

SI Appendix

Jin Wang,^{1,2,3*} Ronaldo J. Oliveira,^{4,5+} Xiakun Chu,^{1,2+}
Paul Whitford,⁶ Jorge Chahine,⁴ Wei Han,²
Erkang Wang,^{1*} José N. Onuchic,^{6*} Vitor B.P. Leite^{4*}

¹ State Key Laboratory of Electroanalytical Chemistry
Changchun Institute of Applied Chemistry, Chinese Academy of Sciences
Changchun, Jilin 130022, China

² College of Physics & State Key Laboratory of Superhard Materials
Jilin University, Changchun, Jilin 130012, China

³ Department of Chemistry, Physics & Applied Mathematics
State University of New York at Stony Brook
Stony Brook, NY 11794-3400, USA

⁴ Departamento de Física - Instituto de Biociências, Letras e Ciências Exatas
Universidade Estadual Paulista
São José do Rio Preto, Brazil, 15054-000

⁵ Laboratório Nacional de Ciência e Tecnologia do Bioetanol
Centro Nacional de Pesquisa em Energia e Materiais
Campinas, SP, Brazil, 13083-970

⁶ Center for Theoretical Biological Physics
Rice University, 6100 Main, Houston, TX 77005-1827, USA.

+ Equal Contributions

*Corresponding Authors: jin.wang.1@stonybrook.edu, ekwang@ciac.jl.cn,
jonuchic@ucsd.edu, vleite@sjrp.unesp.br

Contents

1	Materials and methods	3
1.1	Structure-based model without energetic roughness	3
1.2	Obtaining the density of states	3
1.3	Calculating T_f and T_g	5
1.4	Structure-based model with energetic frustration included	7
1.5	Folding descriptors	8
1.6	Graphing Folding Funnels	8
2	Results	11
2.1	Topological roughness	11
2.1.1	Thermodynamics and kinetics with different sizes: size effect	12
2.1.2	Thermodynamics and kinetics with same size: topology effect	18
2.1.3	Folding energy landscape	23
2.1.4	Free energy profile	25
2.2	Energetic roughness	26
2.2.1	Protein: Villin headpiece subdomain (1YRF)	26
2.2.2	Protein: Albumin binding domain (1PRB)	30
2.2.3	Protein: chymotrypsin inhibitor 2 CI2 (1YPA)	34
2.2.4	Protein: ARR10-B (1IRZ)	38
2.2.5	Protein: Bubble protein (1UOY)	42

1 Materials and methods

1.1 Structure-based model without energetic roughness

To quantify the impact of topology on the effective roughness, we used a C_α structure-based model [1] to explore the density of states. For these calculations, the structure-based Hamiltonian only included stabilizing interactions between residues that are in contact in the native configuration. The Hamiltonian for a C_α structure based model with configuration Γ is given by expression:

$$\begin{aligned} H(\Gamma, \Gamma_N)_{SBM} &= \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 \\ &+ \sum_{dihedral} K_\phi^{(n)} [1 + \cos(n \times (\phi - \phi_0))] \\ &+ \sum_{i < j - 3}^{native} \epsilon(i, j) \left[5 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10} \right] \\ &+ \sum_{i < j - 3}^{non-native} \epsilon_2(i, j) \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12} \end{aligned}$$

The total energy is divided into bond stretching, angle bending, torsional and non-bonded interactions. The parameters K_r , K_θ , K_ϕ , ϵ , ϵ_2 weight the relative strength of each type of interaction. r , θ , and ϕ are the bond lengths, the angles, and the dihedral angles, with a subscript zero representing the values adopted in the native configuration, Γ_N . Nonbonded interactions are included between all residue pairs that are separated in sequence by at least three residues, and they are subdivided into native interactions and nonnative interactions. For native contacts, σ_{ij} is the distance between the C_α positions of residues i and j in the native configuration. For non-native contacts, σ_{NC} provides excluded volume repulsion with a value of 4 Å. The native contact map is built by Contacts of Structural Units (CSU) software [2]. Topology files for Gromacs [3] were generated using the smog@ctbp webserver (<http://smog.ucsd.edu>) [4]. Reduced units were used for all calculations. $\epsilon = 1.0$, $K_r = 100.0$, $K_\theta = 20.0$, $K_\phi^{(1)} = 1.0$, $K_\phi^{(3)} = 0.5$ and $\epsilon_2 = 1.0$.

1.2 Obtaining the density of states

To calculate the density of states, we used Replica Exchanged Molecular Dynamics (REMD) to ensure sufficient sampling in energy space [5]. We performed 48 parallel temperatures ranging from 0.2 \sim 2.2, where T_f was typically ≈ 1.0 . Each replica was simulated for 2×10^8 timesteps and exchanges were attempted every 100 steps. Figure S1 shows the temperature evolution of 3 replicas for protein α_3D in the simulations, which demonstrates that the replicas sample the full range of temperatures, ensuring sufficient sampling over the full range of phase space. The distributions of energy at different temperatures significantly overlap with neighboring temperatures (Figure S2A). Average acceptance ratios for swapping ranged from 20% to 50% for all proteins.

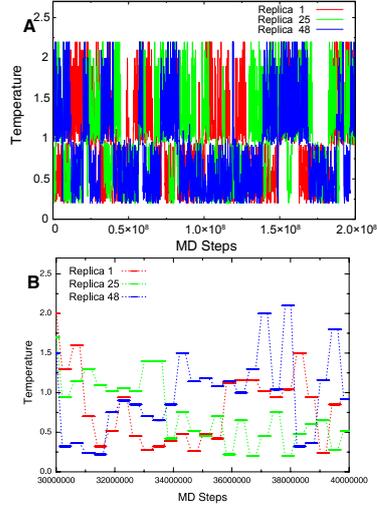


Figure S1: The temperature evolution of replicas 1, 25, 48 for protein α_3D . (A) The whole temperature exchanged evolution. Each replica covers the whole temperature range and walks randomly in temperature space. (B) Zoomed-in perspective of the temperature exchange evolution.

