Universal Initial Thermodynamic Metastable State of Unfolded Proteins

Authors: Ma Xiaoliang, Hou Chengyu, Shi Liping, Li Long, Li Jiacheng, Ye Lin, Yang Lin, Xiaodong He, Yang Lin

Date: 2018-12-10T00:00:00+00:00

Abstract

Exploring and understanding the protein folding problem has long been a challenge in molecular biology. Unfolded proteins must exist in a universal initial thermodynamic metastable state; otherwise, it would be impossible to explain how proteins accomplish precise folding amidst intense thermal vibration interference. Through analyzing the correlation between the aqueous environment and protein folding, this study reveals a universal initial thermodynamic metastable state of unfolded proteins arising from the shielding effect of water molecules. The existence of this metastable state is determined by the physical properties of water molecules in the environment and is regarded as a necessary condition for proper protein folding. By investigating published experimental data and constructing molecular models, we have obtained relevant evidence for the existence of this universal initial thermodynamic metastable state, and through theoretical analysis, we have preliminarily explored the physical mechanisms by which this thermodynamic metastable state enables precise protein folding.

Full Text

The Universal Initial Thermodynamic Metastable State of Unfolded Proteins

Xiaoliang Ma[1], Chengyu Hou[2], Liping Shi[1], Long Li[3], Jiacheng Li[1], Lin Ye , Lin Yang[1], *, Xiaodong He[1]

[1]National Key Laboratory of Science and Technology on Advanced Composites in Special Environments, Harbin Institute of Technology, Harbin 150001, China
[2]School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China
[3]School of Electrical and Electronic Engineering, Harbin University of Science

and Technology, Harbin 150080, China
School of Mechanical and Mechatronic Engineering, University of Sydney, Sydney 2006, Australia

*Corresponding author. E-mail: linyang@hit.edu.cn

Abstract

Exploring and understanding the protein folding problem has long been a challenge in molecular biology. Unfolded proteins must possess a universal initial thermodynamic metastable state; otherwise, it would be impossible to explain how proteins achieve precise folding under intense thermal vibration interference. By analyzing the correlation between aqueous environments and protein folding, this paper reveals a universal initial thermodynamic metastable state of unfolded proteins caused by the water molecule shielding effect. The existence of this metastable state is determined by the physical properties of water molecules in the environment and is considered a necessary condition for correct protein folding. We have found evidence for this universal initial thermodynamic metastable state by studying published experimental data and establishing molecular models, and we have preliminarily explored the relevant physical mechanisms through which this thermodynamic metastable state leads to precise protein folding.

Keywords: protein folding, thermodynamic metastable state, water molecule shielding effect
Subject Classification: Q7

Proteins and their products are the foundation of life on Earth, with nearly all known important biochemical reactions and biological phenomena involving proteins [1]. There may be millions of different proteins in Earth' s organisms, and each protein' s biological function and activity are expressed through its three-dimensional shape. The ability of proteins to form three-dimensional structures arises from protein folding, which can therefore be regarded as the foundation of molecular biology [2,3]. Protein folding functionalizes polypeptide chains—for example, by creating the complex surface activity of globular proteins. Protein folding can be considered the most important mechanism and driving force for the existence, evolution, functionalization, and diversity of all life, directly leading to the generation of different protein types [4]. Since Anfinsen revealed that a protein' s three-dimensional structure is determined by its primary structure, research exploring the physical relationship between amino acid sequences and protein three-dimensional structures has become a major frontier field in molecular biology [5].

The protein folding problem was posed 60 years ago, and as one of the most important questions in molecular biology, extensive research has been conducted on protein folding mechanisms and principles, with related studies receiving multiple Nobel Prizes in Chemistry [5-8]. However, progress at the biophysical level remains limited, and we still cannot explain how nature accomplishes and

ensures precise protein folding, which restricts the development and application of protein research. The protein folding problem is considered a legacy of basic science, and its solution would broadly and profoundly advance the level and progress of molecular biology and drug development [1]. Ken Dill refined the protein folding problem into three classic questions at different levels [1]: (i) What is the physical folding code embedded in protein amino acid sequences that determines three-dimensional structure? (ii) What are the physical driving forces and mechanisms that enable rapid protein folding? (iii) Can protein three-dimensional structures be accurately predicted through computer algorithms? The protein folding process is guided by multiple physical forces: (i) hydrogen bond formation, (ii) van der Waals forces, (iii) electrostatic forces, (iv) hydrophobic interactions, and (v) entropy [1]. These physical forces can be approximately described by "force field functions," making molecular dynamics simulation an important tool for solving the protein folding problem. However, even with today' s highest-performance computers, the protein folding problem remains unresolved.

