THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL Letter

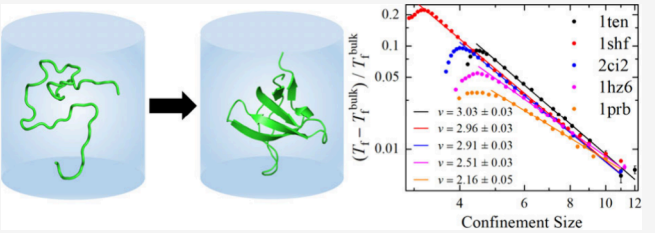
Scaling Laws for Protein Folding under Confinement

Bin Zhu,[#] Chenxi Zhang,[#] Jiwei Wang, Chuandong Jia, Teng Lu,* Liang Dai,* and Tao Chen*

Cite This: J. Phys. Chem. Lett. 2024, 15, 10138−10145 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Spatial confinement significantly affects protein folding. Without the confinement provided by chaperones, many proteins cannot fold correctly. However, the quantitative effect of confinement on protein folding remains elusive. In this study, we observed scaling laws between the variation in folding transition temperature and the size of confinement, $(T_f - T_f^{bulk})/T_f^{bulk} \sim L^{-\nu}$. The scaling exponent ν is significantly influenced by both the protein's topology and folding cooperativity. Specifically, for a given protein, ν can decrease as the folding cooperativity of the model increases, primarily due to the heightened sensitivity of the unfolded state energy to changes in cage size. For proteins with diverse topologies, variations in topological complexity influence scaling exponents in multiple ways. Notably, ν exhibits a clear positive correlation with contact order and the proportion of nonlocal contacts, as this complexity significantly enhances the sensitivity of entropy loss in the unfolded state. Furthermore, we developed a novel scaling argument yielding $5/3 \leq \nu \leq 10/3$, consistent with the simulation results.



Protein folding within the cellular environment is a complex process, significantly influenced by various factors, such as chaperones, crowding agents, and ribosomal exit tunnels.[1−3] Investigating protein folding within confined spaces is crucial, not only for elucidating the fundamental principles of regulation induced by these factors but also for guiding the design of encapsulation systems that aim to stabilize enzymes.[4,5] Consequently, the folding of proteins in confined spaces has attracted considerable attention.[6−50] These studies have significantly advanced our understanding of how confinement and chaperones regulate protein folding. However, several issues still need to be addressed, particularly concerning the quantitative impact of confinement on protein folding.

Utilizing the Gō-like models,[51,52] previous studies have shown that the shift in folding transition temperature relative to the bulk value changes with the size of the confining cage as $(T_f - T_f^{bulk})/T_f^{bulk} \sim L^{-\nu}$. Notably, the scaling exponent ν varies significantly across different studies.[20,24,47] For instance, one study on protein folding under cylindrical confinement reported a scaling exponent of $\nu \approx 3.25$, irrespective of the topological diversity of the proteins.[20] However, another investigation on protein folding under different dimensional constraints observed a uniform scaling exponent of $\nu \approx 5/3$.[24] Furthermore, an earlier study on protein folding under spherical confinement revealed that the scaling exponent ν is relevant to specific proteins and the repulsive confinement model.[47] Therefore, a unified theoretical explanation for the variability in scaling exponents remains to be explored.

The geometry of the confining medium significantly affects protein conformations and interactions, thereby influencing protein folding behaviors.[12,20,24,42,50] An important question is to what extent geometric factors affect the stability and folding dynamics of proteins under confinement. Researchers have explored this issue by examining protein folding in confined spaces of various dimensions, including slit pores, cylindrical pores, and spherical cavities.[12,24,50] These studies have made considerable progress in understanding the differences in thermodynamic stability and folding dynamics caused by the geometric shapes of confining media across different dimensions. However, it remains unclear how the geometric effects of confined media, such as cylindrical and spherical cavities, impact protein folding behavior within the same dimension. Addressing this issue is important to elucidate whether the cylindrical shape of chaperonins offers any advantages.

To date, the simple Gō model has made significant contributions in advancing our understanding of protein folding mechanisms in bulk solutions and within diverse confining environments.[20,24,52−58] However, some studies have indicated that this model does not adequately capture the folding cooperativity and diversity in folding rates of certain proteins as observed in experiments.[59−61] In contrast, incorporating desolvation barriers—typical features of intramolecular and intermolecular interactions—into the model,

Received: July 16, 2024
Revised: September 24, 2024
Accepted: September 25, 2024

https://doi.org/10.1021/acs.jpclett.4c02098
J. Phys. Chem. Lett. 2024, 15, 10138−10145

termed as the db model, can effectively capture these folding behaviors.[59−66] Moreover, the structures of the unfolded state are critical in influencing the thermodynamics of the folding transition, particularly in the context of protein folding under confinement.[45] Notably, the size of the unfolded state derived from the db model, which is generally consistent with experimental results,[67] tends to be larger than that obtained from the common Gō model (Figure S1 and Table S1).

In this study, we will employ the db model, coupled with theoretical analysis, to investigate the effects of confinement on the thermodynamic stability and folding mechanisms of small two-state proteins across diverse topological structures. Our primary objective is to uncover the quantitative impact of confinement on protein folding. Additionally, we aim to determine how the shape of the confining medium affects protein folding behavior.

Models and Methods. In the db model, both the solvation and desolvation barriers are integrated into the native-centric attraction potential of the protein chain. Details of the effective energy ($E_{protein}$) of the intraprotein interactions for a substate protein encapsulated in a cage are provided in the Supporting Information.

The repulsive interactions between the confined protein and the cage (E_{cage}) are given by[26]

$$E_{cage} = \sum_{i=1}^{N} \varepsilon \left(\frac{\sigma}{d_i + \sigma/2} \right)^{12} \quad (1)$$

where N represents the number of residues in a given protein, $\varepsilon = 1$ is the unit of energy, $\sigma = 4$ Å is the unit of length, and d_i represents the distance of the ith residue to the inner wall of the cage. Here, the cutoff distance d_c for the repulsive interactions is designated as $d_c = 1.0\sigma$, at which the repulsive energy between the residue and the wall is $\sim 0.0077\varepsilon$. Accordingly, the total potential energy E of the system is described as

$$E = E_{protein} + E_{cage} \quad (2)$$

The Langevin dynamics method is utilized to investigate the folding and unfolding dynamics of substrate proteins, both in bulk and when confined within cages.[66,68] To ensure efficient and systematic sampling, the replica-exchange method (REM) is employed in thermodynamic simulations.[69]

In the present study, unless otherwise stated, the cage is assumed to be cylindrical, analogous to the shape of GroEL/GroES chaperonin cavity (Figure 1a). This investigation utilizes five small single-domain proteins which encompass a range of topological structures, such as β proteins, α/β proteins, and α proteins (Figure 1b).