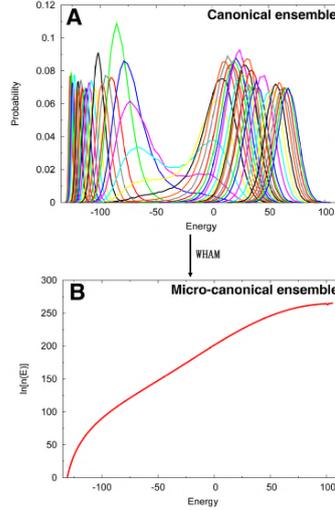


Figure S2: The energy distribution of canonical ensemble and micro-canonical ensemble for protein α_3D . (A) The energy distribution from REMD simulations. Different curves correspond to different temperatures. (B) The density of states represents the energy distribution of the micro-canonical ensemble. Using WHAM, we transform the canonical ensemble into the micro-canonical ensemble.

Each simulated replica samples a canonical ensemble. In order to probe the density of states, we convert the distributions from the canonical ensemble $n(E, T)$ to obtain the distribution of energies in the micro-canonical ensemble $n(E)$:

$$n(E, T) \sim n(E)e^{-E/k_B T}$$

This way, we directly obtain the underlying density of states and therefore the intrinsic

effective energy landscape of protein folding. Trajectories at different temperatures were collected and analyzed using the Weighted Histogram Analysis Method (WHAM) [6]. For analysis, only the non-bonded and dihedral energies were used. The bond energy terms only change and contribute at very high temperatures. However, bond stretching is less relevant to folding, so it is excluded from analysis here.

Since REMD leads to individual replicas making transitions between temperatures, the constant temperature kinetics can not be directly extracted, although there have been some efforts to do so elsewhere [7–11]. For our kinetic analysis, we simply performed 200 trajectories from different initial unfolded configurations and random velocities, for each temperature and protein. The mean first passage time (MFPT) was then calculated.

1.3 Calculating T_f and T_g

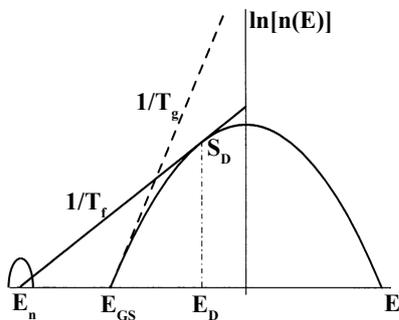


Figure S3: Schematic diagram of idealized log density of states in the random energy model (REM) [12].

Figure S3 shows a schematic diagram of the log of the density of states in a random energy model (REM) [12]. The density of states is divided into two distinct regions: native state ensemble and non-native state ensemble. The slope connecting the native ensemble E_n to the tangent of the density of state for the non-native ensemble E_D measures the folding temperature, according to the relation $1/T_f = S_D/(E_D - E_n)$. At the folding temperature T_f , the lowest energy ensemble (E_n) and the unfolded ensemble are equally probable. The slope of the curve at the ground-state energy (E_{GS}) of the non-native ensemble gives the reciprocal of the trapping transition temperature (T_g), which is the temperature at which the entropy of the system goes to zero. The foldability of proteins is ensured by the criterion $T_f > T_g$, which is clearly shown in the figure. (Figure adapted from [13] with minor modifications.)

In atomistic models, the energy gap between native state energy E_n and ground state energy of the non-native ensemble E_{GS} is more difficult to evaluate since the native state spreads from a single point to a continuous distribution. The ground state energy of non-native states, relative to the native stability (E_n and E_{GS}), has been demonstrated to determine the foldability of proteins through the use of lattice models. The large gap, which leads to more stable folding thermodynamics and faster kinetics has been extensively explored through theoretical and simulation studies [14–19]. However, when the distribution of non-native states is not considered in the evaluation of the energy gap, it is insufficient to determine the foldability [20–22]. Other lattice model simulations found a new criterion that a protein-like sequence has a large ratio of energy gap between native states and non-native states divided by the energy dispersion of non-native states [23–25]. In our work, we used a ratio

between the energy gap δE , roughness ΔE and entropy S of system to describe the folding thermodynamics and kinetics of each protein. These three quantities characterize the topography of the protein folding energy landscape in terms of the steepness, roughness and size of the funneled landscape and yield the dimensionless ratio: $\Lambda = \frac{\delta E}{\Delta E \sqrt{2S}}$ [14, 15, 26–30]. δE denotes the energy gap between the native and non-native ensembles, ΔE is the variance in energies of the non-native states and S is the entropy of the non-native ensemble.

It is challenging, yet important, to define the location of the native and non-native basins within the density of states. We use the deviation of each configuration from the native structure to identify the non-native ensemble. The RMSD criterion is proposed in our analysis instead of fraction of the native contacts Q , because Q mainly monitors tertiary structure changes, while the secondary structure arrangement is mostly related to the rotation of torsional angles. We first plot the 1D free energy profile along Q at folding temperature; then, we find the location of minimal free energy within the native ensemble; finally, we extract the mean RMSD value: $RMSD_C$. For one-state downhill folders, we set $RMSD_C = 0.15$. When the structure has $RMSD > RMSD_C$, it is referred to as a non-native configuration. Then, we can divide the density of states into the native ensemble and non-native ensemble. The non-native ensemble can provide information on the trapping transition temperature T_g and entropy S , and therefore the topological landscape roughness $\Delta E_{Top} = \sqrt{2ST_g}$ (Figure S4 and Figure S5). The energy gap δE is calculated from the difference between native ensemble energy and the average energy of the non-native configurations. Accordingly, we then obtain $\Lambda = \frac{\delta E}{\Delta E_{Top} \sqrt{2S}}$.

