

Article

The “Beacon” Structural Model of Protein Folding: Application for Trp-Cage in Water

Qiang Sun *, Xian He  and Yanfang Fu

Key Laboratory of Orogenic Belts and Crustal Evolution, Ministry of Education, The School of Earth and Space Sciences, Peking University, Beijing 100871, China; xianhe@stu.pku.edu.cn (X.H.); yanfang.fu@stu.pku.edu.cn (Y.F.)

* Correspondence: qiangsun@pku.edu.cn

Abstract: Protein folding is a process in which a polypeptide must undergo folding process to obtain its three-dimensional structure. Thermodynamically, it is a process of enthalpy to overcome the loss of conformational entropy in folding. Folding is primarily related to hydrophobic interactions and intramolecular hydrogen bondings. During folding, hydrophobic interactions are regarded to be the driving forces, especially in the initial structural collapse of a protein. Additionally, folding is guided by the strong interactions within proteins, such as intramolecular hydrogen bondings related to the α -helices and β -sheets of proteins. Therefore, a protein is divided into the folding key (FK) regions related to intramolecular hydrogen bondings and the non-folding key (non-FK) regions. Various conformations are expected for FK and non-FK regions. Different from non-FK regions, it is necessary for FK regions to form the specific conformations in folding, which are regarded as the necessary folding pathways (or “beacons”). Additionally, sequential folding is expected for the FK regions, and the intermediate state is found during folding. They are reflected on the local basins in the free energy landscape (FEL) of folding. To demonstrate the structural model, molecular dynamics (MD) simulations are conducted on the folding pathway of the TRP-cage in water.

Keywords: protein folding; water; hydrophobic interactions; hydrogen bonding; necessary pathway



Citation: Sun, Q.; He, X.; Fu, Y. The “Beacon” Structural Model of Protein Folding: Application for Trp-Cage in Water. *Molecules* **2023**, *28*, 5164. <https://doi.org/10.3390/molecules28135164>

Academic Editor: Takeshi Kikuchi

Received: 30 May 2023

Revised: 30 June 2023

Accepted: 30 June 2023

Published: 2 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Proteins play essential roles in biological processes, functioning as exquisite catalysts, inhibitors, and sensors. To perform the functions, most proteins are necessarily folded into compact three-dimensional structures, native states, and remain stably folded. To predict the native structure of a protein from its primary structure, it is important to understand the process of protein folding. Thus far, many experimental and theoretical works have been devoted to investigating the mechanism of protein folding.

Protein folding is a process of molecular self-assembly, during which a disordered polypeptide chain will collapse to form a compact and well-defined three-dimensional structure. It is one of the most important problems of molecular biology. From an Anfinsen experiment on ribonuclease A, this states that the information required to form folded conformation resides in its polypeptide amino acid sequence as the denatured enzyme re-folds into native conformation without assistance of any other protein [1]. According to the Levinthal study [2], it means that protein folding cannot take place via a random sampling of all possible conformations. Additionally, many physical models have been proposed to explain the high speed of protein folding, such as the nucleation growth model [3,4], framework model [5,6], diffusion–collision mechanism model [7,8], hydrophobic collapse model [9,10], and jigsaw model [11]. In these “classical views” of protein folding, it was assumed that proteins could find their predestined native states within the vast space of possible folds only by traveling through some predetermined pathway (or intermediates), which are considered to be partially unfolded conformations that are stable enough to be detected.

In recent years, the free-energy folding funnel model, derived from hypothetical energy landscape and statistical mechanical considerations [12–18], was developed to investigate protein folding. In this model, proteins are in disordered states at the highest energy level. As the proteins fold into the organized, native-like conformation, they shift to a lower energy phase. The energy landscape theory states that a folding of a protein does not follow a singular, specific pathway but occurs through statistical description of the topography of the free-energy landscape [19]. Therefore, it suggests that predefined pathways with compulsory intermediates simply do not exist. However, this is incompatible with the work of Englander, Mayne, and co-workers [20]. Based on hydrogen exchange studies, they suggested that a protein will fold by foldons if the protein contains foldons because this route is the fastest route for folding [21–26]. The foldon evidence indicates that protein folding follows a single folding pathway because the pathway is foldon-directed. To date, there is a strong debate on the folding pathways of proteins in water, with single or multiple pathways.

Hydrophobic interactions play an important role in protein folding. Therefore, it is important to study the physical origin of hydrophobic interactions. Historically, the classical mechanism for hydrophobicity, proposed by Frank and Evans [27] and advanced by Kauzmann [28] and many others, was based on the “iceberg” model. This means that the water immediately surrounding the hydrophobic group is more “ordered” than in bulk water as a small hydrophobic solute (such as argon or methane) is embedded into water. Later, it was recognized that hydrophobic interactions may be related to the solute size. When large hydrophobic solutes are inserted into water, as suggested by Stillinger [29], a distorted hydrogen bond network cannot be maintained in the water layer next to an extended hydrophobic surface, and the hydrogen bond network is then disrupted. In Lum, Chandler, and Weeks (LCW)’s [30–32] works, they provided a quantitative description of structural and thermodynamic aspects of hydrophobic hydration over the entire small to large length scale region. The crossover from small- to large-scale behavior occurs as the solute radius is about 1 nm. In our recent works [33–39], hydration-free energy is derived and used to understand the mechanism of hydrophobic effects. They are ascribed to the structural competition between interfacial and bulk water.

In combination with our recent studies [33–39] on hydrophobic interactions, this work is devoted to investigating the folding mechanisms of proteins in water. Thermodynamically, it is an enthalpic process to overcome the loss of conformational entropy in folding. Protein folding is primarily related to hydrophobic interactions and intramolecular hydrogen bondings. Hydrophobic interactions play an important role in the whole folding process, especially in the initial structural collapse of a protein. Additionally, based on thermodynamic analysis, folding is reasonably guided by the strong interactions within proteins, such as intramolecular hydrogen bondings related to the α -helices and β -sheets of the secondary structures. Therefore, proteins are divided into FK regions related to intramolecular hydrogen bondings and non-FK regions. During folding, various conformations are expected for FK and non-FK regions, such as specific and dynamic conformations. Different from non-FK regions, it is necessary for FK regions to form specific conformations during folding, which are regarded as the necessary pathways (“beacons”) in folding. Additionally, they are reflected on the local basins in the FEL of folding. To demonstrate the structural model, MD simulations are conducted on the folding pathway of the TRP-cage in water.

2. Hydrophobic Interactions

Protein folding is a physical process, during which a disordered polypeptide chain is collapsed and folded to attain a compact and well-defined three-dimensional structure. According to Anfinsen’s study [1], this means that the three-dimensional structure of a native protein may be the one in which the Gibbs free energy of a whole system is at its lowest. Therefore, under physiological conditions, a protein tends to be in its most stabilized form and remains biologically active in its three-dimensional structures. During

folding from the extended to folded conformations, the lowest Gibbs energies may be expected for the folding pathways of proteins in water.

Protein folding may be related to various forces exerted on the atoms of the amino-acid chain, such as hydrophobic interactions, hydrogen bondings, van der Waals interactions, etc. These forces may be due to the interactions within a protein itself (direct forces), as well as related to the solvent (solvent-induced forces). Thermodynamically, as the solutes are embedded into water, the thermodynamic functions include water–water, solute–water, and solute–solute interaction energies,

$$\Delta G = \Delta G_{\text{Water-water}} + \Delta G_{\text{Solute-water}} + \Delta G_{\text{Solute-solute}} \quad (1)$$

In fact, it is necessary for solutes to approach each other before they are affected by the direct solute-solute interactions ($\Delta G_{\text{Solute-solute}}$). This is closely related to $\Delta G_{\text{Water-water}}$, and $\Delta G_{\text{Solute-water}}$. Therefore, to understand the driving force during protein folding, it is necessary to investigate the structure of water, and the effects of solutes on water structure, respectively.

Numerous works have been carried out to understand the structure of water. To date, various structural models have been proposed, which are generally partitioned into the mixture and continuum models [40,41].

$$HB_{\text{Mixture model}} = (HB)_{\text{Low-density type}} + (HB)_{\text{High-density type}} \quad (2a)$$

$$HB_{\text{Continuum model}} = (HB \text{ networks})_{\text{Various bond-length and bond-angle}} \quad (2b)$$

In the mixture model, two distinct structural types are regarded to simultaneously exist in ambient water. For the continuum model, water comprises a random, three-dimensional hydrogen-bonded network. It may be characterized by a broad distribution of O-H...O hydrogen bond distances and angles. However, the structural networks cannot be “broken” (or separated) into distinct molecular species as in the mixture model. Thus far, liquid water is usually regarded as a tetrahedral fluid, which is based on the first coordination number $N_c = 4\pi\rho \int_{r_{\min}}^{r_{\max}} r^2 g_{\text{OO}}(r) dr$, where ρ means the density of water, and r_{\min} and r_{\max} are the lower and upper limits of integration in $g_{\text{OO}}(r)$. For ambient water, N_c is determined to be 4.3 [42] and 4.7 [43], respectively.

Water is generally regarded as an anomalous liquid, which is related to the hydrogen bondings of water. The OH vibrations are sensitive to hydrogen bondings of water and widely used to study the structure of water. From the Raman spectroscopic studies [44–46], as three-dimensional hydrogen bondings appear, various OH vibration frequencies correspond to different hydrogen-bonded networks in the first shell of a water molecule (local hydrogen bonding), and the effects of hydrogen bonding beyond the first shell on OH vibrations are weak. Therefore, when three-dimensional hydrogen bondings occur, different OH vibrations may be due to various local hydrogen-bonded networks of a water molecule.

At ambient conditions, the Raman OH stretching band of water may be fitted into five sub-bands, which are ascribed to OH vibrations engaged in various local hydrogen-bonded networks, such as DDAA (double donor–double acceptor, tetrahedral hydrogen bonding), DDA (double donor–single acceptor), DAA (single donor–double acceptor), DA (single donor–single acceptor), and free OH vibrations, respectively [44–47]. Therefore, a local statistical model (LSM) is proposed for ambient water, which means that a water molecule interacts with neighboring water molecules (in the first shell) through various local hydrogen bondings. Additionally, the hydrogen bondings of water may be influenced by the changes in temperature, pressure, dissolved salt, confined environment, etc.

When a solute is embedded into water, the solute–water interface occurs, which undoubtedly affects the structure of water. The OH vibrations are primarily dependent on the local hydrogen bondings of a water molecule. Therefore, the dissolved solute mainly affects the structure of interfacial water (the topmost water layer at the solute–water interface) [48]. In comparison with bulk water, no DDAA (tetrahedral) hydrogen bondings

are found in interfacial water [48]. Therefore, the Gibbs-free energy of interfacial water, incurred by the solute, is expressed as

$$\Delta G_{\text{Solute-water}} = n_{\text{DDAA}} \cdot \Delta G_{\text{DDAA}} \cdot R_{\text{Interfacial water}} \quad (3)$$

where n_{DDAA} is the hydrogen bondings per water molecule of DDAA hydrogen bondings, ΔG_{DDAA} means the Gibbs energy of DDAA hydrogen bondings, and $R_{\text{Interfacial water}}$ is the ratio of interfacial water to bulk water.

From the Raman spectroscopic studies [44–46], DDAA (tetrahedral) and DA are the predominant hydrogen-bonded networks in ambient water. Additionally, they are related to the structural changes across the solute–water interface. It is important to understand the characteristics of DDAA and DA hydrogen bondings. From our recent study [39], in comparison with a DDAA hydrogen-bonded network, the DA structural motif owns a lower enthalpy and a higher entropy and density.

Hydration-free energy means the change in Gibbs energy as a solute is transferred from a vacuum (or the gas phase) to a solvent. After the solute is simply regarded as a sphere, the $R_{\text{Interfacial water/volume}}$ is $4 \cdot r_{\text{H}_2\text{O}}/R$, where R is the radius of solute. Therefore, hydration-free energy is given as (Figure 1)

$$\begin{aligned} \Delta G_{\text{Hydration}} &= \Delta G_{\text{Water-water}} + \Delta G_{\text{Solute-water}} \\ &= \Delta G_{\text{Water-water}} + \frac{8 \cdot \Delta G_{\text{DDAA}} \cdot r_{\text{H}_2\text{O}}}{R} \end{aligned} \quad (4)$$

where $\Delta G_{\text{Water-water}}$ is the Gibbs energy of water, and $r_{\text{H}_2\text{O}}$ is the average radius of a H_2O molecule. At 293 K and 0.1 MPa, $\Delta G_{\text{Water-water}}$ is -1500 cal/mol [49]. At ambient conditions, the average volume per water molecule is 3×10^{-29} m³. After it is treated as a sphere, $r_{\text{H}_2\text{O}}$ is determined to be 1.9 Å.

In thermodynamics, the lower the hydration-free energy the more stable the system is. Hydration-free energy is related to $\Delta G_{\text{Water-water}}$ and $\Delta G_{\text{Solute-water}}$, and it may be dominated by $\Delta G_{\text{Water-water}}$ or $\Delta G_{\text{Solute-water}}$. This means that the structural transition may take place as $\Delta G_{\text{Water-water}}$ being equal to $\Delta G_{\text{Solute-water}}$,

$$\Delta G_{\text{Water-water}} = \Delta G_{\text{Solute-water}} \quad (R_c = \frac{8 \cdot \Delta G_{\text{DDAA}} \cdot r_{\text{H}_2\text{O}}}{\Delta G_{\text{Water-water}}}) \quad (5)$$

where R_c is the critical radius of the solute [33,39]. At 293 K and 0.1 MPa, R_c is 6.5 Å for a sphere solute [33]. With increasing the solute size (or concentrations), it is divided into initial and hydrophobic solvation processes (Figure 1). Additionally, $\Delta G_{\text{Solute-water}}$ is proportional to the ratio of surface area to volume of solute ($1/R$), and various dissolved behaviors of solutes are expected in different solvation processes.