Effects of Confinement on Thermodynamic Stability. To investigate the influence of confinement on thermodynamic stability of proteins, the change in folding transition temperature relative to the bulk value, $(T_f - T_f^{bulk})/T_f^{bulk}$, is computed across a wide range of cage sizes. Here, T_f and T_f^{bulk} represent the folding transition temperatures in cages and in bulk, respectively, which are identified by the peak positions in the heat capacity $C_v(T)$ curves (Figure S2). As the cage size (L) reduces, the $(T_f - T_f^{bulk})/T_f^{bulk}$ increases until a critical cage size (L_C) is reached, and then it begins to decrease (Figure 2a).

As depicted in Figure 2a, the extent to which confinement affects protein stability is related to specific proteins and cage size. According to the melting temperatures measured by experiments,[70,71] the folding transition temperatures of the Fyn SH3 domain and the albumin binding domain are increased by ~78.5 and ~13.0 °C, respectively, at the maximum value of $(T_f - T_f^{bulk})/T_f^{bulk}$. Notably, the location of the crossover is expected to depend on the topological features of proteins, as the critical size of cage strongly correlates with the longest residue−residue distance (R_{max}) in the native PDB structure (Figure S3). When the cage size L is smaller than $R_{max}/2$, the folded structure deforms, destabilizing the folded state.

Scaling Analysis. Before reaching the crossover region, within a wide range of cage sizes, a scaling relationship is observed, namely $(T_f - T_f^{bulk})/T_f^{bulk} \sim (L/\sigma)^{-\nu}$. It appears that the scaling exponent ν is related to specific proteins: the values of ν for the Fibronectin type III, Fyn SH3 domain, Chymotrypsin inhibitor 2, Protein L, and Albumin-binding

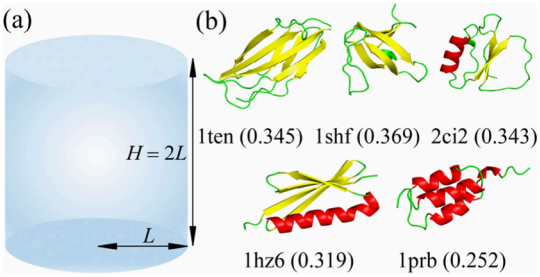


Figure 1. (a) Schematic representation of the cylindrical cage with a radius of L and a height of $H = 2L$. (b) Native structures of the proteins investigated, arranged from top left to bottom right: Fibronectin type III, Fyn SH3 domain, Chymotrypsin inhibitor 2, Protein L, and Albumin-binding domain. The corresponding PDB codes are listed below each structure, with the contact order (CO) of the native structure provided in parentheses.

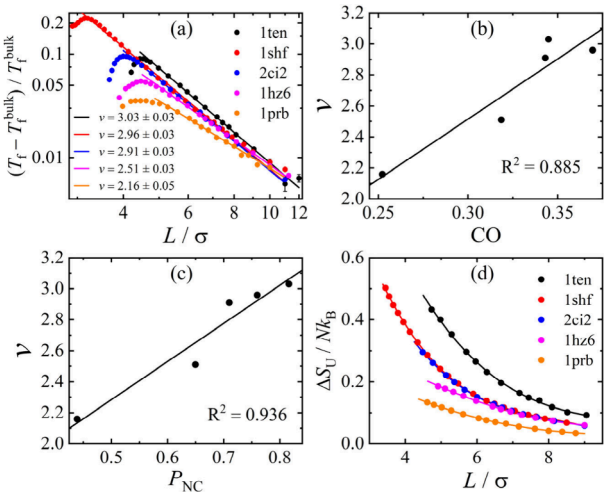


Figure 2. (a) The change in the folding transition temperature relative to the bulk value, $(T_f - T_f^{bulk})/T_f^{bulk}$, as a function of the cage size, L, for the five proteins under study. Lines represent the scaling relationship given by $(T_f - T_f^{bulk})/T_f^{bulk} \sim (L/\sigma)^{-\nu}$, where σ denotes the unit of length. (b) Relationship between the scaling exponent ν and the contact order (CO) of the native structure. (c) Correlation between the scaling exponent ν and the proportion of the nonlocal contact number P_{NC} of the native structure. (d) The entropic loss of the unfolded state ΔS_U, as a function of L, corresponding to the scaling region, at the proteins' respective bulk transition temperatures. Here, $\Delta S_U = S_U(\text{bulk}) - S_U(L)$, with $S_U(\text{bulk})$ and $S_U(L)$ representing the entropy of the unfolded state in bulk solutions and within the cage of size L, respectively.

https://doi.org/10.1021/acs.jpclett.4c02098
J. Phys. Chem. Lett. 2024, 15, 10138−10145

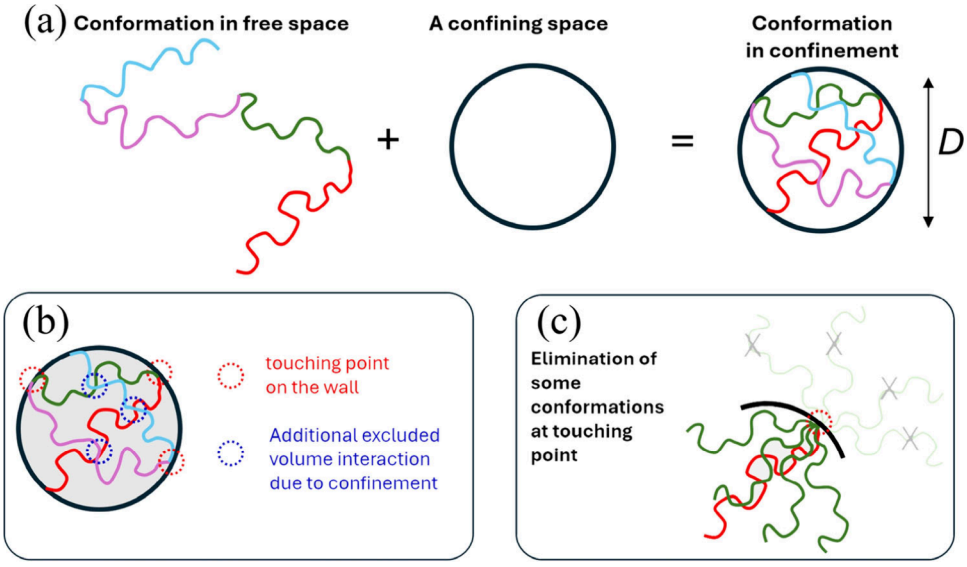


Figure 3. (a) Illustration of how a confining space affects polymer conformations. The reflection points at confining wall separate the entire chain to many segments, which are labeled by different colors. (b) Illustration of reflection points of polymer conformations and additional excluded volume interactions caused by confinement. (c) At each reflection point, some possible conformations (light green) are eliminated due to the existence of the confining wall, and some possible conformations (dark green) are not affected by the confinement.

domain are ~3.03, 2.96, 2.91, 2.51, and 2.16, respectively (Figure 2a).