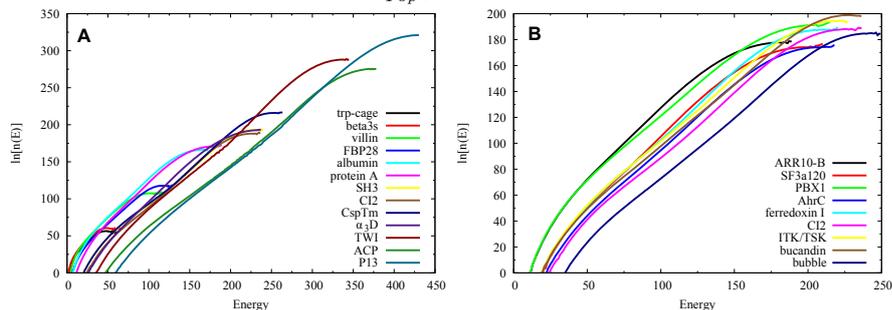


Figure S4: Logarithm of the density of states of the non-native ensemble, the energy landscape funnel in one dimension (1d-funnel), as a function of energy for the (A) 13 proteins with different sizes and (B) 9 proteins with same size. The ground states of the non-native ensembles, as well as the inverse slope of the log density of states (trapping transition temperature T_g), are identified for each protein.

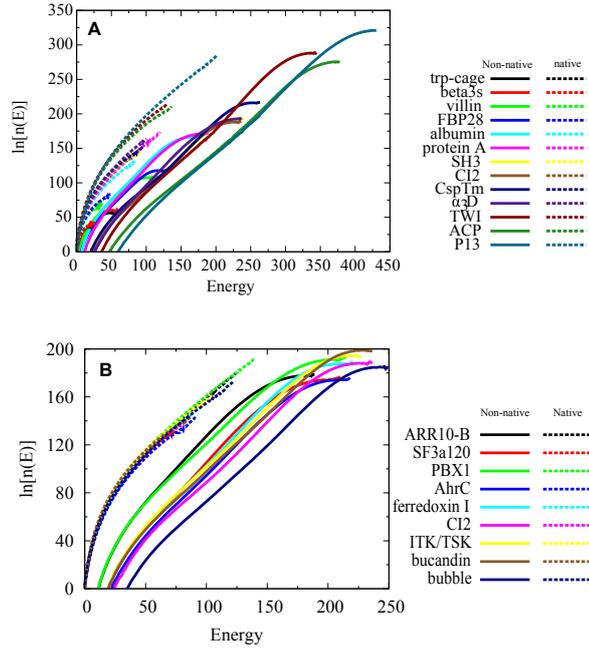


Figure S5: Logarithm of the density of states of the native and non-native ensembles. The energy landscape funnel in one dimension (1d-funnel), as a function of energy for (A) 13 proteins with different sizes and (B) 9 proteins with same size.

1.4 Structure-based model with energetic frustration included

Energetic roughness is incorporated into the model by adding non-native attractive interactions with heterogeneous strengths into the structure-based model. The Hamiltonian of a configuration Γ is then given by

$$H(\Gamma, \Gamma_N) = H(\Gamma, \Gamma_N)_{SBM} + \sum_{non-native(i,j)} V_{NN}(r_{ij}, \eta_b)$$

where $V_{NN}(r_{ij}, \eta_b)$ is of the form [31, 32]:

$$V_{NN}(r_{ij}, \eta_b) = \begin{cases} \eta_{ij} [1 - \frac{1}{2} (\frac{r_{ij}}{r_N})^{20}] & \text{if } r_{ij} < r_N \\ \frac{\eta_{ij}}{2} (\frac{r_N}{r_{ij}})^{20} & \text{if } r_{ij} \geq r_N \end{cases}$$

where the strength η_{ij} for non-native pair (i,j) is randomly assigned from a Gaussian distribution with mean ϵ_{NN} and variance b^2 . When $\epsilon_{NN} = 0$ and $b^2 = 0$, the model simplifies to a purely structure-based model. In our simulations, the parameter $r_N = \frac{4}{3}\sigma_{NC}$, σ_{NC} represents the excluded volume repulsion used in the standard structure based model. The mean of non-native interactions ϵ_{NN} was set to 0, and we scaled the variance b^2 to change the magnitude of the stabilizing non-native interactions.

In an analytical model of protein folding, Bryngelson et al. proposed that energetic roughness will lead to a glass trapping temperature (T_g), at which the protein becomes frozen in a low energy configuration, the entropy vanishes and the kinetics become very slow. The trapping transition temperature (T_g) induced by energetic roughness

was determined as: $T_g = \sqrt{\Delta E_{Ene}^2/2S}$, where ΔE_{Ene} is the energetic fluctuations in non-native states and S is the entropy of non-native states. The energetic roughness introduced by the non-native interactions is expressed as $\sqrt{A_{max}b^2}$, where A_{max} is the maximum number of non-native contacts the protein can form.

The landscape roughness of proteins ΔE_{Total} is contributed both by topological and energetic roughness. With the assumption that altering energetic roughness does not change topological roughness, we can calculate $\Delta E_{Total}^2 = \Delta E_{Top}^2 + \Delta E_{Ene}^2$. The trapping transition temperature (T_g) will be: $T_g = \sqrt{\Delta E_{Total}^2/2S}$. So the folding landscape measure $\Lambda = \delta E/(\Delta E_{Total}\sqrt{2S})$. The energy gap δE , entropy S can be calculated directly from density of states.