In the initial solvation process, $\Delta G_{\text{Solute-water}}$ is less than $\Delta G_{\text{Water-water}}$, and hydration-free energy is dominated by $\Delta G_{\text{Solute-water}}$. To be more thermodynamically stable, this is fulfilled through maximizing $\Delta G_{\text{Solute-water}}$. In other words, it may be achieved through the maximization of the surface area to the volume ratio of the solutes in the water. Therefore, the dissolved solutes tend to be dispersed in solutions, and water molecules are found between them. Additionally, in the initial solvation process, the driving force is thermodynamically due to the increase in entropy arising from interfacial water.

In the hydrophobic solvation process, the Gibbs-free energy of interfacial water is higher than bulk water ($\Delta G_{\text{Solute-water}} > \Delta G_{\text{Water-water}}$). To be more thermodynamically stable, this may be fulfilled through maximizing $\Delta G_{\text{Water-water}}$. In fact, it is accompanied with the minimization of $\Delta G_{\text{Solute-water}}$. As $\Delta G_{\text{Solute-water}}$ is proportional to the surface area to volume ratio of solute ($1/R$), the solutes may be aggregated to minimize the ratio of surface area to volume of them (Figure 2). Therefore, hydrophobic effects may be expected in the hydrophobic solvation process. In thermodynamics, the hydrophobic solvation process is ascribed to be an enthalpic process related to maximizing the hydrogen bondings of water.

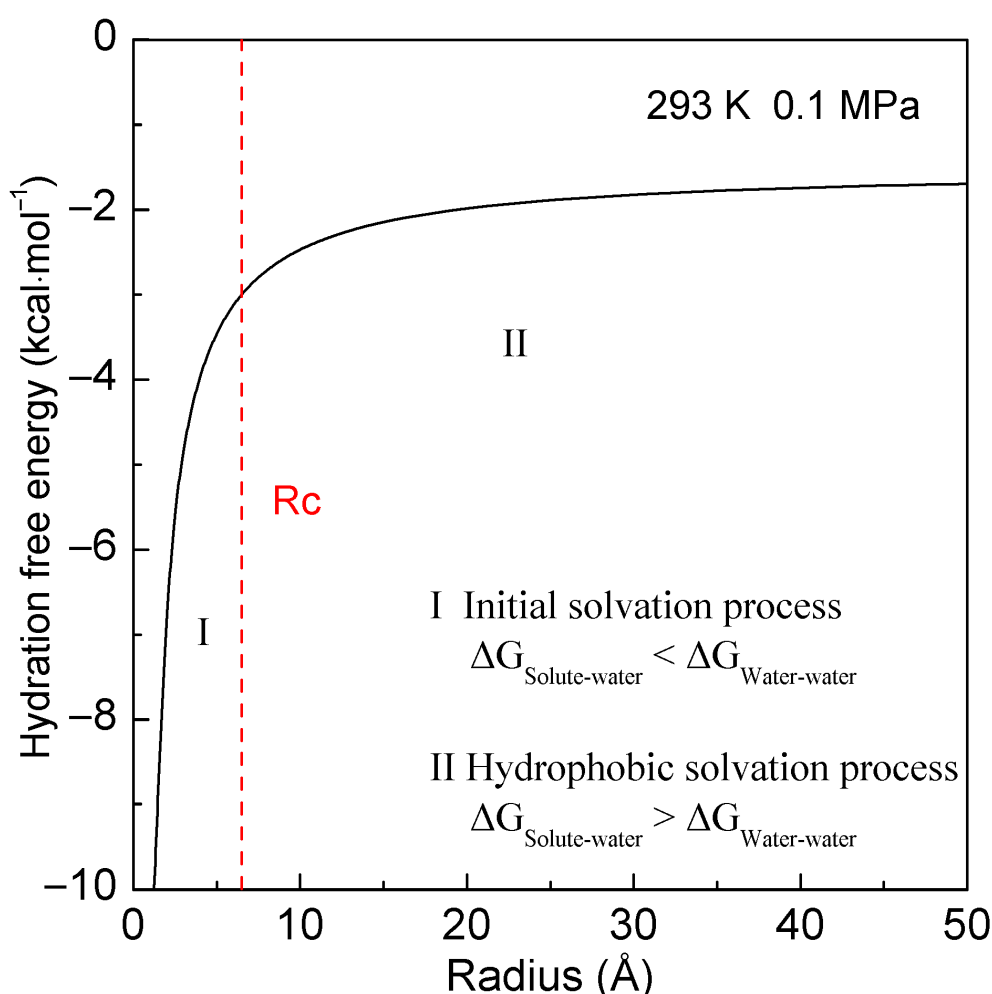


Figure 1. Hydration-free energy at 293 K and 0.1 MPa. With increasing the solute size, it is divided into the initial and hydrophobic solvation processes. The critical radius of solute (R_c) is shown by the dashed line.

The dissolved solutes mainly affect the hydrogen-bonded networks of interfacial water. Owing to hydrophobic interactions, they are attracted and aggregated in solutions to maximize the hydrogen bondings of water. In fact, as the solutes come into contact in solutions, this decreases the solute surfaces available for interfacial water. Therefore, during the solutes' accumulations in water, the Gibbs energy of the interfacial water may be expressed as

$$\Delta G_{\text{Interfacial water}} = \gamma \cdot \Delta G_{\text{Solute-water}} \left(\gamma = \frac{\left(\frac{\text{Surface area}}{\text{Volume}} \right)_{\text{Aggregate}}}{\left(\frac{\text{Surface area}}{\text{Volume}} \right)_{\text{Non-aggregate}}} = f\left(\frac{1}{R_{\text{Separation}}}\right) \right) \quad (6)$$

where γ is the geometric factor, which is used to reflect the changes in solute surfaces while the solutes are aggregated in water. In fact, the solutes are rarely rigid. It may be accompanied with the changes in volume while they are accumulated in water. Therefore, γ is generally given as the above equation, where $R_{\text{Separation}}$ is the separation between the solutes. In our recent work [34], when the solute surfaces came into contact, the distances between them were termed as R_H (hydrophobic radius).

While the solutes are accumulated in water, they are divided into H1w and H2s hydrophobic solvation processes [34]. In the H1w hydrophobic process, the distances between the solutes are larger than R_H , or γ is 1 [34]. In other words, no accumulation

of solute surfaces may be expected, and water molecules are found between the solutes. However, as the solutes come into contact in the H2s hydrophobic process, the distances between the solutes are less than R_H , or γ is less than 1 ($\gamma < 1$). Additionally, while the solutes are aggregated in water, a dewetting transition process, similar to the liquid–gas phase transition, may be observed. From our recent study [34], dewetting is closely related to the H2s hydrophobic process, in which a single water layer between solutes may be expelled into bulk water, and the solute surfaces come into contact in solutions.

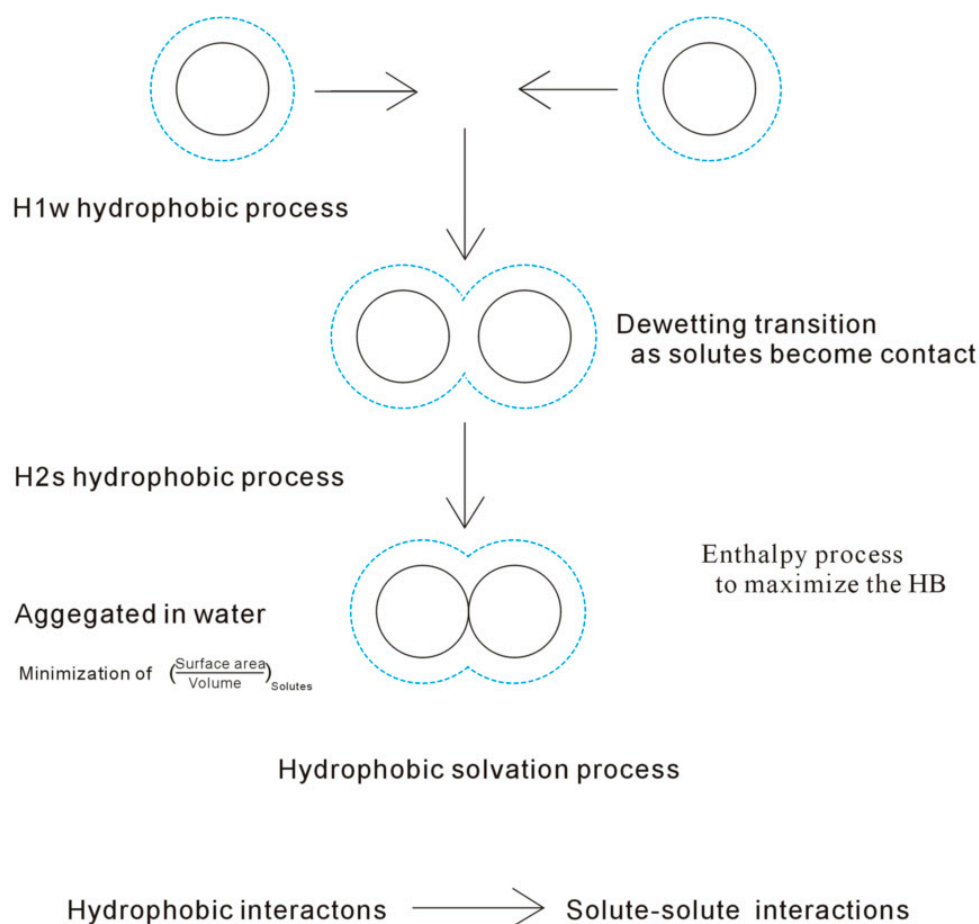


Figure 2. The mechanism of hydrophobic effects. The solutes mainly affect the structure of interfacial water, which is drawn by the dashed line. To maximize the hydrogen bondings of water, the dissolved solutes are attracted and tend to be aggregated in solution. During their association in water, they are divided into H1w and H2s processes. With decreasing the separation between solutes, the solute–solute interactions become important, especially in the H2s process.

Additionally, based on our recent work [36], various directional natures are expected in the H1w and H2s processes. In the H1w hydrophobic process ($>R_H$), the solutes tend to approach the specific direction with the lowest energy barrier, in which less water molecules are expelled. However, as the solutes come into contact in the H2s process ($<R_H$), this decreases the solute surfaces available for interfacial water. The Gibbs-free energy of the interfacial water is proportional to the surface area to volume ratios of the solutes. Therefore, the solutes are expected to be aggregated in the specific direction to minimize the surface area to volume ratio. These may be used to understand the mechanism of molecular recognition, especially the specificity of molecular recognition.

In the hydrophobic solvation process, the solutes are expected to be aggregated to maximize the hydrogen bondings of water. While the solutes are associated in solutions, the interfacial water molecules in the region between the solutes may be expelled into bulk water, which may be closely related to the hydrophobic interactions. In other words,

the strengths of the hydrophobic interactions may be dependent on the water molecular numbers transformed from the interfacial to bulk water. From this, it is expressed as

$$\begin{aligned}\Delta G_{\text{Hydrophobicity}} &= \gamma \cdot \frac{8 \cdot \Delta G_{\text{DDAA}} \cdot r_{\text{H}_2\text{O}}}{R_{\text{Solute-solute}}} - \sum_{i=1}^m \frac{8 \cdot \Delta G_{\text{DDAA}} \cdot r_{\text{H}_2\text{O}}}{R_i} \\ &= n_{\text{Interfacial} \rightarrow \text{bulk water}} \cdot \Delta G_{\text{DDAA}}\end{aligned}\quad (7)$$

where the first (second) item is Gibbs energy of interfacial water after (before) solutes are aggregated in water. Additionally, $n_{\text{Interfacial} \rightarrow \text{bulk water}}$ is the water molecular number changed from interfacial to bulk water during the solutes' associations in solutions. Based on Equation (7), hydrophobic interactions may be related to not only the solute size and shape but also temperature, pressure, and dissolved salt, etc.

In general, hydrophobic effects mean the tendency of non-polar molecules (or molecular surfaces) to be aggregated in water. According to our recent studies [33,39], they are reasonably described as the tendency for minimization of the ratio of surface area to the volume of the solutes to maximize the hydrogen bondings of water. This is because the dissolved solute mainly affects the structure of the interfacial water, and the hydrogen bondings of the interfacial water are weaker than the bulk water. Additionally, owing to the hydrophobic interactions, the solutes are attracted and tend to be aggregated in water. When decreasing the separation between solutes, the direct solute-solute interactions become stronger, especially when the solute surfaces come into contact in the H₂s hydrophobic solvation process (Figure 2), which undoubtedly affects the dissolved behaviors of the solutes. Therefore, regarding the association of the solutes in water, it is ascribed to be driven by hydrophobic interactions.

3. Structure-Thermodynamics Relationship during Folding (“Beacon” Model)

In the process of protein folding, the interactions may include (1) hydrogen bondings, (2) hydrophobic interactions, (3) van der Waals interactions, (4) electrostatic interactions, etc. [50]. The process of protein folding is closely related to the forces exerted on the atoms of the amino acid chain, and the native folded structure is reasonably ascribed to the combined effects of the above interactions. In fact, these forces arise from the interactions with other parts of the protein itself (interactions within protein), as well as those related to water (water-induced forces). Therefore, various interactions may play a different role in the process of protein folding, which is related to the characteristics of each force.