Protein folding primarily leads to the formation of numerous hydrogen bonds and hydrophobic bonds within the molecule, and the formation of these bonds releases substantial free energy while reducing the potential energy of the molecular structure. Therefore, protein folding might be viewed as a free energy release process or molecular structure relaxation process [9]. Hydrogen bonds are the primary contributors to protein structural stability. Notably, increasing temperature destabilizes protein three-dimensional structures and causes irreversible denaturation, with denatured protein structures potentially representing stable folded conformations. The reason why most proteins can only fold correctly within specific temperature ranges remains unknown.

A protein has an enormous number of potentially possible folding conformations. Since these potential conformations typically form numerous hydrogen bonds and van der Waals forces, the free energy of each folded conformation is far lower than that of the unfolded protein. If protein folding were simply a free energy release process, how could proteins rapidly fold into their unique native structures without misfolding [10]? Considering that some proteins fold within microseconds, how do they avoid misfolding without searching folding pathways? There is still no quantitative microscopic understanding of folding pathways for arbitrary amino acid sequences.

Experimental techniques such as laser temperature-jump and single-molecule experiments have revealed important principles of protein folding kinetics. Experimental results have confirmed that protein folding first leads to secondary structure formation, with tertiary structure formation occurring later [2]. Additionally, experiments have shown that secondary structure formation is extremely fast, completing within the microsecond timescale [11,12]. The native structure is often considered the most stable structure with minimum free energy [13], suggesting that protein secondary structures represent locally most stable structures with minimum potential energy. Classic secondary structures

— -helices and -sheets—indeed maximize the number of hydrogen bonds formed between backbone groups, as carbonyl oxygen atoms (C-O) and amide hydrogen atoms (N-H) on the backbone all form hydrogen bonds with each other [14,15]. The formation of these numerous hydrogen bonds ensures that the formation of -helices and -sheets can effectively release free energy and reduce molecular potential energy.

However, many protein secondary structures are not -helices or -sheets [14,15], meaning many protein secondary structures do not effectively form more hydrogen bonds to substantially reduce their free energy. Consequently, these secondary structures are not stable structures with minimized local molecular potential energy, and their stability is not high because the number of hydrogen bonds maintaining them is not large. Their structural stability is clearly incomparable to that of -helices and -sheets. Since the natural process of protein folding cannot guarantee that every segment of the amino acid sequence forms a potential energy-minimized stable structure, protein folding cannot be simply understood as a relaxation or free energy minimization process. It is difficult to explain how protein folding abandons numerous locally more stable structures to pursue the stability and potential energy minimization of tertiary structure. The formation of a protein' s hydrophobic core occurs late in protein folding, and hydrophobic bond formation requires hydrophobic amino acids to approach distances where surface tension interactions between hydrophobic regions occur. Therefore, it cannot be determined that hydrophobic interactions dominate the protein folding process. In summary, the formation of different secondary structures from different amino acid sequence segments cannot be simply explained as ordinary free energy release phenomena. The mechanism causing proteins to form specific secondary structures in specific amino acid sequence segments remains unknown—that is, the physical folding code of proteins is unknown.

Moreover, protein folding typically requires an aqueous environment and appropriate temperature. The environmental temperature required for protein folding is close to room temperature, and the thermal motion at this temperature is extremely intense for nanoscale unfolded protein molecules. One could even consider that unfolded protein molecules complete precise folding amidst violent thermal vibrations. In such an environment of intense thermal motion, precise protein folding is almost certainly not entirely determined by structural relaxation processes or hydrophobic interactions. Because unfolded protein molecule surfaces contain numerous hydrophilic functional groups, intense random thermal vibrations would cause these groups to randomly encounter each other and form hydrogen bonds, leading to misfolding. From this, we can speculate that unfolded protein molecules may possess some thermodynamic metastable state that prevents random folding during vibrations. Protein folding might utilize thermal vibrations to precisely disrupt this thermodynamic metastable state in certain amino acid sequence segments to form precise folding, which also explains the necessity of appropriate temperature for protein folding. Therefore, finding the thermodynamic metastable state of unfolded proteins is key to answering the protein folding problem.