1.5 Folding descriptors

To measure the foldability of each protein, Thirumalai and coworkers proposed an intrinsic thermodynamic parameter $\sigma = (T_\theta - T_f)/T_\theta$, where T_θ is the polymer collapse temperature and T_f is the folding temperature [20, 22, 33–35]. T_θ , considered to be the temperature at which the protein collapses from a random coil to a compact structure [36], is measured from the peak of temperature-dependent specific heat curve. To quantify T_f , Thirumalai et al. introduced an “overlap” function, which measures the similarity of each configuration to the native state.

$$\chi = 1 - \frac{1}{N^2 - 3N + 2} < \sum_{i \neq j, j \pm 1} \delta(r_{ij} - r_{ij}^0) >$$

where the superscript 0 refers to the native state and N is the residue number. Thus the folding temperature T_f corresponds to the peak of temperature-dependent fluctuation $\Delta\chi$. In structure based model, the collapsed and folding process are often close to each other, such that $\sigma \approx 0$ [37]. Therefore in this work, σ is not used for the topological roughness analysis.

For the analysis of the kinetics of folding, it is necessary to perform simulations under consistent conditions for each proteins. Kinetic simulations were performed at the temperature which $\chi = 0.2$ (T_χ) for each protein. T_f is the folding temperature, where the probability of the protein being folded, or unfolded, is equal. T_χ can be interpreted as at which the folding population equals 80%, which mimics folding under physiological conditions.

1.6 Graphing Folding Funnels

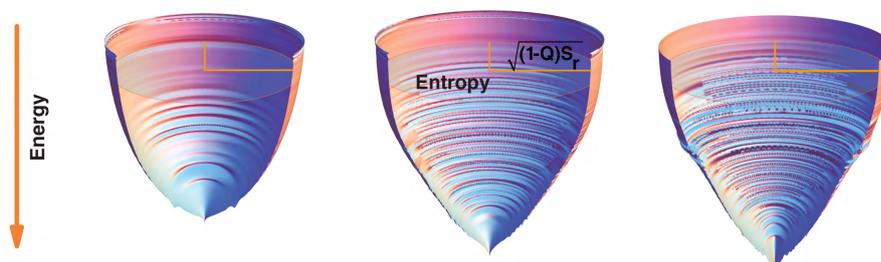


Figure S6: The realistic folding funnels for villin headpiece(left), CI2(middle), P13(right).

Figure S6 shows the three folding energy landscapes, calculated directly from simulations. The funnels are graphed in the following way. The vertical axis corresponds to energy. For each energy, the average Q , $\langle Q(E) \rangle$ is calculated. The semi-major axis of the ellipsoidal cross section is set to $\sqrt{S_r(1-Q)}$. The other semi-axis of the cross section is then set, such that the area of the cross section is equal to the entropy $S(E)$. Then we shift $S(Q)$, $S(Q)=0$. That is, the native configuration is depicted as having an entropy of zero.

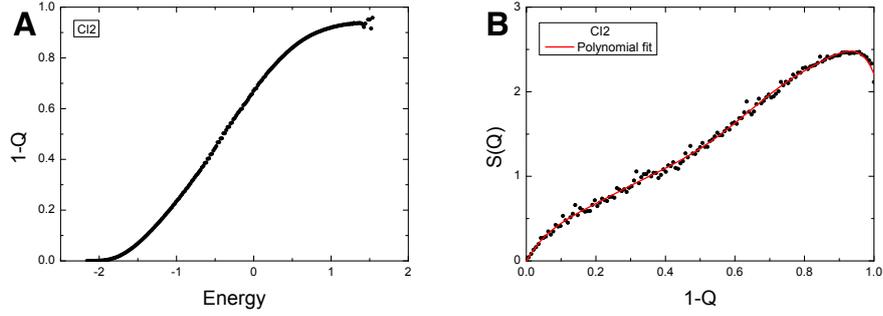


Figure S7: The relationship used to plot the realistic funnels. (A) The relationship between $1 - \langle Q(E) \rangle$ and energy. (B) The relationship between $1-Q$ and $S(Q)$. The red line is the polynomial fit of the grid data. Energy and entropy has been scaled by protein size N . The relation shown here is similar for the other 12 proteins.

The roughness in each funnel is different for each protein (Figure S6). To depict the relationship between the bumpiness on the funnel and the roughness of the energy landscape, $S(Q)$ was plotted against $1-Q$ and then fitted to a polynomial function (Figure S7B). Then mean square error (MSE) between the $S(Q)$ (i.e. the rugged funnel) and the smooth fit measures the “bumpiness” of the funnel, which we will show to be related to the roughness for specific cases (Figure S7)

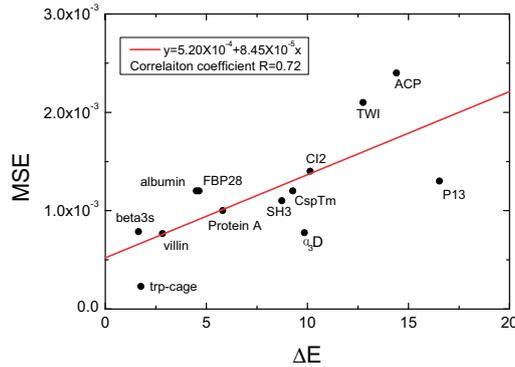


Figure S8: The relationship between the MSE and the roughness of the energy landscape. The abbreviations of proteins are labeled beside the data.

The bumpiness of the funnels monotonically correlates with the roughness of the energy landscapes for proteins of different sizes (Figure S8). As expected, as the roughness increases, the funnels becomes more rugged. However, we did not find a

correlation between the bumpiness of the funnels and the roughness of the landscapes, when comparing proteins of the same size. Since the bumpiness of the funnels and the roughness both correlated with protein size, their correlation is due to the size dependence. In other words, the bumpiness of the funnels can be a representation of landscape roughness for proteins with different sizes while it may not be applicable for proteins with same size.

2 Results

The Temperature Dependence of Protein Folding Rates.