In thermodynamics, Gibbs-free energy (ΔG) may be used to investigate whether a process is likely to occur, which is related to changes in enthalpy (ΔH) and entropy (ΔS),

$$\Delta G = \Delta H - T \cdot \Delta S \quad (8)$$

where ΔH measures the total energy of a thermodynamic system, ΔS is a measure of the number of microscopic states of a system and is commonly used as a metric for disorder. When ΔG is less than zero, the process is spontaneous. Therefore, the process may be dominated by the changes in enthalpy (enthalpic process) or entropy (entropic process), respectively.

Thermodynamically, as a protein is folded into the native three-dimensional structure, the total Gibbs-free energy is reasonably expressed as follows,

$$\begin{aligned}\Delta G &= \Delta G_{\text{Water-water}} + \Delta G_{\text{Protein-water}} + \Delta G_{\text{Protein-protein}} \\ &= \Delta G_{\text{Hydration}} + (\Delta H_{\text{Protein-protein}} - T \cdot \Delta S_{\text{Protein-protein}})\end{aligned}\quad (9)$$

where $\Delta G_{\text{Hydration}}$ is hydration-free energy, $\Delta H_{\text{Protein-protein}}$ are the changes in enthalpy related to the interactions between the atoms (or residues) of protein, and $\Delta S_{\text{Protein-protein}}$ is the conformational entropy, which means the loss of entropy during the protein folding from the extended chain to the compact native structure.

Protein stability refers to the energy difference between the folded and unfolded state of the protein in the solution, which determines whether a protein will be its native folded conformation or a denatured (unfolded or extended) state. Remarkably, the free energy difference between these states is usually between 20 and 60 kJ·mol⁻¹ [51], which is of the magnitude of one to four hydrogen bondings. Therefore, folded proteins are only marginally stable [52]. This means that the unfolding–folding processes involve only the formations and break-ups of weak, non-covalent interactions. To investigate the mechanism of protein folding, it is important to correctly evaluate the dominant contributions among the many energy terms related to the free energy of protein folding.

The unfolded states of a protein possess an enormous number of degrees of freedom. During protein folding, accompanied with the decrease in conformational flexibility, this leads to an enormous loss of conformational entropy, which is a measure of the degrees of conformational freedom available to a protein (or part thereof). The loss of conformational entropy is a main destabilizing force in the thermodynamics of protein folding [53]. Experimental estimates of the entropy change on folding, and ΔS_{fold} gives a comparable range of -2.6 to -9 cal·mol⁻¹·K⁻¹ per residue [54–56]. In fact, both the backbone and side-chain of each residue in a protein will have their freedom of motion restricted in the final folded structure. In Baxa et al.'s work [57], the loss of conformational entropy is largely due to the loss of backbone entropy. From Towse et al.'s study [58], the side chain entropy shows wider distributions on increasing side chain lengths or bulks.

Under physiological conditions, a polypeptide chain of a protein is spontaneously folded to a native three-dimensional structure. In thermodynamics, this means that the total Gibbs energy (ΔG_{Total}), related to $\Delta G_{Hydration}$ and $\Delta G_{Protein-protein}$, may be less than zero. In our recent study [39], hydrophobicity may have been closely related to the enthalpy-entropy compensation (EEC), which meant if ΔH and ΔS for the particular reaction were changing in one direction (either increase or decrease), their changes being transformed into ΔG were mutually compensated, and there was little change in the value of ΔG . Of course, this was due to the competition between the interfacial and bulk water [33]. To attain the native structure of a protein, it is essential to outweigh the loss of conformational entropy arising from folding. Therefore, it is necessary to obtain enthalpy enough to overcome the entropic penalty. This means that the protein folding is facilitated by maximizing the enthalpy, and it is reasonably described to be an enthalpic process.

There is considerable evidence that hydrophobic interactions must play a major role in protein folding [59]. According to Pace et al.'s study [60], the average contribution of hydrophobic interactions to a protein's stability is 60%. The importance of methyl groups in modulating biological activity for small molecules is well documented. Experimentally, the benefit of burying a solvent-exposed methyl group on a ligand into a hydrophobic pocket of a protein is about 0.7 kcal·mol⁻¹, or a 3.2-fold increase in binding constant per methyl group ($-CH_3$). In Pace et al.'s work [60], burying a $-CH_2$ group on folding may contribute, on average, 1.1 ± 0.5 kcal·mol⁻¹ to protein stability. Additionally, based on Pace et al. studies [60,61] of 151 hydrogen bonding variants in 15 proteins, these mean that hydrogen bonding contributes to the protein stability about 40%, the net contribution of hydrogen bonding to overall protein stability is 1.1 kcal·mol⁻¹, and is largely independent of the size of the protein [62]. Therefore, it can be derived that protein folding may be primarily affected by hydrophobic interactions and hydrogen bondings.

In general, as a protein folds, 81% of the nonpolar side chains, 70% of the peptide groups, 63% of the polar side chains, and 54% of the charged side chains are buried in the interior of the protein out of contact with water. Based on our recent studies [33,39], hydrophobic interactions play important roles in protein folding, especially in the initial (structural collapse) and final folding stages. To maximize the hydrogen bondings of water, a protein may be folded from the extended conformation into a three-dimensional structure. In kinetic experiments, an initial collapse in the size of the polypeptide chain is usually observed upon changing solvent conditions from being denaturing to renaturing [63–66].

Additionally, owing to hydrophobic interactions, this leads to the structural collapse of a protein in water. It may be transformed from an extended coil to a more compact, globular structure in order to minimize the surface area/volume ratio of protein. By decreasing the separations between the atoms (or residues) of proteins, the direct interactions between residues of proteins (interactions within protein) become stronger, which may affect the folding process of proteins in water. Therefore, it is found that protein folding may reasonably be regarded to be driven by hydrophobic interactions.

In thermodynamics, the loss of conformational entropy is a major destabilizing force in the process of protein folding. It has been considered that the entropic penalty can be compensated for by an energy gain through the formation of intramolecular hydrogen bonds in proteins [67,68]. Other intramolecular interactions, such as salt bridging, van der Waals attraction, etc., may also stabilize the native structures of proteins [69]. In fact, approximately two-thirds of the intramolecular hydrogen bonds are within repetitive elements of secondary structure [70] of the folded proteins. Based on the experimental studies [71,72], the hydrogen bond energy of the hydrogen bonds between the N-H groups and C=O groups of the main chains in the secondary structures is about $-3.47 \text{ kcal}\cdot\text{mol}^{-1}$. Therefore, the secondary structures of a protein, such as α -helices and β -sheets, are stabilized by the formation of intramolecular hydrogen bonds between the acceptor CO and the donor NH groups [73]. It is derived that the folding process may be guided by the formation of intramolecular hydrogen bondings within the secondary structures of protein. In fact, a backbone-based theory of protein folding is proposed [70,74], which is based on the energetics of backbone hydrogen bonds dominating the folding process. Of course, it is necessary that the hydrogen bonds between water and the peptide NH and CO groups must be broken before peptide hydrogen bonds are formed. This is undoubtedly related to the structural collapses of proteins driven by hydrophobic interactions.

Recently, high-resolution experimental methods sensitive to population distributions have been developed and applied to investigate the structural characteristics at different stages of folding reactions. Significant conformational heterogeneity is found during protein folding, including the unfolded state, collapsed intermediate states, and even the native state [75–77]. In fact, the heterogeneity in protein folding and unfolding reactions may be closely related to the various kinds of physicochemical interactions between various structural elements of a protein and between a protein and solvent [77].

From the discussion on the relationship between structure and thermodynamics during protein folding, a protein may be reasonably divided into the folding key (FK) and the remaining non-folding key (non-FK) regions, respectively. The $\Delta G_{\text{Protein-protein}}$ is expressed as

$$\begin{aligned} \Delta G_{\text{Protein-protein}} &= \Delta G_{\text{FK Regions}} + \Delta G_{\text{non-FK Regions}} + \Delta G_{\text{Between FK and non-FK regions}} \\ &= \sum_{i=1}^n \Delta G_{\text{FK}_i} + \Delta G_{\text{non-FK Regions}} + \Delta G_{\text{Between FK and non-FK regions}} \end{aligned} \quad (10)$$

where $\Delta G_{\text{FK regions}}$, $\Delta G_{\text{non-FK regions}}$, and $\Delta G_{\text{Between FK and non-FK regions}}$ mean the Gibbs energies within the FK regions, non-FK regions, and between the FK and non-FK regions, and n is the number of FK regions.

Intramolecular hydrogen bondings are expected to form within the FK regions of a protein, such as the α -helix and β -sheet of a secondary structure of a protein. Thermodynamically, protein folding is an enthalpic process. Therefore, folding may be guided by the strong intramolecular hydrogen bondings within the FK regions. Additionally, due to the formation of the intramolecular hydrogen bondings in the FK regions, a specific conformation may be expected for the FK regions in the folding process. This means that a higher structural order may be expected for the FK regions, which is related to the formation of intramolecular hydrogen bondings within the regions, in comparison with the non-FK regions. In addition, after the specific conformations are formed, they are usually preserved in the remaining folding time. In the process of protein folding, the FK regions

may be regarded as the necessary pathways where folding tends to pass through so that the protein is folded to form the final three-dimensional native structure.

In comparison with the FK regions, other interactions are expected in the non-FK regions of a protein, such as van der Waals forces, salt bridges, hydrogen bondings between side chains of a protein, etc. They may play an important role in the final folding stage. Different from the specific conformation of FK regions, dynamic conformation is found for the non-FK regions in the folding process. In other words, evident conformational changes and fluctuations, such as in the distances, volumes, etc., may be expected for the non-FK regions of a protein during folding. Additionally, after the specific conformations of the FK regions are formed during folding, this may lead to decreases in the conformational changes and fluctuations related to the non-FK regions.

As a protein is folded from the extended chain to the native three-dimensional conformation in water, the protein tends to be engaged into the conformation with lower Gibbs energy to become more thermodynamically stable. In thermodynamics, it is facilitated by maximizing enthalpy to overcome the loss of entropy. During folding, this is related to not only folding time but also the spatial distributions of intramolecular interactions within a protein, especially FK regions. During folding, the Gibbs energy of $\Delta G_{protein-protein}$ at time t may be reasonably expressed as

$$\begin{aligned}\Delta G_{Protein-protein}(t) &= f\left(t, \sum_{i=1}^m x_i\right) = f\left(\sum_{i=1}^m x_i(t)\right) \\ &= f(FK\ region(t), non - FK\ region(t))\end{aligned}\quad (11)$$

where m is the atomic number of the protein, and $x_i(t)$ is the position of atom i in the protein. Therefore, at any folding time t , the protein is reasonably divided into the FK and non-FK regions.

From the above, folding may be divided into the following stages, such as the initial structural collapse, the folding of FKs related to intramolecular hydrogen bondings, and the last folding stage of the non-FKs. If several FKs exist, sequential formations (sequential folding) may be expected for them during folding, which are expressed as FK1, FK2, ..., etc. Additionally, the corresponding local basins are expected in the FEL in the folding pathway (Figure 3). In addition, after the specific conformations are formed within the FK regions, they tend to be preserved in the remaining folding time. Therefore, the sequential formations of specific conformations related to intramolecular hydrogen bondings, such as the α -helices and β -sheets of secondary structures of a protein, are expected during folding.

Following the formation of the specific conformations related to the FK regions, it is engaged to the final folding stage. To attain the lowest Gibbs energy of a native three-dimensional structure during this stage, the structure may be modulated by various forces within the non-FK regions and those between the FK and non-FK regions, such as van der Waals forces, salt bridges, hydrogen bondings between side chains, and hydrophobic interactions. Generally, it is necessary for the protein to "breathe" so that water molecules in the interior may be repelled into the bulk. Of course, this is related to the dewetting transition of the H2s hydrophobic process.

Based on the experimental measurements [78–80], the molten globules (MG) [81,82] may be found in a specific region of a protein during folding. MGs are compact, partially folded conformations of proteins that have near-native compactness properties, substantial secondary structures, little detectable tertiary structures, and increased solvent-exposed hydrophobic surface areas relative to their native states, which are thought to be common intermediates in protein folding [83]. Later, there was a growing realization that the "dry" molten globule (DMG) [84] is another distinct state along a graduated MG spectrum. The defining difference between a DMG and a conventional MG is that the water has been squeezed from the core of a DMG [85]. From this work, a protein is divided into FK and non-FK regions in the folding process. Due to the intramolecular hydrogen bondings within the FK regions, higher structural orders may be expected for FK regions than the rest of the

protein (non-FK regions) during folding. Therefore, this may be applied to understand the formation of MGs and DMGs in the folding process.