To elucidate the physical basis underlying the variation in scaling exponents among different proteins, we quantified the topological features of these proteins by utilizing the contact order (CO) of the native structure, where $\mathrm{CO} = \frac{1}{NM}\sum_{j>i+3} l_{ij}$, with $l_{ij} = |j - i|$ and M denotes the number of contacts in the native structure.[72] As illustrated in Figure 2b, a moderate correlation exists between ν and CO, indicating that ν is associated with the topological structures of proteins and generally increases with enhancing topological complexity. Notably, a stronger correlation is observed between ν and P_{NC}, where P_{NC} is the proportion of the nonlocal contact number in the native structure (Figure 2c). Here, a nonlocal contact refers to the native contact formed between residues i and j with $|j - i| > 10$.[73] This result suggests that long-distance contacts play an important role in influencing the scaling exponent ν (Figure S4).

As mentioned above, the scaling exponent ν is significantly influenced by the proportion of nonlocal contacts (P_{NC}). This is likely due to the fact that the formation of long-distance contacts in the unfolded state induces a larger entropic loss, compared to short-distance contacts. As depicted in Figure 2d, with the reduction of cage size, the entropic loss of the unfolded state (ΔS_U) increases, and the ΔS_U for the protein with a higher P_{NC} is essentially larger than that for the protein with a lower P_{NC}. Particularly, the rate of increase in ΔS_U for the protein with a higher P_{NC} is greater than that for the protein with a lower P_{NC}, consequently resulting in a larger scaling exponent ν.

We observed discrepancies in scaling exponents between our results and other studies, which may be attributed to differences in folding cooperativity among the models. The introduction of the desolvation barrier in the db model significantly narrows the potential well, compared to the simple Gō model. With a decrease in the width of the potential well in the Gō model, the ratio of the full width at half-maximum (FWHM) of the heat capacity peak to the folding transition temperature (T_f) decreases, while the height of free-energy barrier increases, indicating increased folding cooperativity (Figure S5).[74,75] In the modified Gō model, similar potential well width to the db model leads to comparable folding cooperativity for the Fyn SH3 domain, suggesting that the narrow width of the potential well is likely a critical factor promoting cooperative folding in the db model (Figure S5).[59] It is observed that, for a given protein, an increase in the folding cooperativity of the model generally results in a decrease in the scaling exponent (Figure S6). The db model, with enhanced folding cooperativity, exhibits more extended conformations in the unfolded state in bulk solution, making entropy loss in the unfolded state more sensitive to cage size changes, potentially increasing the scaling exponent (Figure S7a). Conversely, the variation in energy of the unfolded state also shows heightened sensitivity to changes in cage size, leading to a decrease in the scaling exponent (Figure S7b). Therefore, the smaller scaling exponent in the db model compared to the simple Gō model likely results from the more sensitive response of the unfolded state energy to changes in cage size.

For proteins with diverse topologies, increased topological complexity can enhance folding cooperativity,[61] potentially reducing ν. However, an increase in this complexity markedly heightens the sensitivity of entropy loss in the unfolded state to variations in cage size (Figure 2d), thereby raising ν (Figures 2b and 2c). Therefore, variations in topological complexity influence scaling exponents in multiple ways.

Here, we introduce a theoretical approach to explore the scaling relationship of confinement free energy. For a polymer chain confined in a spherical cavity with diameter of D, we derived the scaling behavior for the confinement free energy using the following approximation. We assume that this polymer performs random walk except touching the confining wall, from the red segment to the green, purple, and blue segments (Figure 3a). Each touching point with the confining wall results in an increase in confinement free energy on the order of k_BT, due to the restriction of the polymer's

https://doi.org/10.1021/acs.jpclett.4c02098
J. Phys. Chem. Lett. 2024, 15, 10138−10145

conformation (Figure 3b). We can understand it in this way. In the absence of a confining wall, the polymer can choose any direction during random walking, including both the light green and dark green segments in Figure 3c. However, in the case with a confining wall, the polymer can only choose roughly half of the available directions at the touching point (Figure S8), i.e., the light green segments are prohibited by the confining wall (Figure 3c). The number of such touching points can be obtained by

$$N_{\text{touch}} \approx \frac{L_{\text{chain}}}{L_{\text{seg}}} \tag{3}$$

where L_{chain} represents the total length of the polymer chain, and $L_{\text{seg}} \sim D^{5/3}$ is the average contour length of the segment between two touching points.[76] Thus, the confinement free energy caused by these touching points is expressed as

$$F_{\text{touch}} \approx N_{\text{touch}} k_B T \sim L_{\text{chain}} D^{-5/3} \tag{4}$$

Furthermore, the confinement also induces excluded volume (EV) interactions among the segments (red, green, purple, and blue in Figure 3b), contributing additional confinement free energy proportional to the square of the number of segment pairs, as described by

$$F_{EV} \sim N_{\text{touch}}^2 \sim L_{\text{chain}}^2 D^{-10/3} \tag{5}$$

By combining the above two contributions, we obtained the total confinement free energy

$$\begin{aligned} F_{\text{confine}} &= F_{\text{touch}} + F_{\text{EV}} \\ &\sim L_{\text{chain}} D^{-5/3} + L_{\text{chain}}^2 D^{-10/3} \end{aligned} \tag{6}$$

Due to the mixture of two scaling terms, the apparent scaling exponent is expected to vary between −5/3 and −10/3, depending on the relative weight of two terms. Notably, the predicted range of $5/3 \le \nu \le 10/3$, agrees with the data reported in some previous studies and is further validated by our simulation results (see section S3 in the Supporting Information (SI)). We can anticipate that L_{chain} affects the relative weights of F_{touch} and F_{EV}, thereby affecting the apparent scaling exponent ν. For a larger L_{chain}, the relative weight of F_{EV} increases, and the scaling exponent is expected to be closer to −10/3. For more explanation and discussion about the scaling analysis, see section S4 and Figure S9 in the SI).

