rapid shift in pH value, the presence of organic solvents miscible with water, chaotropes or detergents (Herczenik and Gebbink 2008; Hamada et al. 2009; Rašković et al. 2015a, 2015b). Although, there was a general belief that denaturation is a synonym for protein unfolding (transition to an unordered chain), it has been unequivocally shown that during denaturation protein misfolding occurs, not unfolding (except in the case of chaotrope denaturants) (Gianni and Jemth 2016). Denatured protein structures are unstable in solution and tend to aggregate in a hydrophobic manner, leading to the formation of elongated intermolecular β-sheets (Herczenik and Gebbink 2008; Rašković et al. 2015a, 2015b). Thus, the most accurate definition of denaturation is the transition of a native structure to any of the non-native structures (ordered and aggregated or disordered), followed by the loss of the protein's activity/function.

The extended version of the energy landscape theory can provide a comprehensive principle for the transitions of protein conformational states, such as protein non-native aggregation, denaturation and amyloid formation. Fig. 3 shows all of the conformational states that a polypeptide

can have. Folding intermediates with a predominant content of β-sheets can form a nucleus of aggregation (under appropriate conditions) rather than fold into a native structure, since all forms of aggregates are more stable than the native structure itself (Jahn and Radford 2005). In a living cells, chaperones and the proteasome-ubiquitin system prevent extensive aggregate formation. However, the cells' protective mechanism potential declines with age and is especially manifest in several forms of dementia and Alzheimer's disease (folding diseases) where amyloid aggregates can be detected in brain tissue (Dobson 2003; Hartl et al. 2011).

Protein structure prediction

The main task in theoretical protein biophysics is to develop a computer algorithm that can predict a protein's 3D structure from its amino acid sequence. In 2017, in the Uniprot data base there are about 85 million reported polypeptide sequences, but only about 130000 solved protein 3D structures (in the PDB base). Having in mind that there are many depositions of 3D structures of the same or homologous proteins, the PDB base contains only about 1900 3D structures of different protein folds.

Currently, all successful structure prediction algorithms are based on template modeling. The less accurate ones are based on sequence-sequence alignments, followed by choosing a template of maximal sequence homology with a target sequence and then fitting a target protein sequence into a structure of the chosen template. The main disadvantage of this type of modeling is its high inaccuracy if sequential homology between the target and template is not extremely high. More accurate algorithms are based on the assumption that the type of fold is more conserved in evolution then a simple amino acid sequence of a whole protein. Fold recognition as a concept was introduced 27 years ago as a strategy to model the structure of proteins with a limited number of homologs (Jones et al. 1992). Nowadays, fold recognition is further improved by fragment assembly. The fragment assembly approach was first tried in 2001 and currently it has several phases: fragmentation of the target protein into ~20 amino acid overlapping peptides; alignment of the fragment sequences with the structures of homologs or remote proteins in PDB; building the model, assessing and refining the model (Schwede et al. 2003; Zhexin 2006; Bordoli et al. 2008; Kelley et al. 2015).

The accuracy of protein structure prediction algorithms has rapidly advanced since 1994 as a consequence of a competition between protein modelers through the Critical Assessment of Protein Structure Predictions (CASP) network, whereby novel tools, such as molecular dynamics and the use of NMR experimental data were introduced in 2014, which has led to an improvement in models' accuracy (Moult et al. 2016).

Protein folding and stability in practice

Optimizing protein stability by altering conditions in solution

The presence of additives in a protein solution can have a positive or negative impact on protein stability. Additives can be roughly divided into kosmotropes (stabilizing agents) and chaotropes (destabilizing agents).

Kosmotropes do not show any binding affinity to the protein backbone or surface (they are preferentially excluded from the protein surface), ensuring the preservation of a protein's hydration shell and preservation of its structure and activity (regardless of temperature, pH value, etc.) The most common kosmotrope stabilizers used in protein biochemistry are saccharides, polyols, polymers, free amino acids (except arginine), phosphate and potassium ions (Rašković et al. 2016). Contrary to them, chaotropes (urea, guanidinium ion) are preferential binders to proteins' polypeptide backbones, leading to protein unfolding and solubilization of the unordered polypeptide (Moelbert et al. 2004).

In vitro refolding of recombinant proteins

Recombinant DNA technology provides a set of efficient techniques to produce rare or inaccessible proteins in an unlimited and inexpensive way. The main bottleneck in the production of heterologous proteins in bacteria (such as overexpressing *Escherichia coli*) is the aggregation of misfolded recombinant polypeptides into inclusion bodies. As inclusion bodies usually contain almost pure, intact recombinant polypeptides, protein refolding strategies could greatly increase the yield of native (and active) proteins. *In vitro* refolding methodology of overexpressed aggregated polypeptides should comprise 4 different steps: isolation and solubilization of inclusion bodies, protein refolding and protein purification.