The kinetics of protein folding is highly dependent on the environment, especially temperature in simulations. In Figure S14 and Figure S20, we see the kinetics of folding of different proteins with respect to temperatures with the different sizes and with same size but different structural topologies. For the proteins studied here, there is a U-shape dependence of the mean first passage time (MFPT) τ , when plotted as a function of a) absolute simulation temperature T , or b) temperatures normalized by $T = T_0$. T_0 is the optimal temperature for folding (i.e., the temperature where the protein folding time τ is minimal.). The U shape dependence of folding kinetic rate versus temperature can be understood in terms of thermodynamic considerations. At higher temperatures, the non-native states are preferred over the folded states, making more difficult to reach folded configurations. In contrast, when the temperature is lower, the native state are preferred, however the traps associated with folding become prominent, which slows down the kinetics. Therefore, there is an optimal temperature for each protein where the kinetic rate is minimal, which has also been observed experimentally [21, 27, 41, 42, 45–47, 49, 51–68].

Each protein has specific thermodynamic stability and therefore a unique folding temperature T_f . The kinetic rates measured in experiments are often at different temperatures near or below their folding temperatures. For the analysis of the kinetics of folding, it is necessary to perform simulations under consistent conditions for each protein. A key question is: What is the appropriate reference temperature for comparing the kinetic behavior of proteins? We obviate the need for an absolute temperature scale for comparison, here, we choose the T_χ , which is the temperature at which the native ensemble of each protein accounts for 80% of the population for measuring the kinetic rate for each protein [22]. This is a reasonable choice of the temperatures for comparison of the intrinsic folding kinetics of different proteins. In addition, T_χ being lower than T_f can be interpreted as mimicking folding under physiological conditions, our results thereby have a strong connection to the experimental measurements.

2.1 Topological roughness

The following figures and tables provide the details of all calculated quantities and detailed descriptions of each protein studied.

2.1.1 Thermodynamics and kinetics with different sizes: size effect

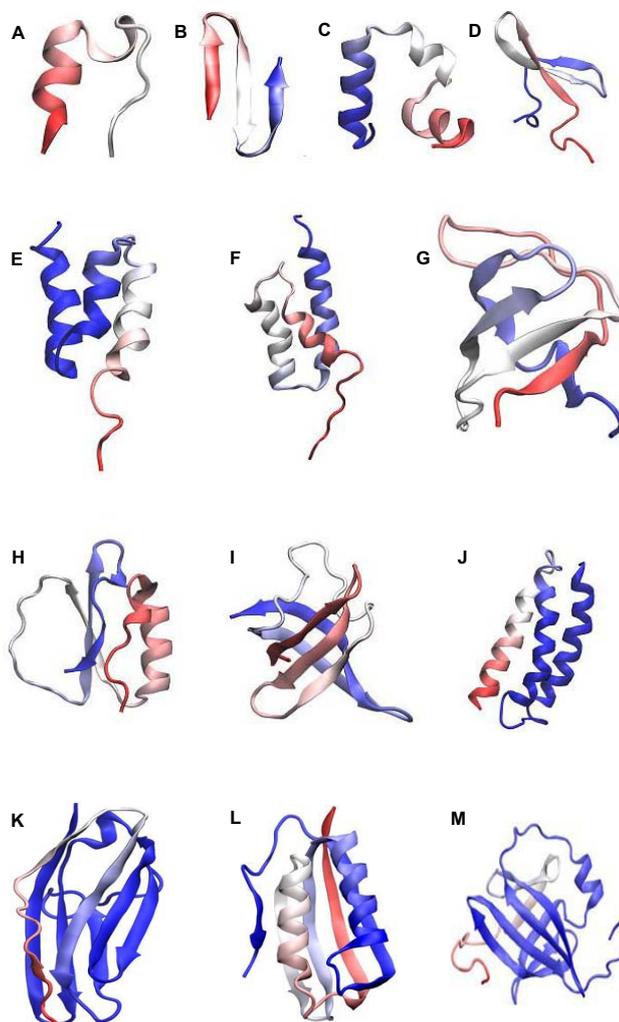


Figure S9: The native structure of 13 proteins with different sizes used in this study for the folding simulations are shown in cartoon representation with its Protein Data Base code (PDB) (A) tryptophan cage (1L2Y) (B) beta3s B3S (no PDB code) (C) villin headpiece subdomain (1YRF) (D) ww domain FBP28 (1E0L) (E) albumin binding domain (1PRB) (F) protein A (1BDD) (G) src homology 3 domain SH3 (1FMK) (H) chymotrypsin inhibitor 2 CI2 (1YPA) (I) cold shock protein CspTm (1G6P) (J) α_3D (2A3D) (K) twitchin 18th igsf module TWI (1WIT) (L) acylphosphatase ACP (2VH7) (M) oncoprotein P13MTCP1 P13 (1QTU). The structures are created using the package Visual Molecular Dynamics (VMD) [38] and are colored by an index along the chain from red (N-terminus) to blue (C-terminus).

Table S1: Quantities extracted from the density of states for the 13 proteins with different sizes.