Effects of Confinement on the Specific Entropy Terms. Within a wide range of cage sizes, a reduction in cage size results in a noticeable loss of conformation entropy in the unfolded state (Figure 4a). However, it remains unclear how confinement influences the specific entropy components. To address this issue, the conformation entropy at a given fraction of native contacts (Q) is expressed as[77−79]

$$S = Ns_0 + S_{\text{bond}} + S_{\text{route}} \tag{7}$$

with $Ns_0 = S(Q = 0) - S(Q = 1)$.[77] S_{bond} in eq 7 is the bond entropy, resulting from the entropy loss due to the formation of contacts, which is given by[77−79]

$$S_{\text{bond}} = S_{\text{MF}} - \frac{3}{2} k_B M \langle \delta Q \delta \log l \rangle \tag{8}$$

where k_B is the Boltzmann constant, M denotes the number of contacts in the native structure, and l is the loop length. The first term S_{MF} in eq 8 represents the bond entropy derived from a mean field approximation for $l_i = \overline{l}$, where l_i and $\overline{l}$ represent

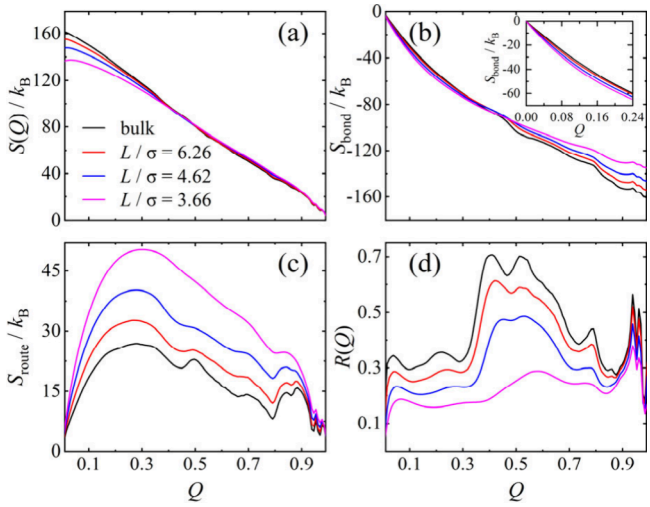


Figure 4. Effects of confinement on specific entropy components and folding routes. (a) Conformation entropy $S(Q)$, (b) bond entropy S_{bond}, (c) route entropy S_{route}, and (d) route measure $R(Q)$ as functions of the fraction of native contacts Q for the Fyn SH3 domain in bulk and confined in cylindrical cages of different sizes at their respective folding transition temperatures. The inset in panel (b) provides a magnified representation of bond entropy in the unfolded state.

the loop length of the ith contact and the average loop length, respectively. The detailed form of S_{MF} is expressed as[77−79]

$$\begin{aligned} S_{\text{MF}} = &-QNs_0 - \frac{3}{2} k_B MQ \left(\frac{\overline{l} \log \overline{l}}{\overline{l} - 1} \right) \\ &+ \frac{3}{2} k_B M \left(\frac{1}{\overline{l} - 1} \right) [1 + (\overline{l} - 1)Q] \log[1 + (\overline{l} - 1)Q] \end{aligned} \tag{9}$$

The second term in eq 8 represents the variation in entropy loss due to fluctuations in loop length, which is given by[77−79]

$$M \langle \delta Q \delta \log l \rangle = \sum_{i=1}^{M} (Q_i - Q)(\log l_i - \overline{\log l}) \tag{10}$$

where Q_i denotes the formation probability of the ith contact at a given Q.

S_{route} is the route entropy obtained from counting all different arrangements of MQ contacts at a given Q, which is expressed as[77−79]

$$S_{\text{route}} = k_B \lambda(Q) \sum_{i=1}^{M} [-Q_i \log Q_i - (1 - Q_i) \log(1 - Q_i)] \tag{11}$$

where $\lambda(Q)$ quantifies the entropy loss associated with the chain connectivity.[77] Considering the difficulty in obtaining the analytical form of $\lambda(Q)$, we directly calculate S_{route} via $S_{\text{route}} = S - Ns_0 - S_{\text{bond}}$, assuming that $S(Q = 1)$ is constant in large cage sizes.

As Q increases, there is an enhanced entropic loss resulting from contact formation, reflected by a decrease in the bond entropy S_{bond} (Figure 4b). Notably, with a decrease in cage size, the value of S_{bond} decreases in the unfolded basin while increasing in the folded basin (Figure 4b), thereby contributing to both thermodynamic stability and folding kinetics. In the unfolded basin, a decrease in cage size promotes the formation of nonlocal contacts (Figure S10), leading to an increase in

https://doi.org/10.1021/acs.jpclett.4c02098
J. Phys. Chem. Lett. 2024, 15, 10138−10145

bond entropy loss and, consequently, a decrease in S_{bond} value. Conversely, in the folded basin, a smaller cage size induces increasingly collapsed conformations, thereby reducing bond entropy loss upon contact formation. Thus, changes in cage size have opposing effects on S_{bond} in the unfolded and folded basins, highlighting the complex interplay between molecular confinement and conformation state.

As shown in Figure 4c, the route entropy S_{route} generally exhibits an initial increase followed by a decrease as Q increases. Notably, a decrease in cage size leads to an increase in S_{route}, which should be associated with the change in potential folding routes. To quantitatively assess the impact of confinement on folding routes, we computed the route measure $R(Q)$, which is defined as $R(Q) = \frac{1}{MQ(1-Q)}\sum_{i=1}^{M}(Q_i - Q)^2$.[60,61,78] According to this definition, the range of $R(Q)$ is $0 \le R(Q) \le 1$. As discussed previously, a higher $R(Q)$ indicates fewer pathways through a given Q value, while a lower $R(Q)$ suggests more pathways.[61,78] Thus, $R(Q)$ also reflects the topography of energy landscape. Generally, the $R(Q)$ exhibits a double-hump feature, akin to the results observed in other db models.[60,61] Remarkably, as the cage size decreases, the value of $R(Q)$ reduces (Figures 4d and S11), indicating that a smaller cage size will lead to an increase in folding routes and a decrease in folding route specificity, which may be beneficial for protein folding. Moreover, an increase in folding routes with reducing the cage size reinforces the evidence that confinement enhances route entropy (Figure 4c). It is worth noting that the discrepancy between the peak locations of $R(Q)$ and S_{route} should be attributed to the effect of chain connectivity (Figure S12).