Inclusion bodies are usually isolated by centrifugation of homogenized bacterial cells by French press or ultrasonication. Protein contaminants can be extracted from the crude preparation of inclusion bodies using detergents or low concentrations of chaotropes (Vallejo and Rinas 2004). In the next step, the misfolded recombinant polypeptides need to be solubilized, usually by high concentrations of chaotrope denaturants (6-8 M guanidinium ion or urea). The crucial step that limits proper renaturation of acceptable amounts of native protein is the removal of denaturants. Factors that should be taken into consideration are listed below.

It is essential to avoid aggregation of folding intermediates

Despite its extensive use over decades, recovery yields of dialysis are often below 40% due to aggregation of misfolded

proteins. On the other hand, simple dilution provides recoveries of more than 80% of protein, but its main disadvantage is an increase in sample volume of at least 10-fold (Yamaguchi and Miyazaki 2014). Chromatographic methods for the removal of denaturants are based on size-exclusion chromatography (SEC). During chromatography, denaturants can be removed by step or gradient elution, and buffer composition can be altered for further optimization of refolding conditions. Protein aggregation is expected to be reduced during SEC due to chromatographic separation of aggregation-prone folding intermediates and their subsequent refolding. Finally, the most effective refolding can be achieved using a solid support for the solubilized recombinant polypeptide before denaturant removal. His-tagged polypeptides can be immobilized on a metal-affinity matrix or polyanionic supports. Denaturants can then be removed by buffer exchange, and finally, refolded proteins can be detached from the metal-affinity matrix with EDTA or imidazole, or by buffers with high ionic strength in the case of a polyanion matrix (Vallejo and Rinas 2004). Chromatographic methods for denaturant removal provide purification of refolded proteins, as well.

The presence of stabilizing agents can significantly improve the folding process

The addition of polyols or saccharides as well-known kosmotropes that stabilize the protein folded state (like glycerol or sorbitol) to folding buffers leads to structural alterations and changes in water properties so as to assist folding. Kosmotropic salts at low concentrations increase protein solubility and prevent aggregation (Vallejo and Rinas 2004). The amino acid arginine, as a folding additive, produces the most controversial effect. The amino acid moiety of the arginine molecule provides kosmotropic stabilization of folded proteins, while its side chain guanidino-group shows slight chaotropic behavior and acts as an aggregation suppressor (Alibolandi and Mirzahoseini 2011).

Conclusions

Since the protein folding problem was postulated 55 years ago, our understanding has advanced considerably. Thanks to novel experimental and theoretical/computational methods we can now successfully design new proteins and foldamers. Funnel-shaped energy landscapes describe the conformational heterogeneity among non-native protein states and provide a key for understanding the equilibria between native and non-native states and folding kinetics. However, the future of structural biochemistry is very compelling. We know little about the (un)folding of membrane or intrinsically disordered proteins; experimental knowledge of energy landscapes is still very limited; structure prediction algorithms are still inconsistent; we do not have a deep understanding of the folding routes at the amino acid level; we cannot predict a protein's propensity to aggregate and we know little about how folding diseases develop and how to treat them. Fortunately, increasing experimental and theoretical knowledge suggests that modern science will rise successfully to this challenge in the years to come.

Acknowledgments

This work was supported by the Ministry of Education, Science and Technological Development, Republic of Serbia, Grant No. 172049.

References

Alibolandi M, Mirzahoseini H. 2011. Chemical assistance in refolding of bacterial inclusion bodies. Biochemistry Research International. 2011, Article ID 631607.

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

Bai Y. 2003. Hidden intermediates and Levinthal paradox in the folding of small proteins. Biochemical and Biophysical Research Communications. 305:785-788.

Baldwin RL. 2005. Weak interactions in protein folding: hydrophobic free energy, van der Waals interactions, peptide hydrogen bonds, and peptide solvation. In: Buchner J, Kiefhaber T, editors. Protein folding handbook. Weinheim (D): Wiley-VCH Verlag GmbH. p. 127-162.

Berg JM, Tymoczko JL, Stryer L. 2002. Biochemistry. Edition 5th. New York: W.H. Freeman and Co. Chapter 3, Protein structure and function.

Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. 2008. Protein structure homology modeling using SWISS-MODEL workspace. Nature Protocols. 4:1-13.

Branden CI, Tooze J. 1999. Introduction to protein structure. Edition 2nd. New York: Garland science Taylor and Francis group, LLC. Chapter 1, The building blocks. p. 3-12.

Camilloni C, Bonetti D, Morrone A, Giri R, Dobson CM, Brunori M, Gianni S, Vendruscolo M. 2016. Towards a structural biology of the hydrophobic effect in protein folding. Scientific Reports. 6:28285.

Chebaro Y, Ballard A, Chakraborty D, Wales D. 2015 Intrinsically Disordered Energy Landscapes. Scientific Reports. 5:10386

Cruzeiro L, Degrève L. 2017. Exploring the Levinthal limit in protein folding. Journal of Biological Physics. 43:15-30.

Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. Nature Structural Biology. 4:10-19.