Protein	Trp-cage	Beta3s	Villin	FBP28	Albumin	Protein A	SH3	CI2	CspTm	α_3D	TWI	ACP	P13
PDB	1L2Y	(no PDB)	1YRF	1E0L	1PRB	1BDD	1FMK	1YPA	1G6P	2A3D	1WIT	2VH7	1QTU
N	20	20	35	37	53	60	61	64	66	73	93	94	115
Q	23	24	56	70	94	98	148	142	166	136	217	248	278
RCO	0.39	0.36	0.23	0.38	0.28	0.25	0.38	0.35	0.33	0.27	0.38	0.39	0.28
ACO	7.87	7.21	8.07	14.07	14.96	14.95	23.01	22.25	21.85	19.60	35.02	36.94	32.31
RMSD Criterion	0.15	0.15	0.15	0.15	0.26	0.43	0.18	0.15	0.14	0.25	0.17	0.16	0.37
E_n	-22.54	-23.51	-54.38	-68.41	-90.96	-94.42	-144.07	-137.89	-159.88	-130.95	-206.61	-239.82	-266.76
E_{GS}	-19.98	-21.41	-49.98	-62.01	-84.64	-82.94	-124.09	-113.86	-139.33	-105.20	-174.18	-194.46	-207.55
δE	48.47	49.86	104.31	118.56	164.09	180.51	229.64	228.17	254.22	231.09	338.56	370.72	425.03
ΔE	1.76	1.64	2.83	4.64	4.51	5.81	8.74	10.14	9.28	9.85	12.76	14.41	16.53
S_0 (a)	58.51	64.12	103.91	116.45	162.82	160.33	192.27	166.82	211.95	178.73	286.43	267.59	303.89
$\frac{\delta E}{\sqrt{2S_0}}$	4.48	4.40	7.24	7.77	9.09	10.08	11.71	12.49	12.35	12.22	14.15	16.02	17.24
$\frac{\delta E}{\Delta E}$	27.53	30.31	36.83	25.56	36.37	31.08	26.28	22.51	27.40	23.45	26.54	25.73	25.71
T_f (b)	0.82	0.79	0.89	0.98	0.97	0.95	1.11	1.03	1.08	0.93	1.02	1.14	1.09
T_g	0.16	0.15	0.20	0.30	0.25	0.32	0.45	0.55	0.45	0.52	0.53	0.62	0.67
$\frac{T_f}{T_g}$	5.03	5.46	4.54	3.23	3.89	2.94	2.49	1.85	2.39	1.79	1.92	1.83	1.63
Λ	2.54	2.67	2.55	1.67	2.01	1.73	1.34	1.23	1.33	1.24	1.10	1.11	1.04
T_0 (c)	0.59	0.57	0.63	0.67	0.80	0.76	0.88	0.89	0.78	0.83	0.81	0.80	0.80
$\frac{T_0}{T_g}$	3.59	3.91	3.22	2.19	3.18	2.35	1.97	1.61	1.73	1.59	1.52	1.27	1.18
T_χ (d)	0.75	0.74	0.78	0.93	0.92	0.84	1.09	1.02	1.07	0.89	1.01	1.13	1.07
$\frac{T_\chi}{T_g}$	4.63	5.07	3.96	3.07	3.70	2.60	2.45	1.83	2.37	1.72	1.90	1.82	1.59

(a) $S = \sum_E \ln(n(E, Q))$, $n(E, Q)$ is the density of states. S_0 is the configurational entropy at Q_{Min} .

(b) T_f is calculated from the peak of heat capacity curve since the collapsed temperature (T_θ) is same with folding temperature in structure based model [20, 22, 34, 35, 37].

(c) T_0 is the optimum temperature at which the folding of proteins will be the fastest.

(d) T_χ is the temperature where $\chi = 0.2$ and population of native folding is 80%.

(#) Temperature is in the unit of energy scale by multiplying Boltzmann constant k . Length and energy scale are in the reduced units of Gromacs [3, 4].

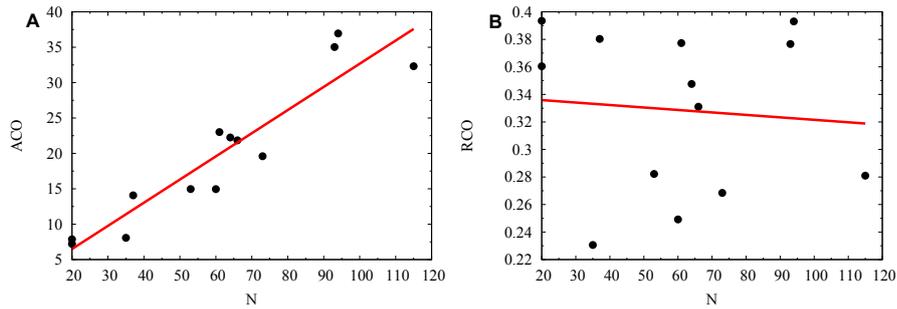


Figure S10: Topology of the native structure varied with different protein sizes. Fits of (A) absolute contact order (ACO), (B) relative contact order (RCO) as a function of protein size N for the 13 proteins. (A) The straight line represents $y = -0.06 + 0.33x$ with correlation coefficient 0.92. (B) The straight line represents $y = 0.33 - 1.8 \times 10^{-4}x$ with correlation coefficient -0.09.

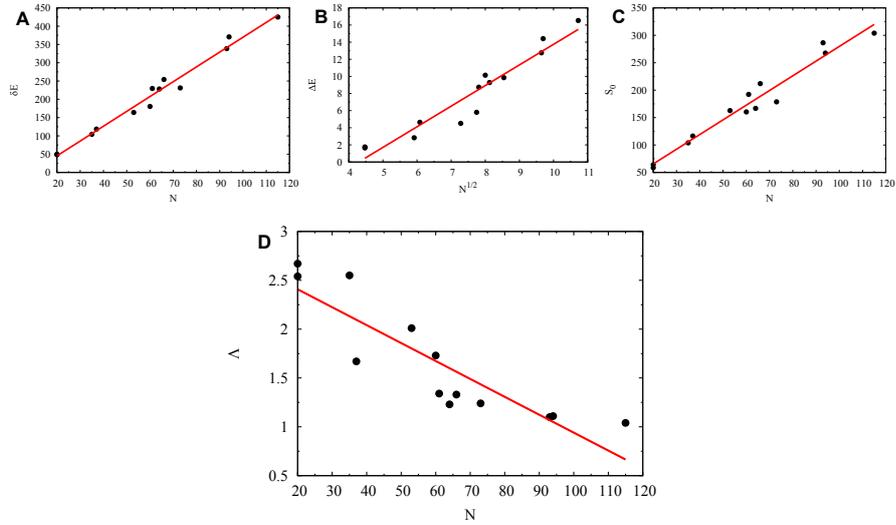


Figure S11: The topography of energy landscape varied with different protein sizes. Fits of (A) energy gap (δE), (C) entropy (S_0), (D) Λ as a function of protein size N and (B) energy standard deviation as a function of $N^{1/2}$ for the 13 different proteins. (A) The straight line represents $y=-35.72+4.06x$ with correlation coefficient 0.99. (B) The straight line represents $y=-10.33+2.41x$ with correlation coefficient 0.96. (C) The straight line represents $y=11.96+2.68x$ with correlation coefficient 0.98. (D) The straight line represents $y=2.77-0.02x$ with correlation coefficient -0.88.