Cage Shapes Influencing Folding Behavior. To get insight into how the shape of the cage influences protein stability, we analyzed the $(T_f - T_f^{bulk})/T_f^{bulk}$ values of proteins confined into cylindrical and spherical cages, respectively, at the same volume fraction of residues η (Figure 5). Here, the volume fraction η is defined as $\eta = N^*(4/3)\pi(a/2)^3/V$, where a = 3.8 Å represents the average $C_\alpha - C_\alpha$ distance of two adjacent residues along the protein chain or the diameter of the bead in the coarse-grained model, and V is the volume of the cylindrical or spherical cage. With increasing the volume fraction of residues η, the value of $(T_f - T_f^{bulk})/T_f^{bulk}$ first increases to a critical value and then decreases in both cylindrical and spherical cages. This indicates that the trend in thermodynamic stability variations induced by confinement is consistent across cages of differing shapes. However, the critical value of η, at which the $(T_f - T_f^{bulk})/T_f^{bulk}$ displays the maximal value, differs between the two differently shaped cages, which should be due to the inherent topological feature of the protein. Notably, at the same η value, the value of $(T_f - T_f^{bulk})/T_f^{bulk}$ is higher in cylindrical cages than in spherical ones (Figures 5a and 5b). This suggests that cylindrical cages can enhance protein stability more effectively than spherical cages. To elucidate the physical basis underlying these stability differences, we calculated the energetic changes for the Fyn SH3 domain and protein L confined in cylindrical and spherical cages, respectively (see Figures 5c and 5d). For the two proteins, when the cage size is sufficiently large to accommodate the native folded structures, a higher stability in the cylindrical cage arises from more entropic loss in the confined unfolded state. This is primarily because the unfolded conformations of proteins within cylindrical cages are more collapsed (Figure S13), leading to a larger entropy loss

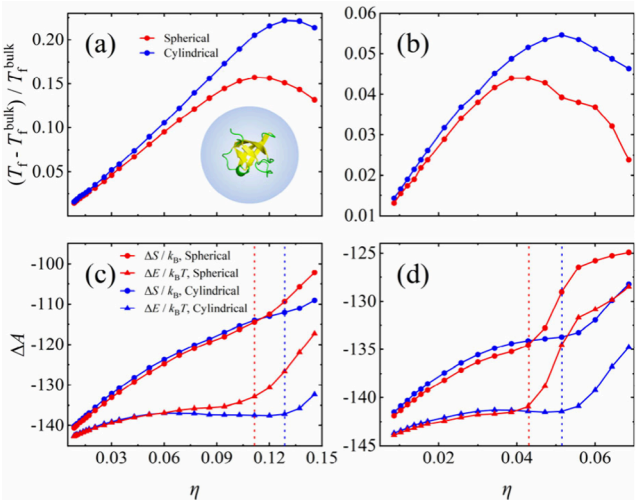


Figure 5. Effects of cage shape on thermodynamic properties. (a, b) The shift in folding transition temperature with respect to the bulk value, $(T_f - T_f^{bulk})/T_f^{bulk}$, as a function of the volume fraction of residues (η) for the Fyn SH3 domain (panel (a)) and protein L (panel (b)) confined in spherical (red symbols) and cylindrical (blue symbols) cages. (c, d) The variations in entropy $\Delta S/k_B$ (solid circles) and in energy $\Delta E/k_BT$ (triangles) as functions of η for the Fyn SH3 domain (panel (c)) and protein L (panel (d)) confined in spherical (red symbols) and cylindrical (blue symbols) cages at their respective bulk transition temperatures. Here, $\Delta S = S_N - S_U$ and $\Delta E = E_N - E_U$. The η values corresponding to the largest increase in transition temperature are indicated by vertical dashed lines using the same color code. A schematic representation of a protein confined in a spherical cage is illustrated in panel (a).

compared to those in spherical cages. However, when the cages are too small to accommodate the native folded structure, the increased energy of the folded state in spherical cages results in lower stability, compared to that in cylindrical cages. It is noteworthy that, despite noticeable differences in thermodynamic stability, the scaling exponents ν are nearly equal for the two distinct shapes of cages (Figure S14).

As stated above, when the cage is able to accommodate the native structure, confinement within a cylindrical cage may induce a higher entropic loss in the unfolded state, which, in turn, destabilizes this state and results in a reduced folding barrier (Figure S15a). According to transition-state theory, $k_f \sim \exp(-\Delta F^{\ddagger}/k_BT)$,[80,81] with k_f and $\Delta F^{\ddagger}$ representing folding rate and folding barrier height, respectively; a lower folding barrier is associated with a higher folding rate (see the change in folding rate in Figure S15b). Therefore, cylindrical cages may be more favorable for protein folding, both thermodynamically and kinetically, compared to spherical ones under certain conditions.

In summary, by utilizing an enhanced cooperative model, we observed scaling laws relating the change in folding transition temperature to the size of confinement, expressed as $(T_f - T_f^{bulk})/T_f^{bulk} \sim L^{-\nu}$, across a broad range of cage sizes. The scaling exponent ν is dependent on both the topology of the protein and its folding cooperativity. Notably, for proteins with diverse topologies, the scaling exponent exhibits a clear positive correlation with both contact order and the proportion of nonlocal contacts. This correlation is mainly due to increased topological complexity significantly enhancing the sensitivity of entropy loss in the unfolded state to variations in cage size, which results in an increase in ν. Additionally, we

https://doi.org/10.1021/acs.jpclett.4c02098
J. Phys. Chem. Lett. 2024, 15, 10138−10145

found that, for a specific protein, enhancing folding cooperativity of the model generally leads to a decrease in ν, primarily due to the increased sensitivity of the unfolded state energy to variations in cage size. Furthermore, we developed a theoretical approach to explore the scaling relationship, and the derived range $5/3 \leq \nu \leq 10/3$ is consistent with the simulation results.

The shape of the confinement also significantly influences protein folding. When the cages are designed to accommodate the native folded structures, proteins confined in cylindrical cages are more favorable for folding in both thermodynamic and kinetic aspects, due to more entropy loss in the confined unfolded state, compared to those in spherical ones. This result implies that the cylindrical form of chaperonins might be a result of natural selection, optimized for enhancing protein folding efficiency. This study deepens our understanding of how chaperonins influence protein folding and can inform the rational design of artificial cages for enzyme stabilization.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.4c02098.