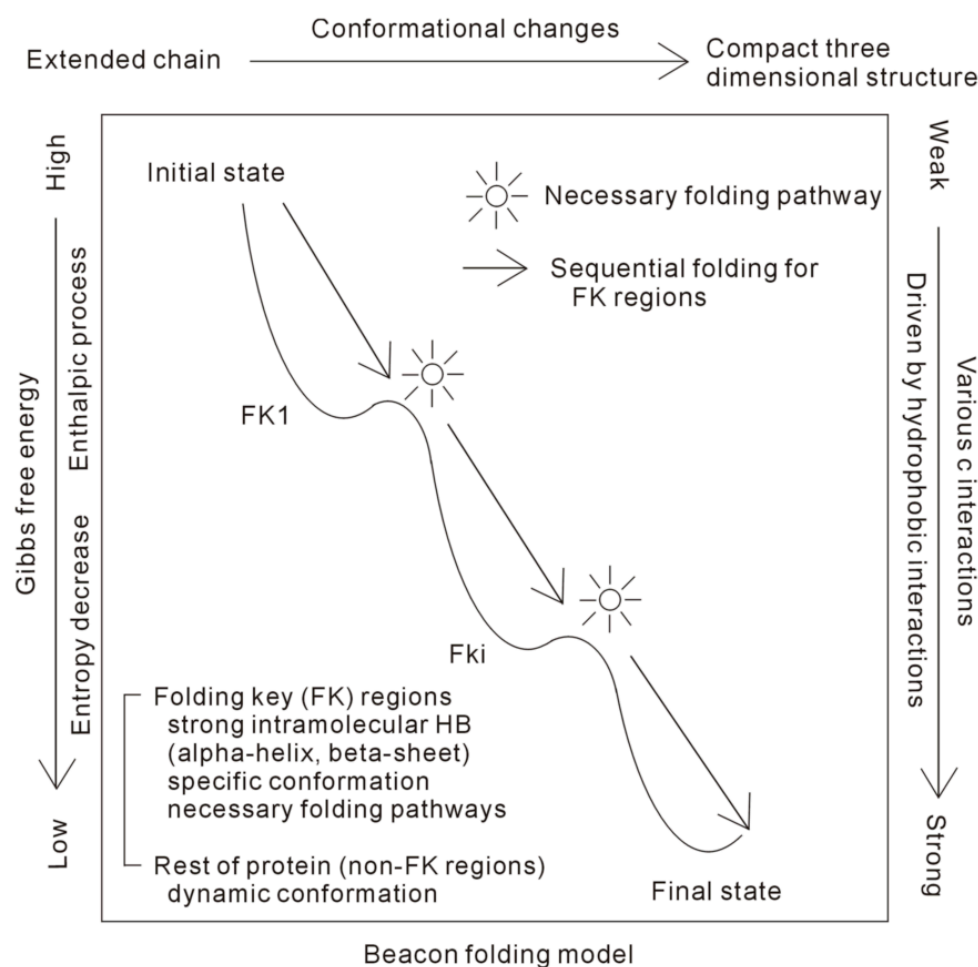


Figure 3. The “beacon” model of protein folding. The protein is divided into the FK and non-FK regions. During folding, the specific conformation is expected for FK, which is regarded as the “beacon” in protein folding. If there are many FK regions, the sequential folding pathway may be expected for them. Additionally, dynamic conformations are expected for the non-FK regions in the folding process.

According to the thermodynamic analysis, a protein is reasonably divided into FK and non-FK regions. In the folding process, specific conformations may be expected for FK regions. These are different from non-FK regions, in which dynamic conformations may be found during folding. From these, various conformational changes are expected for FK and non-FK regions during folding. Therefore, it seems that there exist multiple folding pathways described as free-energy landscapes [19]. Additionally, the conformational changes may be related to not only the regions of protein (spatial distribution) but also the folding time (time dependence). In addition, folding is guided by the intramolecular hydrogen bondings within FK regions. During folding, sequential formations may be expected for the FK regions. In other words, evident conformational changes may be found at different folding times. It seems that there exists a single folding pathway for FK regions. In fact, these may be utilized to understand the strong debate on the folding pathways of a protein in water, either with single or multiple pathways.

In thermodynamics, protein folding is an enthalpic process, which is primarily related to hydrophobic interactions and intramolecular hydrogen bondings. In fact, hydrophobic interactions may play an important role in the whole process of protein folding, especially

in the initial folding stage. A protein is divided into FK and non-FK regions, and various conformational changes are expected for them. Protein folding is guided by the strong intramolecular hydrogen bondings within FK regions, such as the secondary structures of a protein. In the folding process, sequential folding may be expected for the FK regions, which may be regarded as the “beacons”, where the folding might tend to pass through (Figure 4). In the final folding stage, to attain the lowest Gibbs energy of the native state, it is related to the combined effects of various forces related to the non-FK regions, such as van der Waals forces, hydrogen bondings between side chains, salt bridges, and hydrophobic interactions.

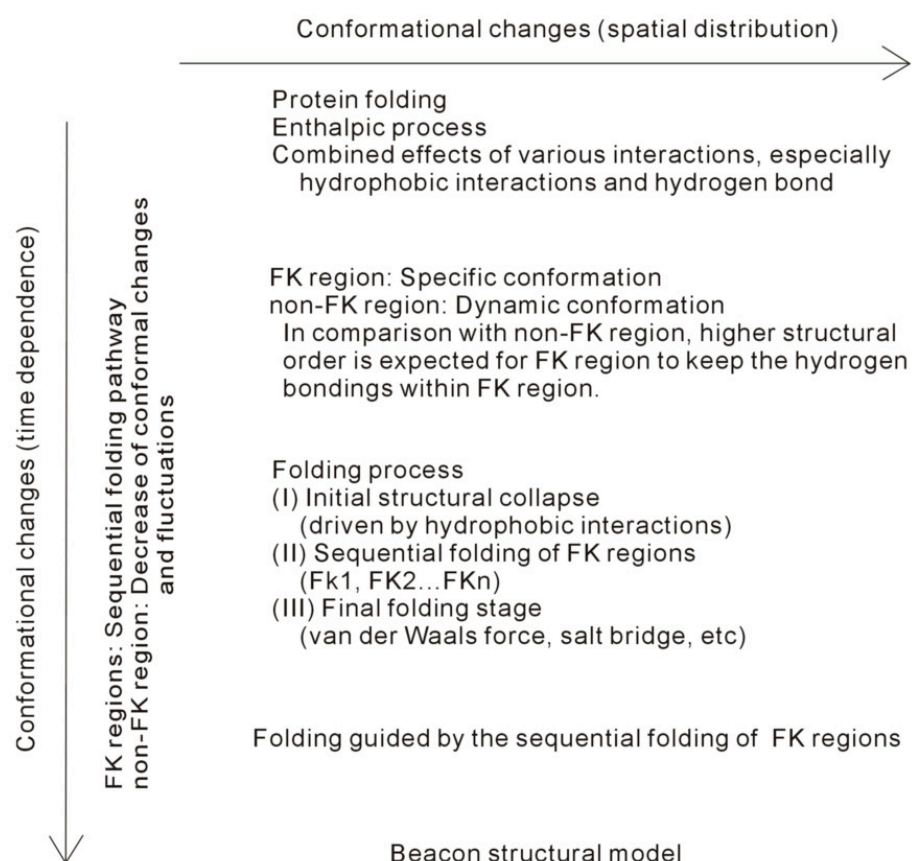


Figure 4. Conformational changes during protein folding.

4. Application for Trp-Cage Folding

Trp-cage is a designed 20-residue protein (Asn1-Leu2-Tyr3-Ile4-Gln5-Trp6-Leu7-Lys8-Asp9-Gly10-Gly11-Pro12-Ser13-Ser14-Gly15-Arg16-Pro17-Pro18-Pro19-Ser20; PDB 1L2Y.pdb) [86]. NMR spectroscopy [86] and X-ray crystallography [87] have been applied to determine the structure of this mini-protein. It contains an α -helix (Leu2-Lys9), a 3^{10} -helix (Gly11-Ser14), and a polyproline II helix (Pro17-Pro19) (Figure 5). The native structure is stabilized by hydrogen bonding between the carbonyl oxygens of Arg16 and He1 of the TRP6 and the salt bridges formed between the Asp9 and Arg16 residues. Additionally, it also consists of a hydrophobic cage formed by the packing of Tyr and Trp6 residues around the Gly11, Pro12, Pro18, and Pro19 residues, which is crucial for maintaining the integrity of the structure [86,88,89].

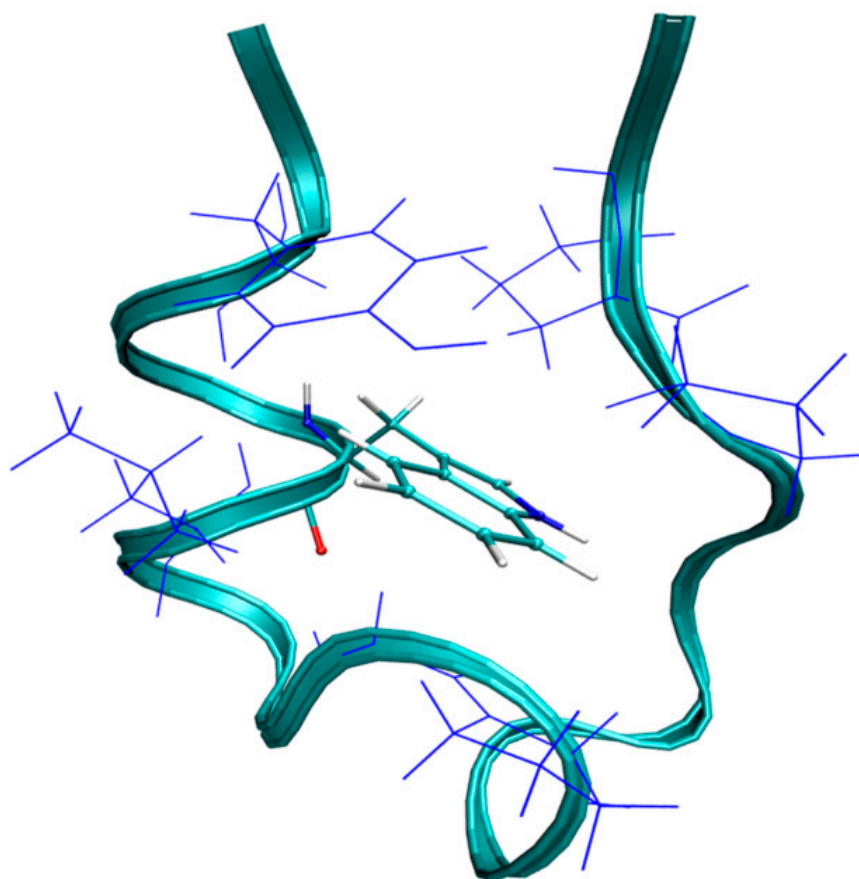


Figure 5. The structure of the TRP-cage. The backbone is shown by a ribbon representation. The hydrophobic core is shown, in which the central residue Trp6 is surrounded by Tyr3, Leu7, Gly11, Pro12, Pro18, and Pro19.

Due to its structural simplicity and rapid folding dynamics, Trp-cage has been extensively studied both experimentally [90–98] and computationally [99–114] in order to understand the folding mechanisms of the protein in water. To date, the folding mechanisms of the TRP-cages still remain elusive. In some experimental [90–93] and simulation [99–102] works, these meant that the folding kinetics may have been two-state. However, from the other experimental [94–98] and simulation [103–106] studies, they suggested that its kinetics were not two-state. Additionally, some experimental works indicated the presence of well-defined on-pathway intermediates [94–98], which was in contrast to the suggestion that the folding was downhill [115,116].

To date, two characteristic folding pathways have been identified for the mini-protein [107–111]. In one pathway (I), the collapse of the hydrophobic core precedes the formation of the α -helix, and, in the other pathway (II), the α -helix is (partially) formed in the initial stage of folding, which is followed by the collapse of the hydrophobic core. Some MD simulations of the TRP-cage folding [103,105,111] suggest that pathway I is dominant, or even exclusive, as in experiment [95]. However, the others [102] mean that pathway II prevails, similar to what was observed experimentally [93,98]. Additionally, folding pathways may be dependent on temperatures. Based on MD simulations, pathway I may be observed at room and nearby temperatures [103,105,111], and pathway II is found at melting and higher temperatures [102].

To understand the folding mechanism of the Trp-cage in water, MD simulations were carried out. Compared to explicit solvents, the implicit solvent models led to folding rates faster than the experimental values, but the relative rates of formation of the secondary structural elements were comparable to the values observed experimentally [107]. In fact, several implicit solvent models have been reported to yield correct refolded structures of

the Trp-cage [117,118]. It should be noted that the melting temperatures obtained in the simulations were typically much higher than the experimental value $T_m \approx 315$ K [86,90], both in the simulations with an explicit solvent (440 K [99] and 455 K [112]) and with an implicit solvent (400 K [113], 450 K [100], and 468 K [102]). It is noted that the specific explicit solvent's force field combination (AMBER99SB or AMBER99SB-ILDN with TIP3P water) provides realistic accuracy in predicting the Trp-cage melting temperature [101]. In this work, the AMBER99SB-ILDN force field with a generalized Born-based implicit solvent model was used to investigate the folding mechanism of the TRP-cage in water (see Methods). In this study, the total 2000 ns MD simulations were, respectively, conducted at 315 K, 320 K, and 350 K.

Based on the "beacon" model of folding, the protein may be divided into the FK and remaining non-FK regions. For the Trp-cage, the FK is the α -helix (Leu2-Lys9), and the rest of the protein is related to the non-FK region. Hydrophobic interactions play an important role in the whole folding process, especially in the initial structural collapse. Additionally, protein folding is an enthalpic process. During folding, it is necessary to form the α -helix, which may be related to the intramolecular hydrogen bondings within the α -helix's structure. It may be regarded as the necessary pathway during folding. In addition, this is reflected in the local free-energy basin in the FEL. Regarding the non-FK region, dynamic conformations are expected in the folding process. In the final folding stage, the mini-protein is folded into the native structure, which is due to the combined effects of various interactions, such as van der Waals interactions, salt bridges, etc. Therefore, the intermediate is expected in the folding process of the TRP-cage in water.