The energy gap δE and entropy S are strongly correlated with protein size N and energetic roughness ΔE scales with the square of protein size $N^{1/2}$ (Figure S11), which is expected from analytic theories [28, 30]. We also find that the dimensionless ratio Λ , which measures of the underlying landscape topography, is also dependent on protein size. In mean field theory, Λ should not explicitly depend on the size of the protein. The size dependence may be due to the surface effects of finite sized proteins, which is not accounted for in mean field theories that assume the infinite size [39].

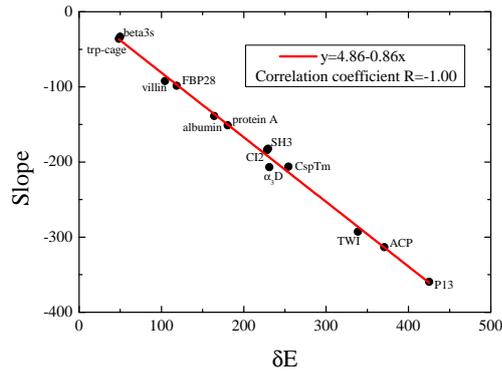


Figure S12: Correlation between the slope of energy landscape and the energy gap. Fits the average slope of energy landscape as a function of energy gap δE the 13 proteins with different sizes with correlation coefficient -1.00

Analytical energy landscape theoretical investigations of protein folding have previously shown that the energy gap is strongly correlated with the slope of energy

versus Q [40]. In Figure S12, we show the correlation between the average slope and the energy gap is strong in our simulations.

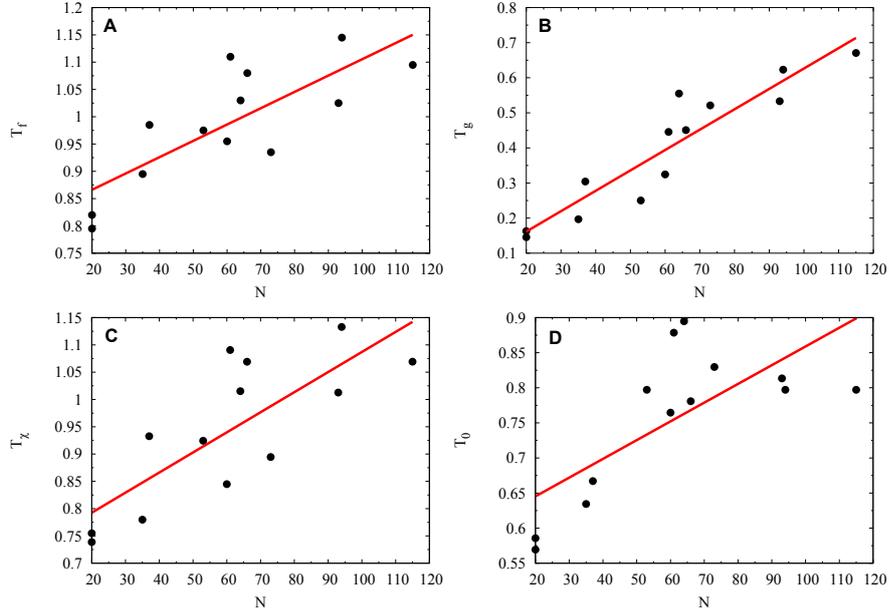


Figure S13: The typical temperature varied with different protein sizes. Fits of (A) folding temperature T_f , (B) glass trapping transition temperature T_g , (C) T_χ of 80% native population and (D) optimum kinetic temperature T_0 as a function of protein size N for the 13 different proteins. (A) The straight line represents $y=0.81+0.003x$ with correlation coefficient 0.80. (B) The straight line represents $y=0.05+0.006x$ with correlation coefficient 0.93. (C) The straight line represents $y=0.72+0.004x$ with correlation coefficient 0.79. (D) The straight line represents $y=0.59+0.003x$ with correlation coefficient 0.72.

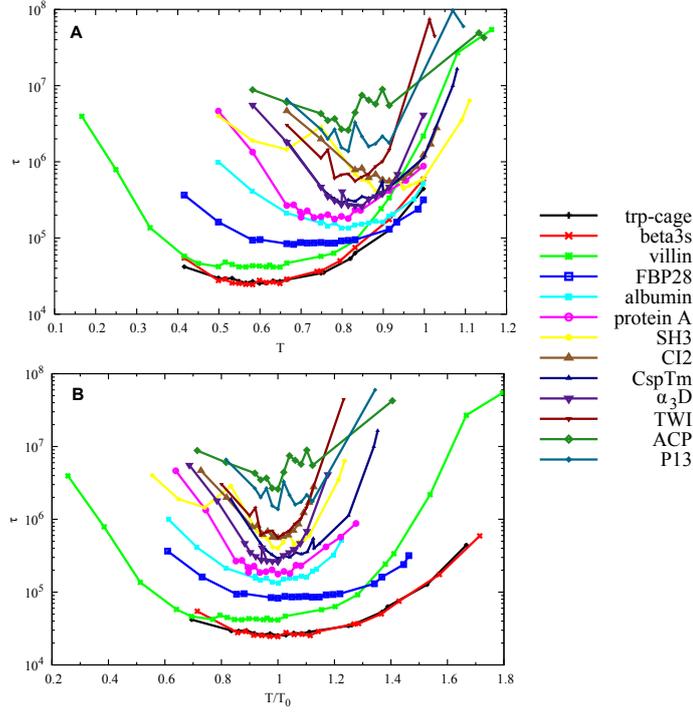


Figure S14: Folding time τ as a function of (A) absolute simulation temperature T and (B) normalized temperature T/T_0 for the 13 proteins with different sizes. The time is in the unit of MD step.