If all unfolded proteins universally possess the same thermodynamic metastable state to ensure folding according to a clear physical folding code, this metastable state likely exists before folding begins—that is, the polypeptide chain molecule itself possesses some thermodynamic stable state. Once protein folding begins, it would be difficult for different types of proteins to form the same type of metastable state during the folding process.

Notably, proteins can only fold in aqueous environments and almost never fold correctly in any non-aqueous solution [16,17]. A comprehensive understanding of the interaction between unfolded proteins and water may lead to discovering the answer to the protein folding problem, or at least important thermodynamic metastable states along the folding pathway. Compared to other solvent molecules, water molecules have unique characteristics: they are very small, and each water molecule' s hydrogen and oxygen atoms carry positive and negative charges, respectively. This allows water molecules to insert into gaps between side chains of unfolded proteins and form hydrogen bonds with every hydrophilic side chain of the unfolded protein. In other words, the ability of these hydrophilic side chains to form hydrogen bonds can be saturated by numerous water molecules, preventing hydrogen bond formation between hydrophilic side chains and between hydrophilic side chains and the backbone. Water molecules in aqueous solutions can shield the hydrophilicity of each hydrophilic side chain. Under conditions where the aqueous environment causes hydrophilic side chains to lose activity, the state of the unfolded protein is likely determined by other adjacent charged atoms on the protein backbone.

[Figure 1: see original paper] The thermodynamic metastable state of unfolded proteins. (a) The stable state of adjacent peptide planes remaining parallel in aqueous environment. (b) The state of adjacent peptide planes remaining parallel widely exists in -sheets.

In aqueous environments, an unfolded protein conformation may first enter a thermodynamic metastable state to prevent misfolding. This metastable state may also be a necessary condition for the protein physical folding code to take effect, as shown in Figure 1a. Peptide bonds on the protein backbone cannot rotate freely due to their double-bond character. The carbonyl oxygen atom and amide hydrogen atom in each peptide plane carry negative and positive charges, respectively. Water molecules cannot shield the electrostatic attraction between two atoms this close. This electrostatic attraction between carbonyl oxygen and amide hydrogen atoms causes the corresponding C-O and N-H bonds to tend toward parallel alignment (see Figure 1a), which simultaneously causes the two adjacent peptide planes containing these bonds to tend toward parallel alignment. Considering that peptide planes are rigid planar structures, we find that the unfolded protein backbone in this state cannot fold freely because rotation of all single bonds is restricted. Without thermal motion disrupting this electrostatic attraction between carbonyl oxygen and amide hydrogen atoms, protein folding cannot occur. Only when adjacent carbonyl oxygen and amide hydrogen atoms move away from each other and escape electrostatic attraction can

chemical bonds on the backbone rotate freely. Thus, the state of adjacent peptide planes remaining parallel represents a universal thermodynamic metastable state of unfolded proteins. The most direct evidence for the existence of this thermodynamic metastable state is that the molecular structure of -sheets features adjacent peptide planes remaining parallel, meaning this metastable state extensively exists in protein native structures, as shown in Figure 1b. Moreover, even in protein native structures that are not -sheets, this state of adjacent peptide planes remaining parallel is also widespread (as shown in Figure 2 [Figure 2: see original paper]). This demonstrates that this state of adjacent peptide planes remaining parallel is an important thermodynamic stable state for polypeptide chains.

[Figure 2: see original paper] The state of adjacent peptide planes remaining parallel widely exists in other protein secondary structures: (a) 1UG1, (b) 1WFI, (c) 1X5H, (d) 1YDU.