To monitor the conformational change in the Trp-cage during folding at 315 K, the root-mean square deviation (RMSD) relative to the NMR structure [86] (PDB: 1L2Y) was determined (Figure 6b). Based on the simulations, the backbone RMSD decreased to 0.35 nm at about 68 ns and started fluctuating up to 0.8 nm. Once the structure fell to 0.3 nm for about 590 ns, it stayed there for the remainder of the simulation. Additionally, it was found that the lowest backbone RMSD value was 0.206 nm. Therefore, the native state was the most-stable state sampled during the simulation. To estimate the effective compactness of the Trp-cage mini-protein, the radius of gyration (R_g) of the backbone atoms of the mini-protein was also calculated. From Figure 6a, the MD trajectory visited both extended conformations ($R_g \geq 0.95$ nm) and compact unfolded structures ($R_g \leq 0.75$ nm) many times before it doled. The value of R_g may remain stable after the native structure is reached.

Regarding the FK region of the Trp-cage, it is related to the α -helix. In comparison to the NMR structure, the calculated RMSD of the α -helix decreases from 0.6 nm to 0.35 nm at about 70 ns, and then it increases up to 0.5 nm at 200 ns (Figure 6c). When the $\text{RMSD}_{\text{FK}(\alpha\text{-helix})}$ decreases to 0.35 nm at 500 ns, it remains stable. Compared with the $\text{RMSD}_{\text{FK}(\alpha\text{-helix})}$, evident changes and fluctuations of the $\text{RMSD}_{\text{non-FK}}$ may be found for the non-FK regions of the Trp-cage before 590 ns (Figure 6d). These may be closely related to the conformational changes in the protein, as shown in $\text{RMSD}_{\text{Backbone}}$ (Figure 6b). Additionally, evident decreases can be observed for the changes and fluctuations of the $\text{RMSD}_{\text{non-FK}}$ after 590 ns (Figure 6d). These mean that the protein is folded into the native three-dimensional structure.

To understand the conformational changes in the ternary structure of the mini-protein, the RMSD of the 3^{10} Helix of the backbone is determined during folding at 315 K. In comparison with the NMR structure of the 3^{10} Helix, the $\text{RMSD}_{3^{10}\text{ Helix}}$ decreases to 0.27 nm at 520 ns and keeps stable in the remaining simulation time (Figure 6e). Therefore, the formation of the 3^{10} Helix α -helix is later than the formation of the secondary structure (α -Helix) of the Trp-cage. From the MD simulations, sequential folding may be found for the Trp-cage in water.

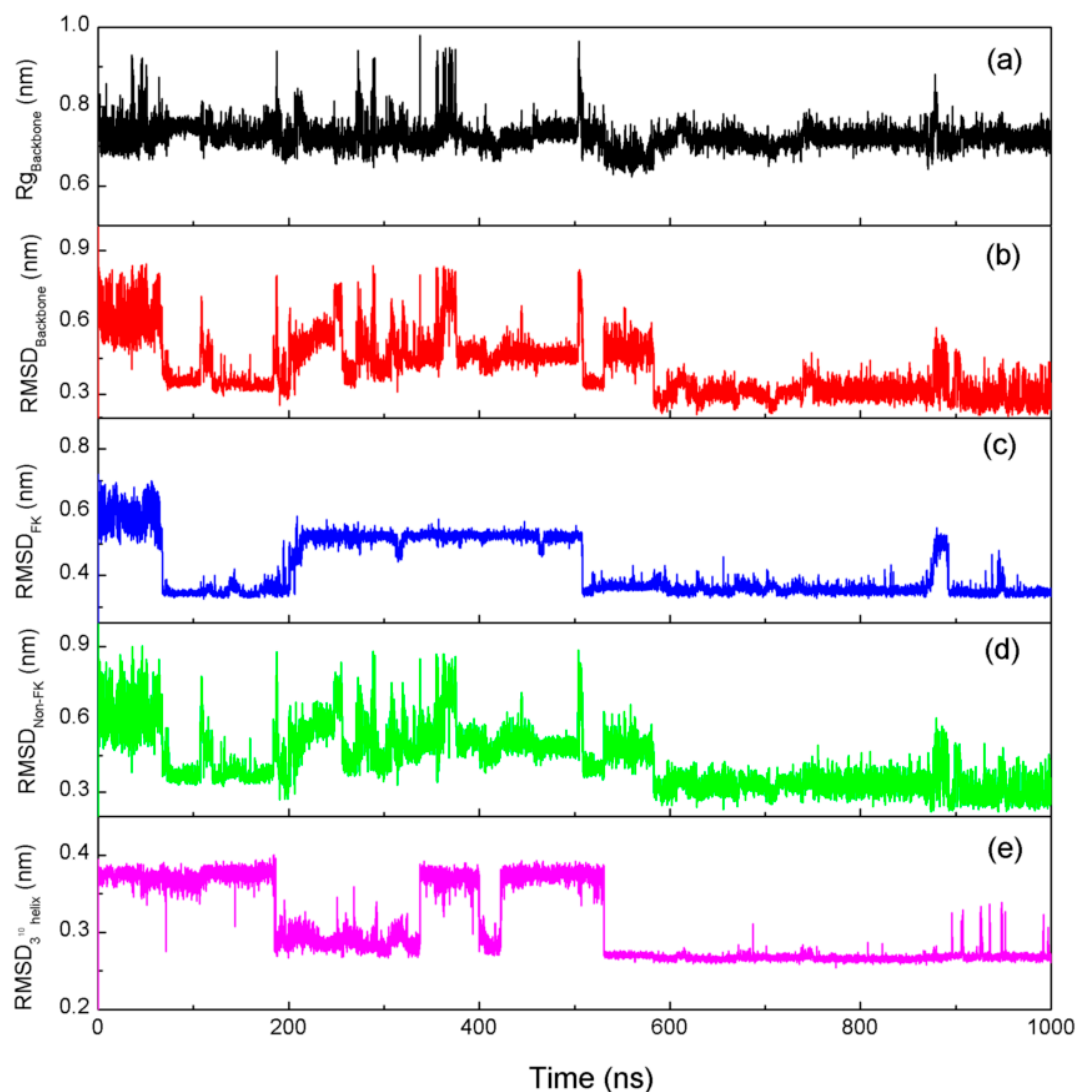


Figure 6. The backbone Rg (a) and RMSD (b) during folding at 315 K. To investigate the conformation changes in the folding process, the RMSD of the backbone related to the FK (α -helix) (c) and non-FK (d) regions of the Trp-cage and the 3^{10} -helix (e) are also calculated.

Additionally, other factors [86,88,89] may also play a very crucial role to keep the native structure of the Trp-cage stable. These factors are (I) the formation of hydrogen bonds (Figure 7a) between the H ϵ 1 of the Trp6 residue and the backbone carbonyl (C=O) of the Arg16 residue, (II) the salt bridge (Figure 7b) between the two residues Asp9 and Arg16, and (III) a hydrophobic core (Figure 7c) containing Tyr3, Gly11, Pro12, Pro18, and Pro19 residues surrounding the central residue Trp6. Based on the MD simulations, the corresponding distances for the center of mass are, respectively, determined for the hydrogen bond, the salt bridge, and hydrophobic core (Figure 7). Evident changes and fluctuations are found for these distances before 590 ns. Additionally, they become relatively stable after 590 ns. This is related to the formation of the native structures of a protein in water. Of course, it is necessary for the corresponding residues to approach before they are affected by these interactions. In fact, this is related to hydrophobic interactions, which lead to the minimization of the surface area to volume ratio to maximize the hydrogen bondings of water.

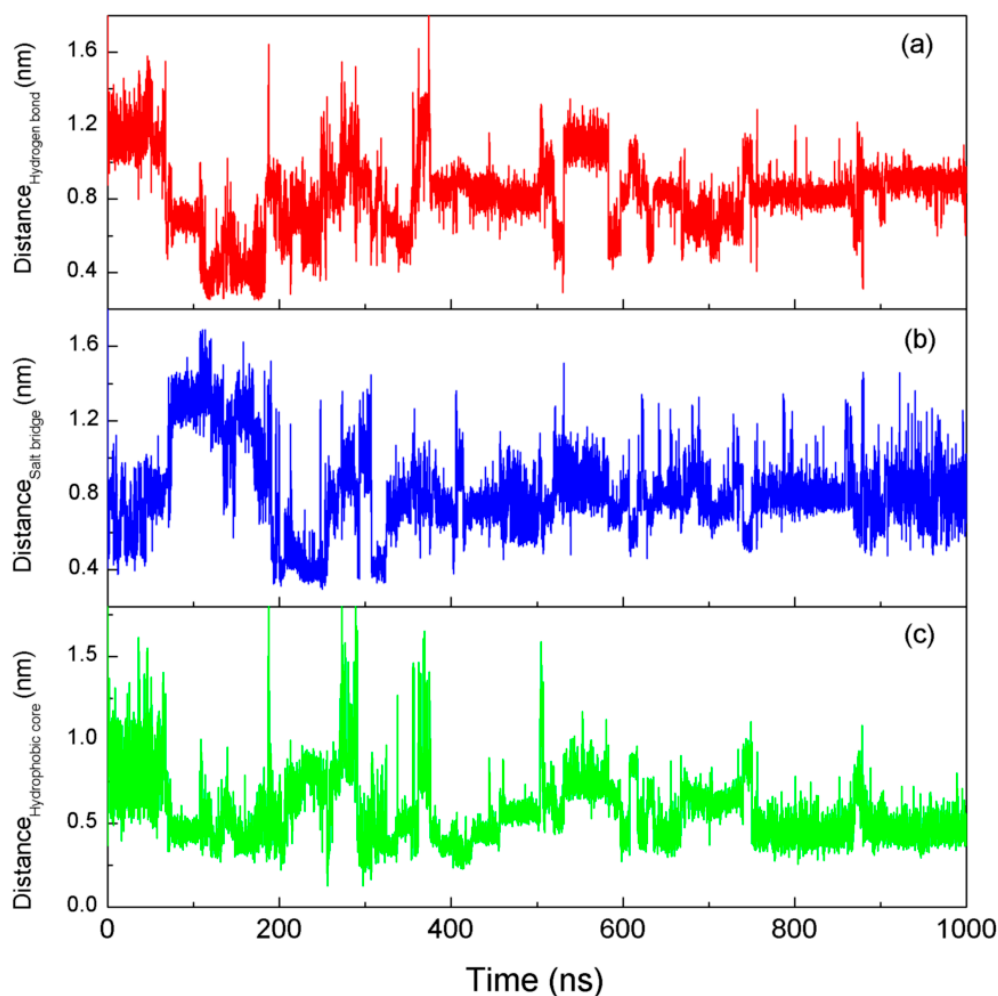


Figure 7. The distances of hydrogen bondings (a), salt bridges (b), and hydrophobic cores (c) during the simulations at 315 K.

To investigate the folding mechanisms of the Trp-cage in water, the secondary structure was analyzed by the DSSP tool in VMD [119] (Figures 8, S1 and S2). It was found that the folding pathway of the Trp-cage may have been dependent on the temperature. However, it was necessary to form the α -helix during folding. This may have been related to the intramolecular hydrogen bondings within the α -Helix. During folding at 315 K, the α -helix of the mini-protein appeared at 67.5 ns and may have persisted in the remaining simulation time (Figure 8). In other words, after the hydrogen bondings within the secondary were formed, they may have been preserved. This was slightly different from the calculated RMSD of the α -Helix (Figure 6c), which meant that the secondary structure was destabilized from 200 ns to 520 ns.

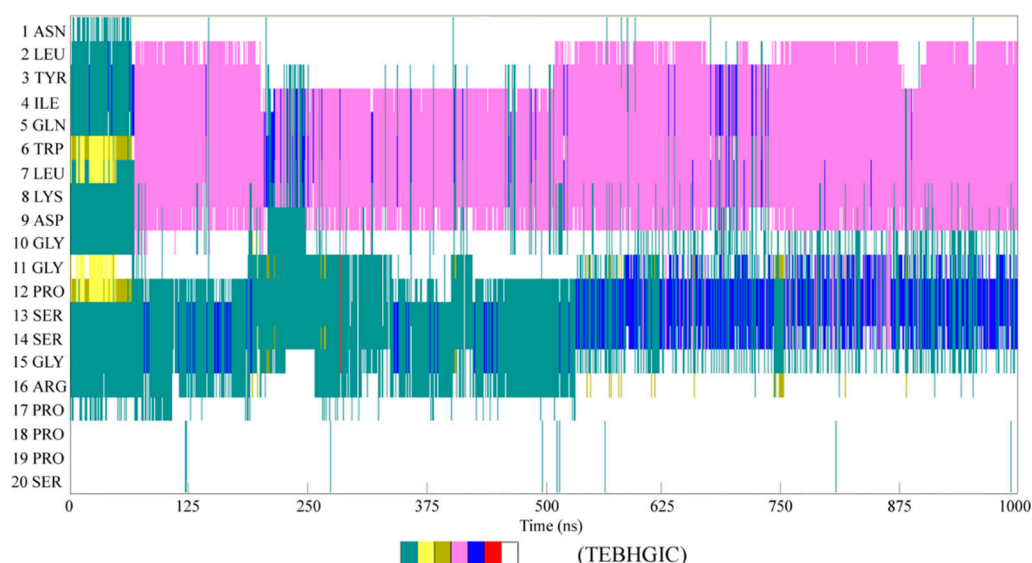


Figure 8. DSSP analysis of the Trp-cage during folding at 315 K.

Based on the above thermodynamic analysis, protein folding is an enthalpic process. Therefore, folding may be guided by the interactions between the atoms of a protein. Due to the strong intramolecular hydrogen bondings within the FK region, it is necessary to form the α -Helix before the mini-protein is folded into the native structure. In the MD simulations, three to four α -helical $i, i + 4$ main-chain hydrogen bonds are found during the formation of the secondary structure of the Trp-cage. Additionally, after the secondary is formed, it may be preserved in the remaining folding time, which may be regarded as the necessary pathway in the folding process. In the final folding stage, the folding process may be modulated by the hydrogen bondings between the side chains, salt bridges, and hydrophobic interactions to attain the native structure.