Comparisons between the common Gō model and the db model (section S1); comparisons of attractive potentials and free-energy profiles between the two models (Figure S1); comparisons of free-energy barrier heights and unfolded state sizes of the Fyn SH3 domain between the two models (Table S1); effects of confinement on thermodynamic stability (Figure S2); relationship between L_C and R_{max} (Figure S3); correlations between ν and P_{NC} (Figure S4); short-range attractive potential of the modified Gō model (S2); effect of the potential well width on the folding cooperativity (Figure S5); effect of the potential well width on scaling exponents (Figure S6); entropy loss and energy change in the unfolded state caused by confinement (Figure S7); effect of confining medium shape on the available directions of the polymer chain (Figure S8); scaling exponent derived from theoretical analysis (S3); confinement free energy of a polymer chain confined within a tube (S4); schematic representation of a semiflexible polymer chain confined within a tube of diameter D (Figure S9); $\langle CO \rangle$ as a function of Q (Figure S10); $R(Q)$ as functions of Q at the bulk transition temperature (Figure S11); route entropy without considering the chain connectivity effect (S^*_{route}) as functions of Q (Figure S12); distributions of radius of gyration in different states (Figure S13); comparison of the scaling exponents for the Fyn SH3 domain confined within spherical and cylindrical cages (Figure S14); effects of cage shape on free-energy barriers and folding rates (Figure S15) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Tao Chen − *College of Chemistry and Materials Science, Northwest University, Xi'an 710127, China; Key Laboratory of Polymer Processing Engineering (South China University of Technology), Ministry of Education, Guangzhou 510641, China;* orcid.org/0000-0001-6686-1821; Email: tchen@nwu.edu.cn

Liang Dai − *Department of Physics, City University of Hong Kong, Hong Kong 999077, China; Shenzhen Research Institute, City University of Hong Kong, Shenzhen 518057, P. R. China;* orcid.org/0000-0002-4672-6283; Email: liangdai@cityu.edu.hk

Teng Lu − *Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083, China;* orcid.org/0000-0002-7982-3370; Email: luteng@sccas.cn

Authors

Bin Zhu − *College of Chemistry and Materials Science, Northwest University, Xi'an 710127, China*

Chenxi Zhang − *College of Chemistry and Materials Science, Northwest University, Xi'an 710127, China*

Jiwei Wang − *College of Chemistry and Materials Science, Northwest University, Xi'an 710127, China*

Chuandong Jia − *College of Chemistry and Materials Science, Northwest University, Xi'an 710127, China;* orcid.org/0000-0001-7543-2535

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.4c02098

Author Contributions

Authors B. Zhu and C. Zhang contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Nos. 22171226, 22273080) and the Natural Science Foundation of Shaanxi Province (Nos. 2023JCYB132, 2022JC06). The partial supercomputing resources were provided by the China Scientific Computing Grid of Chinese Academy of Sciences and Gansu Computing Center.

■ REFERENCES

(1) Balchin, D.; Miličić, G.; Strauss, M.; Hayer-Hartl, M.; Hartl, F. U. Pathway of actin folding directed by the eukaryotic chaperonin TRiC. *Cell* **2018**, *174*, 1507−1521.

(2) Ellis, R. J.; Minton, A. P. Cell biology—Join the crowd. *Nature* **2003**, *425*, 27−28.

(3) Holtkamp, W.; Kokic, G.; Jäger, M.; Mittelstaet, J.; Komar, A. A.; Rodnina, M. V. Cotranslational protein folding on the ribosome monitored in real time. *Science* **2015**, *350*, 1104−1107.

(4) Sun, Q.; Fu, C.; Aguila, B.; Perman, J.; Wang, S.; Huang, H.; Xiao, F.; Ma, S. Pore environment control and enhanced performance of enzymes infiltrated in covalent organic frameworks. *J. Am. Chem. Soc.* **2018**, *140*, 984−992.

(5) Chen, G. S.; Tong, L. J.; Huang, S. M.; Huang, S. Y.; Zhu, F.; Ouyang, G. F. Hydrogen-bonded organic framework biomimetic entrapment allowing non-native biocatalytic activity in enzyme. *Nat. Commun.* **2022**, *13*, 4816.

(6) Chu, X. K.; Suo, Z. C.; Wang, J. Confinement and crowding effects on folding of a multidomain Y-family DNA polymerase. *J. Chem. Theory Chem.* **2020**, *16*, 1319−1332.

(7) Especial, J. N. C.; Nunes, A.; Rey, A.; Faísca, P. F. N. Hydrophobic confinement modulates thermal stability and assists knotting in the folding of tangled proteins. *Phys. Chem. Chem. Phys.* **2019**, *21*, 11764−11775.

(8) Soler, M. A.; Rey, A.; Faísca, P. F. N. Steric confinement and enhanced local flexibility assist knotting in simple models of protein folding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 26391−26403.