Evidence demonstrates that water molecules indeed shield hydrogen bond formation between hydrophilic side chains and between hydrophilic side chains and the backbone. First, in protein molecular models described by the classic CHARMM [18] potential function, hydrogen atoms in water molecules carry a charge of (0.417e), which is generally greater than the charge of positively charged atoms in protein hydrophilic side chains (as shown in Figure 3 [Figure 3: see original paper]). Therefore, in aqueous environments, electronegative atoms in protein hydrophilic side chains will preferentially form hydrogen bonds with hydrogen atoms in water molecules. For the same reason, positively charged hydrogen atoms in protein hydrophilic side chains will preferentially form hydrogen bonds with negatively charged oxygen atoms in water molecules. Another piece of evidence is that many protein native structures are not compact [14,15]; many local structures could still approach each other to form numerous hydrogen bonds and create more compact, more stable protein structures. The existence of these non-compact protein structures also confirms the shielding effect of water molecules on hydrophilic side chains (as shown in Figure 2). Furthermore, protein secondary structures are stabilized by hydrogen bond formation between backbones, with hydrophilic side chains not participating in secondary structure formation, which also proves that protein hydrophilic side chains are shielded by water molecules.

[Figure 3: see original paper] Comparison of charge properties between water molecules and hydrophilic side chains.

Considering that during protein amino acid sequence generation in the ribosome, electrostatic attraction already occurs between carbonyl oxygen atoms and amide hydrogen atoms in free amino acid molecules, this metastable state exists immediately upon generation of the protein amino acid sequence. From PDB molecular models, we can find that the distance between adjacent carbonyl oxygen and amide hydrogen atoms in -sheets is approximately 2.2 Å. The interaction between carbonyl oxygen and amide hydrogen atoms at this distance already constitutes weak hydrogen bonding, indicating that this thermodynamic

metastable state is quite stable. This weak hydrogen bonding between adjacent peptide planes in the metastable state allows the unfolded protein backbone to bear certain torque and bending moments; rotation of one peptide plane will drive rotation of adjacent peptide planes, meaning torsional waves propagate along the unfolded protein backbone.

Since protein folding requires a certain temperature range, we can speculate that thermal motion caused by temperature disrupts the metastable state at specific amino acids in the sequence, causing the distance between carbonyl oxygen and amide hydrogen atoms at those amino acids to increase to a state where the parallel alignment of the corresponding C-O and N-H bonds cannot be restored, leading to folding at those amino acid positions. So how does an unfolded protein in this metastable state fold accurately? This requires analyzing what causes the electrostatic attraction between carbonyl oxygen and amide hydrogen atoms to be disrupted. Considering that the thermal environment inevitably causes side chain vibrations, the most likely cause is large vibrations or rotations of side chains connected to -carbon atoms between adjacent peptide planes, disrupting the electrostatic attraction state between the corresponding carbonyl oxygen and amide hydrogen atoms.

[Figure 4: see original paper] Hydrophobic collapse leading to destabilization and -helix formation.

What specific physical mechanism does temperature use to disrupt the electrostatic attraction state between carbonyl oxygen and amide hydrogen atoms at certain amino acids on unfolded proteins, thereby initiating folding? The metastable state of adjacent peptide planes remaining parallel prevents each peptide plane from rotating freely, so torsional waves generated by side chain thermal vibrations propagate along the unfolded protein backbone. When differences in torsional resistance between adjacent side chains in the protein amino acid sequence hinder the propagation of torsional waves along the protein backbone, folding is likely to be triggered [19]. Torsional waves propagating along the backbone due to temperature may also initiate electrostatic attraction or repulsion between adjacent side chains, which could also cause protein folding. Furthermore, when a segment of consecutive hydrophobic amino acids exists in the unfolded protein' s amino acid sequence, these consecutive hydrophobic amino acids likely form a continuous hydrophobic region. Under entropy effects, this irregular hydrophobic region experiences hydrophobic interactions that cause it to collapse. The combination of hydrophobic interactions and side chain thermal vibrations can also disrupt the thermodynamic metastable state of these amino acids. When these amino acids lose their metastable state, collapse of their hydrophobic region likely forms these amino acids into a helical structure, as shown in Figure 4. If we can discover that folding codes composed of five consecutive hydrophobic amino acids widely exist in -helices of protein structures, we can identify them as the physical folding code of proteins. By comparing PDB files [20], we searched for 30 such codes and found that 90% of them exist in -helices, proving the existence of this folding code.