To understand the folding mechanism of the Trp-cage in water, the FEL of the miniprotein was calculated using the `g_sham` package in GROMACS v4.5.2, which was expressed as

$$\Delta G = -kT \ln \frac{P(x_i)}{P_{\max}(x)} \quad (12)$$

where ΔG was the free energy, $P(x_i)$ was the probability of being in state i , $P_{\max}(x)$ was the probability of the most observed state, k was the Boltzmann constant, and T was the temperature (315 K). Based on the simulations, the FEL may have been determined as a function of the backbone RMSD and R_g (Figure 9). During the Trp-cage folding from a linear chain to a native structure, there was a decrease in free energy. During folding, three local Gibbs energy basins are found in the FEL. Different from the classical picture with only two states, the intermediate state is found during folding, in which the specific conformation (α -Helix) is formed in the FK region, and dynamic conformation is found for the non-FK region. In fact, the presence of a metastable intermediate state has been observed in many computational and experimental works [94–98].

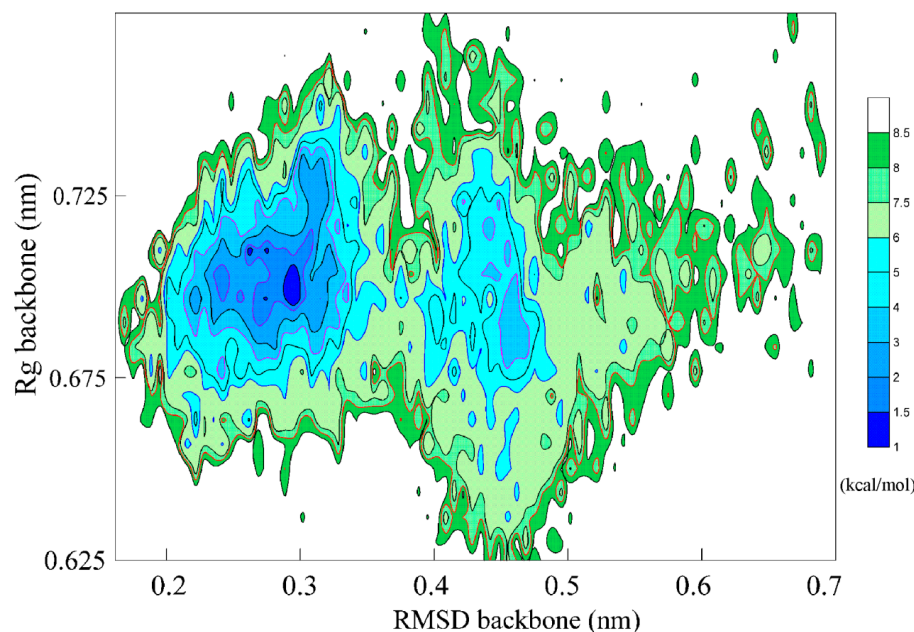


Figure 9. The FEL during folding of the Trp-cage at 315 K. The intermediate is found during folding, which is reflected on the local basin of the free energy.

From the MD simulations, the intermediate was found during folding, which was related to the various conformations of the FK and non-FK regions. From the “beacon” structural model, the specific conformation was expected for the FK region during folding, which was related to the intramolecular hydrogen bondings within the region. Additionally, after the α -Helix was formed, it was preserved in the following simulation time (Figure 10). However, different from the FK regions, dynamic conformation may have been expected for the non-FK regions of the miniprotein. This may have been reflected on the conformational changes and fluctuations during folding (Figure 10a–d). After the RMSD was equilibrated at 590 ns, the representative conformation was shown in Figure 10e. This was obtained through cluster analysis, which was conducted through the `g_cluster` tool of GROMACS. Additionally, the final conformation was also shown (Figure 10f). Additionally, with decreasing the free energy during folding, especially as the α -Helix was formed, this led to the decrease in the conformal changes in non-FK regions until the native structure was reached (Figure 6d).

Two folding pathways have been identified for the Trp-cage in water, which may have been dependent on temperature. This was mainly related to the Gibbs energies of the hydrophobic interactions and the α -helixes of the mini-protein. Hydrophobic interactions may be involved in the whole folding process of the TRP-cage in water. From Equation (7), the strength of the hydrophobic interactions may have been related to $n_{\text{Interfacial} \rightarrow \text{bulk water}}$ and ΔG_{DDAA} . In the folding process, $n_{\text{Interfacial} \rightarrow \text{bulk water}}$ was related to the ratio of the surface area to volume of the folded Trp-cage. From Raman spectroscopic studies [45,46], the increase in temperature may have led to the decrease in ΔG_{DDAA} . With increasing temperature, more interfacial water molecules ($n_{\text{Interfacial} \rightarrow \text{bulk water}}$) were necessarily transformed into bulk water so that the hydrophobic interactions may have been equivalent to the Gibbs energies of the α -helixes. Of course, this was accompanied with the smaller ratio of surface area to volume of the proteins of the TRP-cage, which was related to the structural collapse of the mini-protein in water. Further study may be necessary.

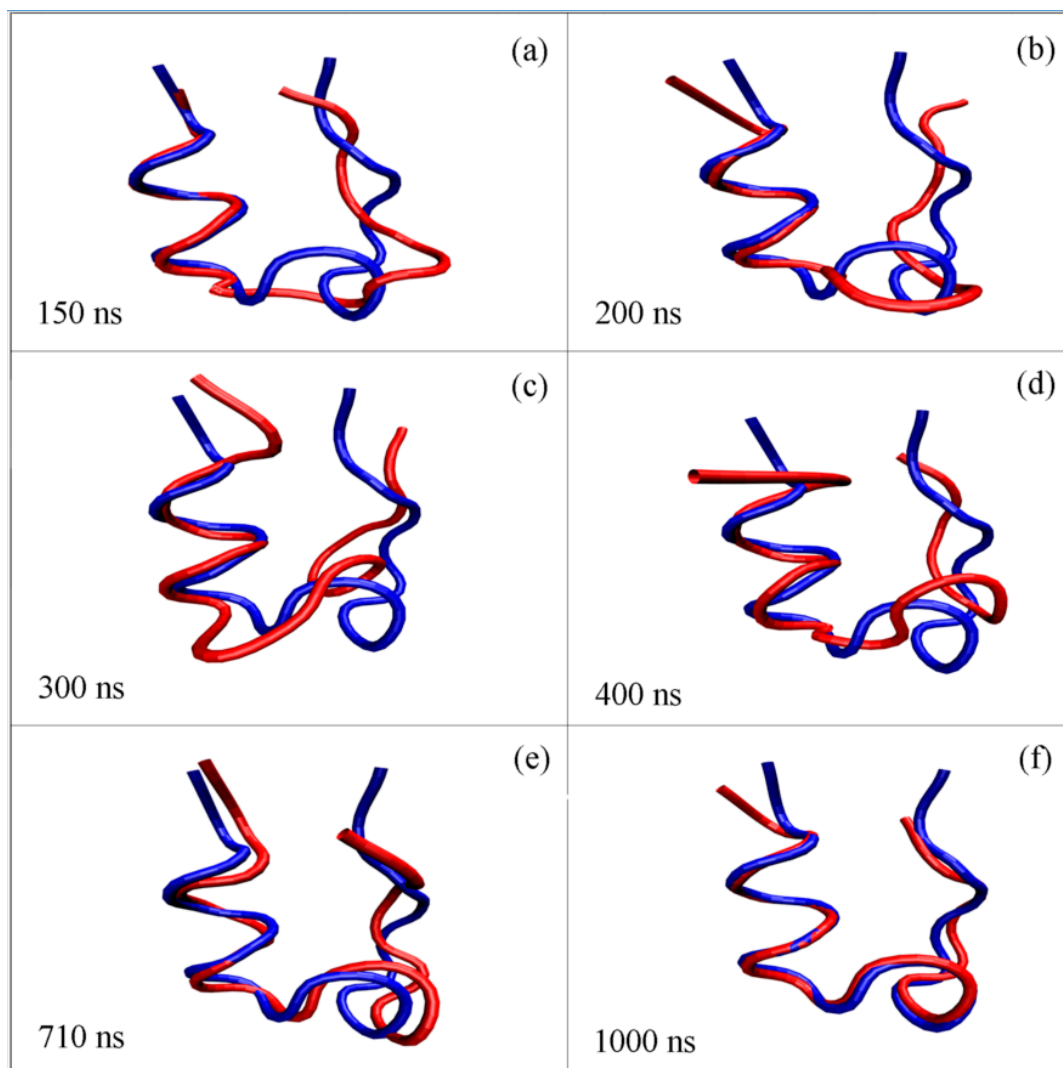


Figure 10. The conformational changes of Trp-cage during folding at 315 K. The NMR structure (1L2Y.pdb) of the Trp-cage is drawn in blue. The conformations of the mini-protein at various times are shown in (a–f), which are shown in red. Different from the specific conformation (α -Helix) of the FK region, dynamic conformations are found for the rest of the Trp-cage (non-FK region).

In this work, the total 2000 ns MD simulations were, respectively, carried out at 315 K, 320 K, and 350 K. Based on the simulations, it was necessary to form the α -helix during folding, which may have been regarded as the necessary pathway. Of course, this was related to the intramolecular hydrogen bondings within the secondary structures of the the TRP-cage. From the work, the intermediate state may have been found during the folding of the Trp-cage from an extended chain to a three-dimensional structure. It was expressed as initial state \rightarrow intermediate state \rightarrow final structure (Figure 11). This may have been reflected in the local basins of the Gibbs-free energy on the FEL of the TRP-cage in water. Additionally, this was also in accordance with the “beacon” structural model discussed above.

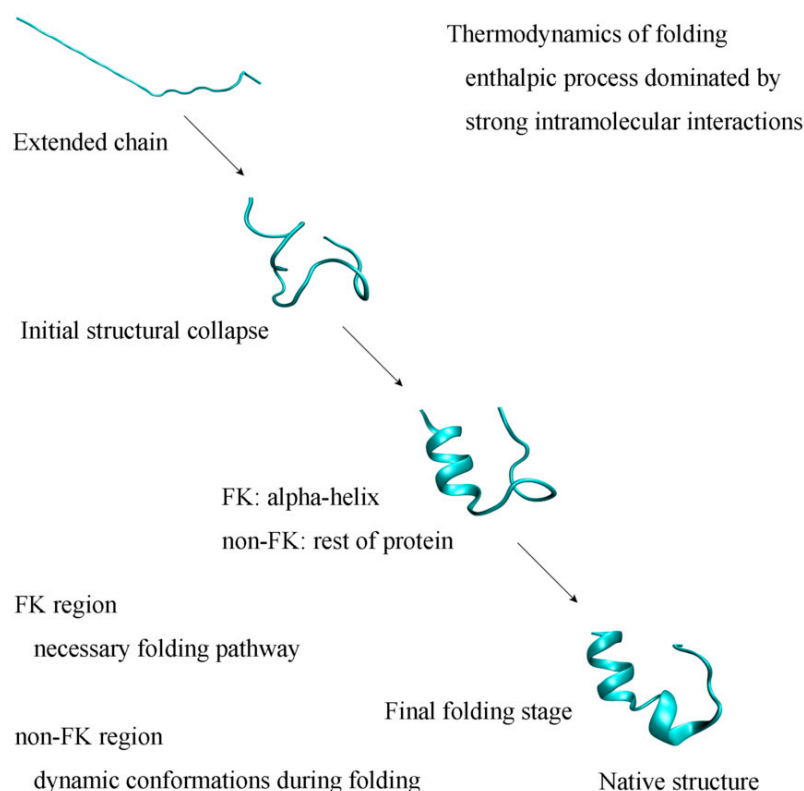


Figure 11. The folding mechanism of the TRP-cage in water.

In thermodynamics, folding is an enthalpic process that is due to various intermolecular interactions, especially hydrophobic interactions and intramolecular hydrogen bondings in proteins. Owing to hydrophobic interactions, folding leads to the structural collapse from a linear structure to a three-dimensional structure. With decreasing the separations between the atoms of proteins, these increase the direct interactions within proteins, which may affect the folding process. Therefore, hydrophobic interactions play an important role in the whole folding process, especially in the initial folding stage. In other words, protein folding may be driven by hydrophobic interactions. Based on the work, protein is divided into the FK regions related to the necessary pathways and the remaining non-FK regions. Due to the formations of intramolecular hydrogen bondings within proteins, specific conformations may be expected for the FK regions, such as α -helices and β -sheets. Proteins may contain several FK regions that are expected to sequentially form during folding. Therefore, they may be regarded as the “beacons” of protein folding. Additionally, dynamic conformations are expected for non-FK regions in the folding process. In the final folding stage, a native structure is obtained through the combined effects of various interactions, such as hydrophobic interactions, salt bridges, van der Waals interactions, hydrogen bondings related to side chains, etc.

5. Methods

MD simulations could provide insight into the folding pathways with unprecedented spatial and temporal resolutions. Therefore, they are widely utilized to investigate the folding mechanisms of proteins in water. In this study, to investigate the folding mechanism of the TRP-cage in water, MD simulations were carried out through the GROMACS 4.5.2 [120,121] package.