(9) Zhao, Y. N.; Dabrowski-Tumanski, P.; Niewieczerzal, S.; Sulkowska, J. I. The exclusive effects of chaperonin on the behavior of proteins with 5_2 knot. *PLoS Comput. Biol.* **2018**, *14*, No. e1005970.
(10) Niewieczerzal, S.; Sulkowska, J. I. Knotting and unknotting proteins in the chaperonin cage: Effects of the excluded volume. *PLoS One* **2017**, *12*, No. e0176744.
(11) Tian, J. H.; Garcia, A. E. Simulation studies of protein folding/unfolding equilibrium under polar and nonpolar confinement. *J. Am. Chem. Soc.* **2011**, *133*, 15157−15164.
(12) Wang, C. Y.; Piroozan, N.; Javidpour, L.; Sahimi, M. Effect of the geometry of confining media on the stability and folding rate of α-helix proteins. *J. Chem. Phys.* **2018**, *148*, 194305.
(13) Javidpour, L.; Sahimi, M. Confinement in nanopores can destabilize α-helix folding proteins and stabilize the β structures. *J. Chem. Phys.* **2011**, *135*, 125101.
(14) Xu, G. H.; Cheng, K.; Wu, Q.; Liu, M. L.; Li, C. G. Confinement alters the structure and function of calmodulin. *Angew. Chem., Int. Ed.* **2017**, *56*, 530−534.
(15) England, J.; Lucent, D.; Pande, V. Rattling the cage: computational models of chaperonin-mediated protein folding. *Curr. Opin. Struct. Biol.* **2008**, *18*, 163−169.
(16) Lucent, D.; Vishal, V.; Pande, V. S. Protein folding under confinement: A role for solvent. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10430−10434.
(17) Macro, N.; Chen, L.; Yang, Y. S.; Mondal, T.; Wang, L. J.; Horovitz, A.; Zhong, D. P. Slowdown of water dynamics from the top to the bottom of the GroEL cavity. *J. Phys. Chem. Lett.* **2021**, *12*, 5723−5730.
(18) Rao, J. S.; Smith, M. D.; Cruz, L. The stability of a β-hairpin is altered by surface-water interactions under confinement. *J. Phys. Chem. B* **2014**, *118*, 3517−3523.
(19) Rao, J. S.; Cruz, L. Effects of confinement on the structure and dynamics of an intrinsically disordered peptide: A molecular-dynamics study. *J. Phys. Chem. B* **2013**, *117*, 3707−3719.
(20) Takagi, F.; Koga, N.; Takada, S. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 11367−11372.
(21) Furuta, T.; Fujitsuka, Y.; Chikenji, G.; Takada, S. In silico chaperonin-like cycle helps folding of proteins for structure prediction. *Biophys. J.* **2008**, *94*, 2558−2565.
(22) Sirur, A.; Knott, M.; Best, R. B. Effect of interactions with the chaperonin cavity on protein folding and misfolding. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6358−6366.
(23) Sirur, A.; Best, R. B. Effects of interactions with the GroEL cavity on protein folding rates. *Biophys. J.* **2013**, *104*, 1098−1106.
(24) Mittal, J.; Best, R. B. Thermodynamics and kinetics of protein folding under confinement. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20233−20238.
(25) Marino, K. A.; Bolhuis, P. G. Confinement-induced states in the folding landscape of the Trp-cage miniprotein. *J. Phys. Chem. B* **2012**, *116*, 11872−11880.
(26) Cheng, C.; Wu, J.; Liu, G.; Shi, S.; Chen, T. Effects of non-native interactions on frustrated proteins folding under confinement. *J. Phys. Chem. B* **2018**, *122*, 7654−7667.
(27) Du, J.; Yin, H.; Lu, Y.; Lu, T.; Chen, T. Effects of surface tethering on the thermodynamics and kinetics of frustrated protein folding. *J. Phys. Chem. B* **2022**, *126*, 4776−4786.
(28) Song, Y.; Thirumalai, D.; Hyeon, C. Moderate activity of RNA chaperone maximizes the yield of self-spliced pre-RNA in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, No. e2209422119.
(29) Chakrabarti, S.; Hyeon, C.; Ye, X.; Lorimer, G. H.; Thirumalai, D. Molecular chaperones maximize the native state yield on biological times by driving substrates out of equilibrium. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E10919−E10927.
(30) Thirumalai, D.; Klimov, D. K.; Lorimer, G. H. Caging helps proteins fold. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 11195−11197.
(31) Caraglio, M.; Pelizzola, A. Effects of confinement on thermal stability and folding kinetics in a simple Ising-like model. *Phys. Biol.* **2012**, *9*, 016006.
(32) Sangha, A. K.; Keyes, T. Protein folding and confinement: Inherent structure analysis of chaperonin action. *J. Phys. Chem. B* **2010**, *114*, 16908−16917.
(33) Wojciechowski, M.; Szymczak, P.; Cieplak, M. The influence of hydrodynamic interactions on protein dynamics in confined and crowded spaces-assessment in simple models. *Phys. Biol.* **2010**, *7*, 046011.
(34) Wang, W.; Xu, W. X.; Levy, Y.; Trizac, E.; Wolynes, P. G. Confinement effects on the kinetics and thermodynamics of protein dimerization. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 5517−5522.
(35) Lu, J. J.; Zhang, X. Y.; Wu, Y. C.; Sheng, Y. B.; Li, W. F.; Wang, W. Energy landscape remodeling mechanism of Hsp70-chaperone-accelerated protein folding. *Biophys. J.* **2021**, *120*, 1971−1983.
(36) Xu, W. X.; Wang, J.; Wang, W. Folding behavior of chaperonin-mediated substrate protein. *Proteins* **2005**, *61*, 777−794.
(37) Shental-Bechor, D.; Levy, Y. Communication: Folding of glycosylated proteins under confinement. *J. Chem. Phys.* **2011**, *135*, 141104.
(38) Zhang, S. Q.; Cheung, M. S. Manipulating biopolymer dynamics by anisotropic nanoconfinement. *Nano Lett.* **2007**, *7*, 3438−3442.
(39) Jewett, A. I.; Baumketner, A.; Shea, J.-E. Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: Creation of an alternate fast folding pathway. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13192−13197.
(40) Baumketner, A.; Jewett, A.; Shea, J.-E. Effects of confinement in chaperonin assisted protein folding: Rate enhancement by decreasing the roughness of the folding energy landscape. *J. Mol. Biol.* **2003**, *332*, 701−713.
(41) Jewett, A. I.; Shea, J.-E. Reconciling theories of chaperonin accelerated folding with experimental evidence. *Cell. Mol. Life Sci.* **2010**, *67*, 255−276.
(42) Kudva, R.; Tian, P. F.; Pardo-Avila, F.; Carroni, M.; Best, R. B.; Bernstein, H. D.; von Heijne, G. The shape of the bacterial ribosome exit tunnel affects cotranslational protein folding. *Elife* **2018**, *7*, No. e36326.
(43) Tian, P. F.; Steward, A.; Kudva, R.; Su, T.; Shilling, P. J.; Nickson, A. A.; Hollins, J. J.; Beckmann, R.; von Heijne, G.; Clarke, J.; Best, R. B. Folding pathway of an Ig domain is conserved on and off the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E11284−E11293.
(44) Nissley, D. A.; Vu, Q. V.; Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. Electrostatic interactions govern extreme nascent protein ejection times from ribosomes and can delay ribosome recycling. *J. Am. Chem. Soc.* **2020**, *142*, 6103−6110.
(45) Zhou, H. X.; Dill, K. A. Stabilization of proteins in confined spaces. *Biochemistry* **2001**, *40*, 11289−11293.
(46) Ping, G.; Yuan, J. M.; Vallieres, M.; Dong, H.; Sun, Z.; Wei, Y.; Li, F. Y.; Lin, S. H. Effect of confinement on protein folding and stability. *J. Chem. Phys.* **2003**, *118*, 8042−8048.
(47) Rathore, N.; Knotts IV, T. A.; de Pablo, J. J. Confinement effects on the thermodynamics of protein folding: Monte Carlo simulations. *Biophys. J.* **2006**, *90*, 1767−1773.
(48) Kmiecik, S.; Kolinski, A. Simulation of chaperonin effect on protein folding: A shift from nucleation-condensation to framework mechanism. *J. Am. Chem. Soc.* **2011**, *133*, 10283−10289.
(49) Piana, S.; Shaw, D. E. Atomic-level description of protein folding inside the GroEL cavity. *J. Phys. Chem. B* **2018**, *122*, 11440−11449.
(50) Caraglio, M.; Pelizzola, A. Effects of confinement on thermal stability and folding kinetics in a simple Ising−like model. *Phys. Biol.* **2012**, *9*, 016006.
(51) Taketomi, H.; Ueda, Y.; Gō, N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effects of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* **1975**, *7*, 445−459.