The rotatable single bonds on the unfolded protein backbone give each amino acid in the sequence at least two degrees of freedom for folding, leading to countless possible folding conformations. Considering that the temperature environment required for protein folding in aqueous solutions causes intense thermal vibrations and misfolding of unfolded proteins, we can speculate that a universal initial thermodynamic metastable state exists to prevent interference from thermal vibrations during folding. By discovering that water molecules can shield the hydrophilicity of hydrophilic side chains, we have revealed a universal initial thermodynamic metastable state of unfolded proteins characterized by adjacent peptide planes remaining parallel. This metastable state enables unfolded proteins to bear certain torque and bending moments without folding, preventing misfolding. We have identified three physical mechanisms that can disrupt the thermodynamic metastable state at specific amino acids in unfolded protein sequences, with torsional waves propagating along the protein backbone being key to initiating these physical mechanisms. Studying these physical mechanisms can decipher the physical code of protein folding. The universal initial thermodynamic metastable state of unfolded proteins revealed here suggests that the protein folding problem may have a Newtonian mechanics-based answer, and the protein folding problem may be a legacy of basic science left for Newtonian mechanics.

References

[1] K. A. Dill, J. L. MacCallum, The Protein-Folding Problem, 50 Years On. Science 338, 1042-1046(2012).
[2] K. A. Dill, S. B. Ozkan, M. S. Shell, T. R. Weikl, The Protein Folding Problem. Annual review of biophysics 37, 289-316 (2008).
[3] Igor K. Lednev, Amyloid Fibrils: the Eighth Wonder of the World in Protein Folding and Aggregation. Biophysical Journal 106, 1433-1435 (2014).
[4] N. V. Grishin, Fold Change in Evolution of Protein Structures. Journal of Structural Biology 134, 167-185 (2001).
[5] C. B. Anfinsen, Principles that Govern the Folding of Protein Chains. Science 181, 223-230 (1973).
[6] J. C. Kendrew et al., Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. Nature 185, 422 (1960).
[7] K. Wakabayashi, [Accomplishment of Dr. Aaron Klug, winner of Nobel prize in chemistry, 1982]. Tanpakushitsu kakusan koso. Protein, nucleic acid, enzyme 28, 156-157 (1983).
[8] M. V. Rodnina, W. Wintermeyer, The ribosome goes Nobel. Trends in biochemical sciences 35, 1-5 (2010).
[9] a. G D Rose, R. Wolfenden, Hydrogen Bonding, Hydrophobicity, Packing, and Protein Folding. Annual Review of Biophysics and Biomolecular Structure 22, 381-415 (1993).
[10] A. C. Clark, Protein folding: are we there yet? Archives of biochemistry and biophysics 469, 1- 3 (2008).
[11] D. T. Clarke, A. J. Doig, B. J. Stapley, G. R. Jones, The -helix folds on

the millisecond time scale. Proceedings of the National Academy of Sciences of the United States of America 96, 7232-7237 (1999).
[12] E. H. L. Chen et al., Directly monitor protein rearrangement on a nanosecond-to-millisecond time- scale. Scientific Reports 7, 8691 (2017).
[13] E. Alm, D. Baker, Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proceedings of the National Academy of Sciences of the United States of America 96, 11305-11310 (1999).
[14] H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank. Nature Structural Biology 10, 980 (2003).
[15] H. Berman, K. Henrick, H. Nakamura, J. L. Markley, The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Research 35, D301-D303 (2007).
[16] O. Collet, How does the first water shell fold proteins so fast? The Journal of Chemical Physics 134, 085107 (2011).
[17] C. M. Soares, V. H. Teixeira, A. M. Baptista, Protein Structure and Dynamics in Nonaqueous Solvents: Insights from Molecular Dynamics Simulation Studies. Biophysical Journal 84, 1628-1641 (2003).
[18] Brooks B R, Bruccoleri R E, Olafson B D, et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations[J]. Journal of Computational Chemistry, 2010, 4(2):187-217.
[19] D. Eisenberg, Three-dimensional structure of membrane and surface proteins. Annual review of biochemistry 53, 595-623 (1984).
[20] Sussman J L, Abola E E, Lin D, et al. The protein data bank[J]. Genetica, 2000.

Note: Figure translations are in progress. See original paper for figures.