In this work, the AMBER99SB-ILDN force field [122] was utilized to describe the interatomic interactions. The water molecules were simulated using the generalized Born (GB) solvation model. This was used without PBC (periodic boundary conditions) and no pressure coupling. Velocity rescaling (v-rescale) thermostat dynamics were used to

control the temperatures. Additionally, the LINCS algorithm [123] was used to constrain the covalent bonds involving the hydrogen atoms. Salt concentrations were 0.2 M. The Trp-cage was, first, energy minimized for 50,000 steps using the steepest descent algorithm. Then, the total 2000 ns MD simulations were, respectively, carried out at 315 K, 320 K, and 350 K. A time step of 2 fs was used.

6. Conclusions

In combination with our recent studies on hydrophobic interactions, this work is devoted to investigating the folding mechanisms of proteins in water. From this work, the following conclusions were derived:

(1) Hydrophobic interactions are regarded as the fundamental driving forces in the folding process, especially in the initial stage. Due to hydrophobicity, this leads to the structural collapse of a protein in water. By decreasing the distances between the residues of proteins, the direct interactions between them become important, which may affect the folding process.

(2) Protein folding is a process of enthalpy to overcome the loss of conformational entropy arising from folding. In fact, various interactions may be involved in protein folding. Based on thermodynamic analysis, folding is reasonably guided by the strong interactions within proteins, such as intramolecular hydrogen bondings related to the α -helices and β -sheets of secondary structures.

(3) Proteins are divided into FK regions related to intramolecular hydrogen bondings and non-FK (the rest of protein) regions. During folding, specific and dynamic conformations are, respectively, expected for the FK and non-FK regions. Different from the non-FK regions, it is necessary for the FK regions to form the specific conformations in the folding process, which are regarded as the necessary pathways (or “beacons”) during folding. Additionally, sequential folding is expected for the FK regions, and an intermediate state is found during folding. In addition, they are reflected on the local basins in the FEL of folding.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules28135164/s1>, To study the folding pathways of the TRP-cage in water, total 2000 ns MD simulations were, respectively, carried out at 315 K, 320 K, and 350 K. To investigate the folding mechanisms of the TRP-cage in water, the secondary structures were analyzed by DSSP tool in VMD [119]. They were, respectively, drawn in Figures 8, S1 and S2. The trajectory files and Figures S1 and S2 are included in supplementary data. Figure S1. DSSP analysis of Trp-cage during the folding at 320 K. Figure S2. DSSP analysis of Trp-cage during the folding at 350 K.

Author Contributions: Q.S.: Conceptualization, Formal Analysis, Investigation, Resources, Data Curation, Writing—Original Draft Preparation, Writing—Review and Editing. X.H.: Writing—Review and Editing. Y.F.: Writing—Review and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The editor and reviewers are greatly appreciated for providing good suggestions to revise the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the compounds are available from the authors.

References

1. Anfinsen, C.B.; Haver, E.; Sela, M.; White, F.H.J. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 1309–1314. [[CrossRef](#)]
2. Levinthal, C. Are there pathways for protein folding? *J. Chem. Phys.* **1968**, *65*, 44–45. [[CrossRef](#)]
3. Lifson, S.; Roig, A. On the theory of helix-coil transition in polypeptides. *J. Chem. Phys.* **1961**, *34*, 1963–1974. [[CrossRef](#)]
4. Zimm, B.H.; Bragg, J.K. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* **1959**, *31*, 526–535. [[CrossRef](#)]
5. Baldwin, R.L. How does protein folding get started? *Trends Biochem. Sci.* **1989**, *14*, 291–294. [[CrossRef](#)] [[PubMed](#)]
6. Ptitsyn, O.B. How does protein synthesis give rise to the 3D-structure? *FEBS Lett.* **1991**, *285*, 176–181. [[CrossRef](#)] [[PubMed](#)]
7. Karplus, M.; Weaver, D.L. Protein-folding dynamics. *Nature* **1976**, *260*, 404–406. [[CrossRef](#)]
8. Karplus, M.; Weaver, D.L. Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Sci.* **1994**, *3*, 650–668. [[CrossRef](#)]
9. Dill, K.A. Theory for the folding and stability of globular proteins. *Biochemistry* **1985**, *24*, 1501–1509. [[CrossRef](#)]
10. Levitt, M.; Warshel, A. Computer simulation of protein folding. *Nature* **1975**, *253*, 694–698. [[CrossRef](#)]
11. Dill, K.A.; Fiebig, K.M.; Chan, H.S. Cooperativity in protein-folding kinetics. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 1942–1946. [[CrossRef](#)]
12. Bryngelson, J.D.; Wolynes, P.G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7524–7528. [[CrossRef](#)] [[PubMed](#)]
13. Leopold, P.E.; Montal, M.; Onuchic, J.N. Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8721–8725. [[CrossRef](#)] [[PubMed](#)]
14. Bryngelson, J.D.; Onuchic, J.N.; Succi, N.D.; Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **1995**, *21*, 167–195. [[CrossRef](#)] [[PubMed](#)]
15. Wolynes, P.G.; Onuchic, J.N.; Thirumalai, D. Navigating the folding routes. *Science* **1995**, *267*, 1619–1620. [[CrossRef](#)]
16. Dill, K.A.; Chan, H.S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997**, *4*, 10–19. [[CrossRef](#)]
17. Plotkin, S.S.; Onuchic, J.N. Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q. Rev. Biophys.* **2002**, *35*, 111–167. [[CrossRef](#)]
18. Sali, A.; Shakhnovich, E.; Karplus, M. Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. *J. Mol. Biol.* **1994**, *235*, 1614–1636.
19. Eaton, W.A.; Wolynes, P.G. Theory, simulations, and experiments show that proteins fold by multiple pathways. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E9759–E9760. [[CrossRef](#)]
20. Baldwin, R.L. Clash between energy landscape theory and foldon-dependent protein folding. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 8442–8443. [[CrossRef](#)]
21. Bai, Y.; Sosnick, T.R.; Mayne, L.; Englander, S.W. Protein folding intermediates: Native-state hydrogen exchange. *Science* **1995**, *269*, 192–197. [[CrossRef](#)] [[PubMed](#)]
22. Hu, W.; Walters, B.T.; Kan, Z.Y.; Mayne, L.; Rosen, L.E.; Marqusee, S.; Englander, S.W. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7684–7689. [[CrossRef](#)] [[PubMed](#)]
23. Hu, W.; Kan, Z.Y.; Mayne, L.; Englander, S.W. Cytochrome C folds through foldon-dependent native-like intermediates in an ordered pathway. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3809–3814. [[CrossRef](#)] [[PubMed](#)]
24. Krishna, M.M.G.; Englander, S.W. A unified mechanism for protein folding: Predetermined pathways with optional errors. *Protein Sci.* **2007**, *16*, 449–464. [[CrossRef](#)]
25. Englander, S.W.; Mayne, L. The case for defined protein folding pathways. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 8253–8258. [[CrossRef](#)]
26. Englander, S.W.; Mayne, L. Reply to Eaton and Wolynes: How do proteins fold? *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E9761–E9762. [[CrossRef](#)]
27. Frank, H.S.; Evans, M.W. Free volume and entropy in condensed systems III. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J. Chem. Phys.* **1945**, *13*, 507–532. [[CrossRef](#)]
28. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **1959**, *14*, 1–63.
29. Stillinger, F.H. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled particle theory. *J. Solution Chem.* **1973**, *2*, 141–158. [[CrossRef](#)]
30. Huang, D.M.; Geissler, P.L.; Chandler, D. Scaling of hydrophobic solvation free energies. *J. Phys. Chem. B* **2001**, *105*, 6704–6709. [[CrossRef](#)]
31. Lum, K.; Chandler, D.; Weeks, J.D. Hydrophobicity at small and large length scales. *J. Phys. Chem. B* **1999**, *103*, 4570–4577. [[CrossRef](#)]
32. Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437*, 640–647. [[CrossRef](#)] [[PubMed](#)]
33. Sun, Q. The physical origin of hydrophobic effects. *Chem. Phys. Lett.* **2017**, *672*, 21–25. [[CrossRef](#)]
34. Sun, Q.; Su, X.W.; Cheng, C.B. The dependence of hydrophobic interactions on the solute size. *Chem. Phys.* **2019**, *516*, 199–205. [[CrossRef](#)]
35. Sun, Q.; Zhang, M.X.; Cui, S. The structural origin of hydration repulsive force. *Chem. Phys. Lett.* **2019**, *714*, 30–36. [[CrossRef](#)]

36. Sun, Q.; Wang, W.Q.; Cui, S. Directional nature of hydrophobic interactions: Implications for the mechanism of molecular recognition. *Chem. Phys.* **2021**, *547*, 111200. [[CrossRef](#)]
37. Sun, Q.; Cui, S.; Zhang, M.X. Homogeneous nucleation mechanism of NaCl in aqueous solutions. *Crystals* **2020**, *10*, 107. [[CrossRef](#)]
38. Sun, Q.; Fu, Y.F.; Wang, W.Q. Temperature effects on hydrophobic interactions: Implications for protein unfolding. *Chem. Phys.* **2022**, *559*, 111550. [[CrossRef](#)]
39. Sun, Q. The hydrophobic effects: Our current understanding. *Molecules* **2022**, *27*, 7009. [[CrossRef](#)]
40. Stanley, H.E.; Teixeira, J. Interpretation of the unusual behavior of H₂O and D₂O at low temperatures: Tests of a percolation model. *J. Chem. Phys.* **1980**, *73*, 3404–3422. [[CrossRef](#)]
41. Nilsson, A.; Pettersson, L.G.M. Perspective on the structure of liquid water. *Chem. Phys.* **2011**, *389*, 1–34. [[CrossRef](#)]
42. Skinner, L.B.; Huang, C.; Schlesinger, D.; Pettersson, L.G.M.; Nilsson, A.; Benmore, C.J. Benchmark oxygen-oxygen pair-distribution function of ambient water from X-ray diffraction measurements with a wide Q-range. *J. Chem. Phys.* **2013**, *138*, 074506. [[CrossRef](#)]
43. Hura, G.; Sorenson, J.M.; Glaeser, R.M.; Head-Gordon, T. A high-quality X-ray scattering experiment on liquid water at ambient conditions. *J. Chem. Phys.* **2000**, *113*, 9140. [[CrossRef](#)]
44. Sun, Q. The Raman OH stretching bands of liquid water. *Vib. Spectrosc.* **2009**, *51*, 213–217. [[CrossRef](#)]
45. Sun, Q. Raman spectroscopic study of the effects of dissolved NaCl on water structure. *Vib. Spectrosc.* **2012**, *62*, 110–114. [[CrossRef](#)]
46. Sun, Q. Local statistical interpretation for water structure. *Chem. Phys. Lett.* **2013**, *568*, 90–94. [[CrossRef](#)]
47. Sun, Q. The effects of dissolved hydrophobic and hydrophilic groups on water structure. *J. Solution Chem.* **2020**, *49*, 1473–1484. [[CrossRef](#)]
48. Sun, Q.; Guo, Y. Vibrational sum frequency generation spectroscopy of the air/water interface. *J. Mol. Liquids* **2016**, *213*, 28–32. [[CrossRef](#)]
49. Dorsey, N.E. *Properties of Ordinary Water Substance*; ACS Monograph No. 81; Reinhold Publishing Corp.: New York, NY, USA, 1940.
50. Dill, K.A.; MacCallum, J.L. The protein-folding problem, 50 years on. *Science* **2012**, *338*, 1042–1046. [[CrossRef](#)]
51. Pace, C.N.; Hermans, J. The stability of globular protein. *CRC Crit. Rev. Biochem.* **1975**, *3*, 1–43. [[CrossRef](#)]
52. Goldenzweig, A.; Fleishman, S.J. Principles of protein stability and their application in computational design. *Annu. Rev. Biochem.* **2018**, *87*, 105–129. [[CrossRef](#)]
53. Baldwin, R.L. Energetics of protein folding. *J. Mol. Biol.* **2007**, *371*, 283–301. [[CrossRef](#)] [[PubMed](#)]
54. D’Aquino, J.A.; Gómez, J.; Hilser, V.J.; Lee, K.H.; Amzel, L.M.; Freire, E. The magnitude of the backbone conformational entropy change in protein folding. *Proteins Struct. Funct. Genet.* **1996**, *25*, 143–156. [[CrossRef](#)] [[PubMed](#)]
55. Makhatazde, G.I.; Privalov, P.L. On the entropy of protein folding. *Protein Sci.* **1996**, *5*, 507–510. [[CrossRef](#)]
56. Fitter, J. A measure of conformational entropy change during thermal protein unfolding using neutron spectroscopy. *Biophys. J.* **2003**, *84*, 3924. [[CrossRef](#)]
57. Baxa, M.C.; Haddadian, E.J.; Jumper, J.M.; Freed, K.F.; Sosnick, T.R. Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15396–15401. [[CrossRef](#)] [[PubMed](#)]
58. Towse, C.; Akke, M.; Daggett, V. The dynamomechanics entropy dictionary: A large-scale assessment of conformational entropy across protein fold space. *J. Phys. Chem. B* **2017**, *121*, 3933–3945. [[CrossRef](#)]
59. Dill, K.A.; Ozkan, S.B.; Shell, M.S.; Weikl, T.R. The protein folding problem. *Annu. Rev. Biophys.* **2008**, *37*, 289–316. [[CrossRef](#)] [[PubMed](#)]
60. Pace, C.N.; Fu, H.; Fryar, K.L.; Landua, J.; Trevino, S.R.; Shirley, B.A.; Hendricks, M.M.; Iimura, S.; Gajiwala, K.; Scholtz, J.M.; et al. Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol.* **2011**, *408*, 514–528. [[CrossRef](#)]
61. Pace, C.N.; Fu, H.; Fryar, K.L.; Landua, J.; Trevino, S.R.; Schell, D.; Thurlkill, R.L.; Imura, S.; Scholtz, J.M.; Gajiwala, K.; et al. Contribution of hydrogen bonds to protein stability. *Protein Sci.* **2014**, *23*, 652–661. [[CrossRef](#)] [[PubMed](#)]
62. Pace, C.N.; Scholtz, J.M.; Grimsley, G.R. Forces stabilizing proteins. *FEBS Lett.* **2014**, *588*, 2177–2184. [[CrossRef](#)] [[PubMed](#)]
63. Agashe, V.R.; Shastri, M.C.R.; Udgaonkar, J.B. Initial hydrophobic collapse in the folding of barstar. *Nature* **1995**, *377*, 754–757. [[CrossRef](#)] [[PubMed](#)]
64. Bhatia, S.; Krishnamoorthy, G.; Dhar, D.; Udgaonkar, J.B. Observation of continuous contraction and a metastable misfolded state during the collapse and folding of a small protein. *J. Mol. Biol.* **2019**, *431*, 3814–3826. [[CrossRef](#)] [[PubMed](#)]
65. Sherman, E.; Haran, G. Coil-globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11539–11543. [[CrossRef](#)]
66. Goluguri, R.R.; Udgaonkar, J.B. Microsecond rearrangements of hydrophobic clusters in an initially collapsed globule prime structure formation during the folding of a small protein. *J. Mol. Biol.* **2016**, *428*, 3102–3117. [[CrossRef](#)]
67. Makhatazde, G.I.; Privalov, P.L. Energetics of protein structure. *Adv. Protein Chem.* **1995**, *47*, 307–425.
68. Myers, J.K.; Pace, C.N. Hydrogen bonding stabilizes globular proteins. *Biophys. J.* **1996**, *71*, 2033–2039. [[CrossRef](#)]
69. Dill, K.A. Dominant forces in protein folding. *Biochemistry* **1990**, *29*, 7133–7155. [[CrossRef](#)]
70. Rose, G.D.; Fleming, P.J.; Banavar, J.R.; Maritan, A. A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16623–16633. [[CrossRef](#)]