Dill KA, MacCallum JL. 2012. The Protein-folding problem, 50 years on. Science. 338:1042-1046.

Dill KA, Ozkan SB, Shell MS, Weikl TR. 2008. The protein folding problem. Annual Review of Biophysics. 37:289-316.

Dobson CM. 2003. Protein folding and misfolding. Nature. 426:884-890.

Doig AJ, Sternberg MJ. 1995. Side-chain conformational entropy in protein folding. Protein Science: A Publication of the Protein Society. 4:2247-2251.

Gianni S, Jemth P. 2016. Protein folding: vexing debates on a fundamental problem. Biophysical Chemistry. 212:17-21.

Granata D, Baftizadeh F, Habchi J, Galvagnion C, De Simone A, Camilloni C, Laio A, Vendruscolo M. 2015. The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. Scientific Reports. 5:15449

Hamada D, Tanaka T, Tartaglia GG, Pawar A, Vendruscolo M, Kawamura M, Tamura A, Tanaka N, Dobson CM. 2009. Competition between folding, native-state dimerisation and amyloid aggregation in β-lactoglobulin. Journal of Molecular Biology. 386:878-890.

Hartl FU, Bracher A, Hayer-Hartl M. 2011. Molecular chaperones in

protein folding and proteostasis. Nature. 475:324-332.
Herczenik E, Gebbink MFBG. 2008. Molecular and cellular aspects of protein misfolding and disease. The FASEB Journal. 22:2115-2133.
Jahn TR, Radford SE. 2005. The yin and yang of protein folding. FEBS Journal. 272:5962-5970.
Jones DT, Taylor WR, Thornton JM. 1992. A new approach to protein fold recognition. Nature. 358: 86-89
Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. 1993. Protein design by binary patterning of polar and nonpolar amino acids. Science. 262:1680-1685.
Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 10:845-858.
Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature. 181:662-666.
Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC, Shore VC. 1960. Structure of myoglobin: a three-dimensional Fourier synthesis at 2 A. resolution. Nature. 185:422-427.
Kennedy D, Norman C. 2005. So much more to know. Science. 309:78-102.
Kresge N, Simoni RD, Hill RL. 2006. The thermodynamic hypothesis of protein folding: the work of Christian Anfinsen. Journal of Biological Chemistry. 281:e11-e11.
Levinthal C. 1968. Are there pathways for protein folding? Journal de Chimie Physique et de Physico-ChimieBiologique. 65:44-45.
Lins L, Brasseur R. 1995. The hydrophobic effect in protein folding. The FASEB Journal. 9:535-540.
Moelbert S, Normand B, De Los Rios P. 2004. Kosmotropes and chaotropes: modelling preferential exclusion, binding and aggregate stability. Biophysical Chemistry. 112:45-57.
Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. 2016. Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. Proteins: Structure, Function, and Bioinformatics. 84:4-14.
Pace CN, Shirley BA, McNutt M, Gajiwala K. 1996. Forces contributing to the conformational stability of proteins. The FASEB Journal. 10:75-83.
Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT. 1960. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature. 185:416-422.
Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of Computational Chemistry. 25:1605-1612.
Rašković B, Babić N, Korać J, Polović N. 2015a. Evidence of β-sheet structure induced kinetic stability of papain upon thermal and sodium dodecyl sulfate denaturation. Journal of the Serbian Chemical Society. 80:613-625.
Rašković B, Popović M, Ostojić S, Anđelković B, Tešević V, Polović N. 2015b. Fourier transform infrared spectroscopy provides an evidence of papain denaturation and aggregation during cold storage. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 150:238-246.
Rašković B, Vatić S, Anđelković B, Blagojević V, Polović N. 2016. Optimizing storage conditions to prevent cold denaturation of trypsin for sequencing and to prolong its shelf life. Biochemical Engineering Journal. 105:168-176.
Roy S, Ratnaswamy G, Boice JA, Fairman R, McLendon G, Hecht MH. 1997. A protein designed by binary patterning of polar and nonpolar amino acids displays native-like properties. Journal of the American Chemical Society. 119:5302-5306.
Schwede T, Kopp Jr, Guex N, Peitsch MC. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research. 31:3381-3385.
Service R. 2008. Problem Solved* (*sort of). Science. 321:784-786.
Spolar RS, Ha JH, Record MT. 1989. Hydrophobic effect in protein folding and other noncovalent processes involving proteins. Proceedings of the National Academy of Sciences of the United States of America. 86:8382-8385.
Vallejo LF, Rinas U. 2004. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microbial Cell Factories. 3:11.
Wolfenden R. 2007. Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins. The Journal of General Physiology. 129:357-362.
Yamaguchi H, Miyazaki M. 2014. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. Biomolecules. 4:235.
Zhexin X. 2006. Advances in homology protein structure modeling. Current Protein and Peptide Science. 7:217-227.
Zwanzig R, Szabo A, Bagchi B. 1992. Levinthal's paradox. Proceedings of the National Academy of Sciences. 89:20-22.