https://doi.org/10.1021/acs.jpclett.4c02098
J. Phys. Chem. Lett. 2024, 15, 10138−10145

(52) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.
(53) Clementi, C. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.* **2008**, *18*, 10−15.
(54) Das, P.; Wilson, C. J.; Fossati, G.; Wittung-Stafshede, P.; Matthews, K. S.; Clementi, C. Characterization of the folding landscape of monomeric lactose repressor: Quantitative comparison of theory and experiment. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14569−14574.
(55) Hills, R. D.; Brooks, C. L. Insights from coarse-grained Gō models for protein folding and dynamics. *Int. J. Mol. Sci.* **2009**, *10*, 889−905.
(56) Takada, S. Gō model revisited. *Biophys. Physicobiol.* **2019**, *16*, 248−255.
(57) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 17874−17879.
(58) Neelamraju, S.; Wales, D. J.; Gosavi, S. Protein energy landscape exploration with structure-based models. *Curr. Opin. Struct. Biol.* **2020**, *64*, 145−151.
(59) Kaya, H.; Uzunoğlu, Z.; Chan, H. S. Spatial ranges of driving forces are a key determinant of protein folding cooperativity and rate diversity. *Phys. Rev. E* **2013**, *88*, 044701.
(60) Chen, T.; Chan, H. S. Effects of desolvation barriers and sidechains on local-nonlocal coupling and chevron behaviors in coarse-grained models of protein folding. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6460−6479.
(61) Ferguson, A.; Liu, Z. R.; Chan, H. S. Desolvation barrier effects are a likely contributor to the remarkable diversity in the folding rates of small proteins. *J. Mol. Biol.* **2009**, *389*, 619−636.
(62) Cheung, M. S.; García, A. E.; Onuchic, J. N. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 685−690.
(63) Fernández-Escamilla, A. M.; Cheung, M. S.; Vega, M. C.; Wilmanns, M.; Onuchic, J. N.; Serrano, L. Solvation in protein folding analysis: Combination of theoretical and experimental approaches. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2834−2839.
(64) Liu, Z. R.; Chan, H. S. Solvation and desolvation effects in protein folding: native flexibility, kinetic cooperativity and enthalpic barriers under isostability conditions. *Phys. Biol.* **2005**, *2*, S75−S85.
(65) Chan, H. S.; Zhang, Z. Q.; Wallin, S.; Liu, Z. R. Cooperativity, local-nonlocal coupling, and nonnative interactions: Principles of protein folding from coarse-grained models. *Annu. Rev. Phys. Chem.* **2011**, *62*, 301−326.
(66) Kaya, H.; Chan, H. S. Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: How adequate is native-centric topological modeling? *J. Mol. Biol.* **2003**, *326*, 911−931.
(67) Hu, J.; Chen, T.; Wang, M.; Chan, H. S.; Zhang, Z. A critical comparison of coarse-grained structure-based approaches and atomic models of protein folding. *Phys. Chem. Phem. Phys.* **2017**, *19*, 13629−13639.
(68) Wu, J.; Cheng, C.; Liu, G.; Zhang, P.; Chen, T. The folding pathways and thermodynamics of semiflexible polymers. *J. Chem. Phys.* **2018**, *148*, 184901.
(69) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **2001**, *60*, 96−123.
(70) Di Nardo, A. A.; Larson, S. M.; Davidson, A. R. The relationship between conservation, thermodynamic stability, and function in the SH3 domain hydrophobic core. *J. Mol. Biol.* **2003**, *333*, 641−655.
(71) Zhu, Y. J.; Fu, X. R.; Wang, T.; Tamura, A.; Takada, S.; Saven, J. G.; Gai, F. Guiding the search for a protein's maximum rate of folding. *Chem. Phys.* **2004**, *307*, 99−109.
(72) Wallin, S.; Chan, H. S. Conformational entropic barriers in topology-dependent protein folding: perspectives from a simple native-centric polymer model. *J. Phys.: Condens. Matter* **2006**, *18*, S307−S328.
(73) Zuo, G. H.; Wang, J.; Wang, W. Folding with downhill behavior and low cooperativity of proteins. *Proteins* **2006**, *63*, 165−173.
(74) Especial, J. N. C.; Faísca, P. F. N. Effects of sequence-dependent non-native interactions in equilibrium and kinetic folding properties of knotted proteins. *J. Chem. Phys.* **2023**, *159*, 065101.
(75) Especial, J. N. C.; Rey, A.; Faísca, P. F. N. A note on the effects of linear topology preservation in Monte Carlo simulations of knotted proteins. *Int. J. Mol. Sci.* **2022**, *23*, 13871.
(76) de Gennes, P. G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.
(77) Suzuki, Y.; Onuchic, J. N. Modeling the interplay between geometrical and energetic effects in protein folding. *J. Phys. Chem. B* **2005**, *109*, 16503−16510.
(78) Plotkin, S. S.; Onuchic, J. N. Structural and energetic heterogeneity in protein folding. I. Theory. *J. Chem. Phys.* **2002**, *116*, 5263−5283.
(79) Bigman, L. S.; Levy, Y. Entropy-enthalpy compensation in conjugated proteins. *Chem. Phys.* **2018**, *514*, 95−105.
(80) Kaya, H.; Chan, H. S. Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: How applicable is the transition state picture to folding and unfolding? *J. Mol. Biol.* **2002**, *315*, 899−909.
(81) Bilsel, O.; Matthews, C. R. Barriers in protein folding reactions. *Adv. Protein Chem.* **2000**, *53*, 153−207.

https://doi.org/10.1021/acs.jpclett.4c02098
J. Phys. Chem. Lett. 2024, 15, 10138−10145