71. Lewis, D.F.V. Hydrogen bonding in human p450-substrate interactions: A major contribution to binding affinity. *Sci. World J.* **2004**, *4*, 1074–1082. [[CrossRef](#)]
72. Newberry, R.W.; Raines, R.T. A prevalent intraresidue hydrogen bond stabilizes proteins. *Nat. Chem. Biol.* **2016**, *12*, 1084–1088. [[CrossRef](#)] [[PubMed](#)]
73. Chatterjee, S.; Roy, R.S.; Balaram, P. Expanding the polypeptide backbone: Hydrogen-bonded conformations in hybrid polypeptides containing the higher homologues of alpha-amino acids. *J. R. Soc. Interface* **2007**, *4*, 587–606. [[CrossRef](#)]
74. Rose, G.D. Reframing the protein folding problem: Entropy as organizer. *Biochemistry* **2021**, *60*, 3753–3761. [[CrossRef](#)]
75. McCarney, E.R.; Werner, J.H.; Bernstein, S.L.; Ruczinski, I.; Makarov, D.E.; Goodwin, P.M.; Plaxco, K.W. Site-specific dimensions across a highly denatured protein; A single molecule study. *J. Mol. Biol.* **2005**, *352*, 672–682. [[CrossRef](#)] [[PubMed](#)]
76. Mishra, P.; Kumar, S. The native state conformational heterogeneity in the energy landscape of protein folding. *Biophys. Chem.* **2022**, *283*, 106761. [[CrossRef](#)] [[PubMed](#)]
77. Bhatia, S.; Udgaonkar, J.B. Heterogeneity in protein folding and unfolding reactions. *Chem. Rev.* **2022**, *122*, 8911–8935. [[CrossRef](#)]
78. Baum, J.; Dobson, C.M.; Evans, P.A.; Hanley, C. Characterization of a partly folded protein by NMR methods: Studies of the molten globule state of guinea-pig α -lactalbumin. *Biochemistry* **1989**, *28*, 7–13. [[CrossRef](#)]
79. Hughson, F.M.; Wright, P.E.; Baldwin, R.L. Structural characterization of a partly folded apomyoglobin intermediate. *Science* **1990**, *249*, 1544–1548. [[CrossRef](#)]
80. Freire, E. Thermodynamics of partly folded intermediates in proteins. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 141–165. [[CrossRef](#)]
81. Ptitsyn, O.B. Molten globule and protein folding. *Advan. Protein Chem.* **1995**, *47*, 83–229.
82. Ptitsyn, O.B.; Uversky, V.N. The molten globule is a third thermodynamical state of protein molecules. *FEBS Lett.* **1994**, *341*, 15–18. [[CrossRef](#)] [[PubMed](#)]
83. Fink, A.L. (Ed.) Molten globule. In *Encyclopedia of Life Sciences*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2001; pp. 1–6.
84. Shakhnovich, E.I.; Finkelstein, A.V. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* **1989**, *28*, 1667–1680. [[CrossRef](#)]
85. Baldwin, R.L.; Rose, G.D. Molten globules, entropy-driven conformational change and protein folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 4–10. [[CrossRef](#)] [[PubMed](#)]
86. Neidigh, J.W.; Fesinmeyer, R.M.; Andersen, N.H. Designing a 20-residue protein. *Nat. Struct. Mol. Biol.* **2002**, *9*, 425–430. [[CrossRef](#)] [[PubMed](#)]
87. Scian, M.; Lin, J.C.; Le Trong, I.; Makhatadze, G.I.; Stenkamp, R.E.; Andersen, N.H. Crystal and NMR structures of a Trp-cage mini-protein benchmark for computational fold prediction. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12521–12525. [[CrossRef](#)]
88. Borgohain, G.; Paul, S. Atomistic level understanding of the stabilization of protein Trp-cage in denaturing and mixed osmolyte solutions. *Comput. Theor. Chem.* **2018**, *1131*, 78–89. [[CrossRef](#)]
89. Pal, S.; Roy, R.; Paul, S. Potential of a natural deep eutectic solvent, glyceline, in the thermal stability of the Trp-Cage mini-protein. *J. Phys. Chem. B* **2020**, *124*, 7598–7610. [[CrossRef](#)]
90. Qiu, L.; Pabit, S.A.; Roitberg, A.E.; Hagen, S.J. Smaller and faster: The 20-residue Trp-cage protein folds in 4 μ s. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953. [[CrossRef](#)]
91. Streicher, W.W.; Makhatadze, G.I. Unfolding thermodynamics of Trp-cage, a 20 residue miniprotein, studied by differential scanning calorimetry and circular dichroism spectroscopy. *Biochemistry* **2007**, *46*, 2876–2880. [[CrossRef](#)]
92. Abaskharon, R.M.; Culik, R.M.; Woolley, G.A.; Gai, F. Tuning the attempt frequency of protein folding dynamics via transition state rigidification: Application to Trp-cage. *J. Phys. Chem. Lett.* **2015**, *6*, 521–526. [[CrossRef](#)]
93. Byrne, A.; Williams, D.V.; Barua, B.; Hagen, S.J.; Kier, B.L.; Andersen, N.H. Folding dynamics and pathways of the Trp-cage miniproteins. *Biochemistry* **2014**, *53*, 6011–6021. [[CrossRef](#)]
94. Ahmed, Z.; Beta, I.A.; Mikhonin, A.V.; Asher, S.A. UV-resonance Raman thermal unfolding study of Trp-cage shows that it is not a simple two-state miniprotein. *J. Am. Chem. Soc.* **2005**, *127*, 10943–10950. [[CrossRef](#)] [[PubMed](#)]
95. Neuweiler, H.; Doose, S.; Sauer, M. A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16650–16655. [[CrossRef](#)]
96. Rovó, P.; Farkas, V.; Hegyi, O.; Szolomájer-Csikós, O.; Tóth, G.K.; Perczel, A. Cooperativity network of Trp-cage miniproteins: Probing salt-bridges. *J. Pept. Sci.* **2011**, *17*, 610–619. [[CrossRef](#)] [[PubMed](#)]
97. Rovó, P.; Stráner, P.; Láng, A.; Bartha, I.; Huszár, K.; Nyitray, L.; Perczel, A. Structural insights into the Trp-cage folding intermediate formation. *Chem. Eur. J.* **2013**, *19*, 2628–2640. [[CrossRef](#)] [[PubMed](#)]
98. Meuzelaar, H.; Marino, K.A.; Huerta-Viga, A.; Panman, M.R.; Smeenk, L.E.; Kettelarij, A.J.; van Maarseveen, J.H.; Timmerman, P.; Bolhuis, P.G.; Woutersen, S. Folding dynamics of the Trp-cage miniprotein: Evidence for a native-like intermediate from combined time-resolved vibrational spectroscopy and molecular dynamics simulations. *J. Phys. Chem. B* **2013**, *117*, 11490–11501. [[CrossRef](#)]
99. Zhou, R. Trp-cage: Folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13280–13285. [[CrossRef](#)]
100. Zheng, W.; Gallicchio, E.; Deng, N.; Andrec, M.; Levy, R.M. Kinetic network study of the diversity and temperature dependence of Trp-cage folding pathways: Combining transition path theory with stochastic simulations. *J. Phys. Chem. B* **2011**, *115*, 1512–1523. [[CrossRef](#)]
101. Day, R.; Paschek, D.; Garcia, A.E. Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 1889–1899. [[CrossRef](#)]

102. Deng, N.J.; Dai, W.; Levy, R.M. How kinetics within the unfolded state affects protein folding: An analysis based on Markov state models and an ultra-long MD trajectory. *J. Phys. Chem. B* **2013**, *117*, 12787–12799. [[CrossRef](#)]
103. Chowdhury, S.; Lee, M.C.; Duan, Y. Characterizing the rate-limiting step of Trp-cage folding by all-atom molecular dynamics simulations. *J. Phys. Chem. B* **2004**, *108*, 13855–13865. [[CrossRef](#)]
104. Du, W.; Bolhuis, P.G. Sampling the equilibrium kinetic network of Trp-cage in explicit solvent. *J. Chem. Phys.* **2014**, *140*, 195102. [[CrossRef](#)] [[PubMed](#)]
105. Han, W.; Schulten, K. Characterization of folding mechanisms of Trp-cage and WW-domain by network analysis of simulations with a hybrid-resolution model. *J. Phys. Chem. B* **2013**, *117*, 13367–13377. [[CrossRef](#)] [[PubMed](#)]
106. Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A kinetic model of Trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comput. Biol.* **2009**, *5*, 1000452. [[CrossRef](#)] [[PubMed](#)]
107. Andryushchenko, V.A.; Chekmarev, S.F. A hydrodynamic view of the first-passage folding of Trp-cage miniprotein. *Eur. Biophys. J.* **2016**, *45*, 229–243. [[CrossRef](#)]
108. Andryushchenko, V.A.; Chekmarev, S.F. Temperature evolution of Trp-cage folding pathways: An analysis by dividing the probability flux field into stream tubes. *J. Biol. Phys.* **2017**, *43*, 565–583. [[CrossRef](#)]
109. Kapakayala, A.B.; Nair, N.N. Boosting the conformational sampling by combining replica exchange with solute tempering and well-sliced metadynamics. *J. Comput. Chem.* **2021**, *42*, 2233–2240. [[CrossRef](#)]
110. Kim, S.B.; Dsilva, C.J.; Kevrekidis, I.G.; Debenedetti, P.G. Systematic characterization of protein folding pathways using diffusion maps: Application to Trp-cage miniprotein. *J. Chem. Phys.* **2015**, *142*, 085101. [[CrossRef](#)]
111. Juraszek, J.; Bolhuis, P. Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15859–15864. [[CrossRef](#)]
112. Paschek, D.; Hempel, S.; García, A.E. Computing the stability diagram of the Trp-cage miniprotein. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17754–17759. [[CrossRef](#)]
113. Pitera, J.W.; Swope, W. Understanding folding and design: Replica-exchange simulations of “Trp-cage” miniproteins. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 7587–7592. [[CrossRef](#)]
114. Ganguly, P.; Shea, J.E. Distinct and nonadditive effects of urea and guanidinium chloride on peptide solvation. *J. Phys. Chem. Lett.* **2019**, *10*, 7406–7413. [[CrossRef](#)]
115. Halabis, A.; Zmudzinska, W.; Liwo, A.; Oldziej, S. Conformational dynamics of the Trp-cage miniprotein at its folding temperature. *J. Phys. Chem. B* **2012**, *116*, 6898–6907. [[CrossRef](#)]
116. Lai, Z.; Preketes, N.K.; Mukamel, S.; Wang, J. Monitoring the folding of Trp-cage peptide by two-dimensional infrared (2DIR) spectroscopy. *J. Phys. Chem. B* **2013**, *117*, 4661–4669. [[CrossRef](#)]
117. Chen, J.; Im, W.; Brooks, C.L. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736. [[CrossRef](#)]
118. Im, W.; Lee, M.S.; Brooks, C.L. Generalized born model with a simple smoothing function. *J. Comput. Chem.* **2003**, *24*, 1691–1702. [[CrossRef](#)]
119. Humphrey, W.; Dalke, A.; Schulten, K. VMD-Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38. [[CrossRef](#)]
120. Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56. [[CrossRef](#)]
121. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. GROMACS: Fast, flexible and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [[CrossRef](#)] [[PubMed](#)]
122. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.L.; Dror, R.O.; Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99sb protein force field. *Proteins Struct. Funct. Genet.* **2010**, *78*, 1950–1958. [[CrossRef](#)] [[PubMed](#)]
123. Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.