In Figure S14, we see a “U-shape” behavior of the folding time versus temperature. Long folding time at high temperature is due to the unstable nature of folded configurations at high T . Long time folding at low temperature is due to the roughness of the landscape. This behavior has been widely observed in the kinetic folding experiments and simulations of different proteins [30, 41–49].

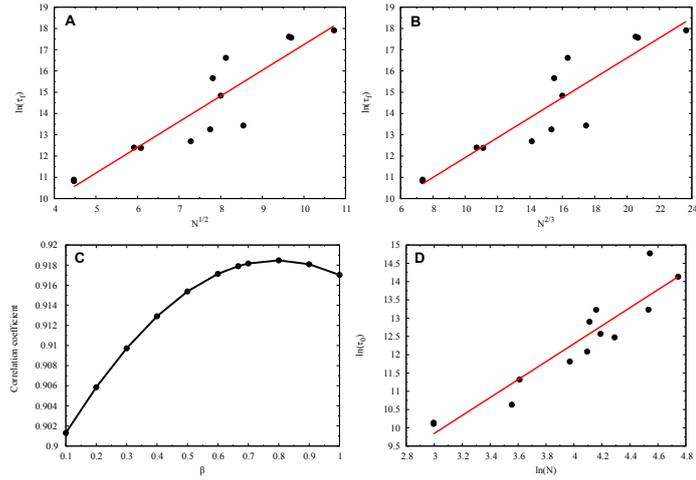


Figure S15: Folding time as a function of protein sizes. (A) Fits $\ln\tau_f$ as a function of $N^{1/2}$ with the straight line $y=5.15+1.21x$ and the correlation coefficient is 0.92.

(B) Fits $\ln\tau_f$ as a function of $N^{2/3}$ with the straight line $y=7.24+0.47x$ and the correlation coefficient is 0.92. (C) Fits of $\ln\tau_f \sim N^\beta$ with straight line and plot the correlation coefficient as a function of β . Because of the limit range of protein size N , it is difficult to determinate a single superscript β but a range of reasonable values. (D) Fits $\ln\tau_0$ as a function of $\ln N$ representing a relationship: $\ln\tau_0 \sim N^{2.48}$ with correlation coefficient is 0.94. τ_f and τ_0 are the folding time at folding temperature T_f and optimum temperature T_0 .

The folding rate is correlated with protein size. Our results in Figure S15 are consistent with the previous theories and simulations [34, 39, 50, 51, 69–77].

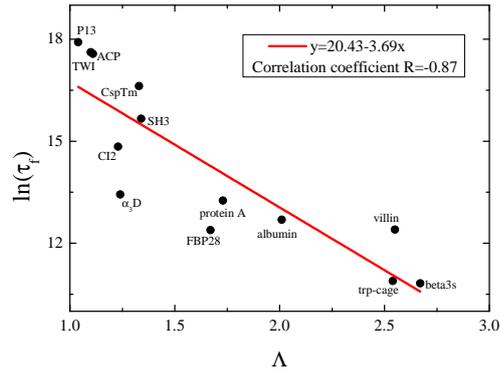


Figure S16: Folding time at folding temperature as a function of Λ . The correlation is strong with correlation coefficient -0.87

Figure S16 shows that folding time at the folding temperature is strongly correlated with Λ . It is to be expected since both $\ln(\tau_f)$ and Λ are dependent on protein size N .

2.1.2 Thermodynamics and kinetics with same size: topology effect

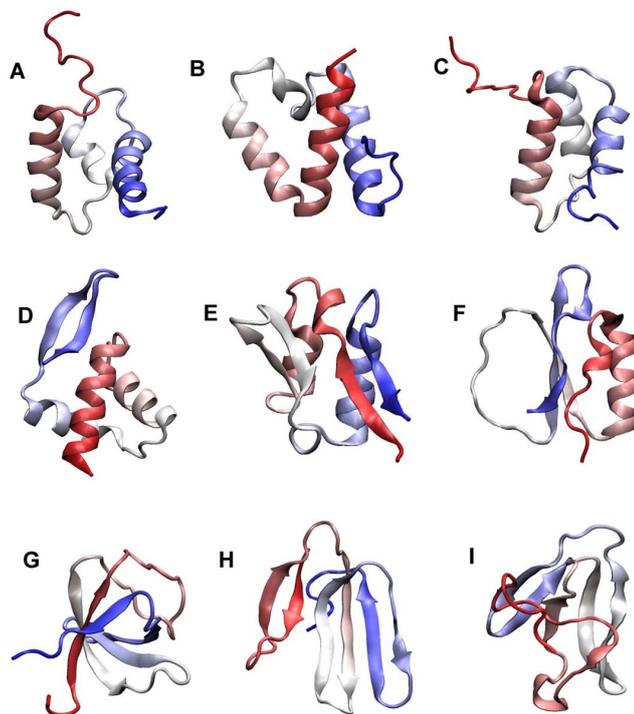


Figure S17: The native structure of 9 proteins with same size used in this study for the folding simulations are shown in cartoon representation with its Protein Data Base code (PDB). (A) ARR10-B (1IRZ) (B) splicing factor SF3a120 (2DT6) (C) homeobox protein PBX1 (1DU6) (D) N-terminal domain of AhrC (2P5K) (E) ferredoxin I (1DFD) (F) chymotrypsin inhibitor 2 CI2 (1YPA) (G) tyrosine-protein kinase ITK/TSK (2RN8) (H) bucandin (1IJC) (I) bubble protein (1UOY). The 9 proteins can be classified into 3 categories by structure topology. (A-C) correspond to all α topology, (D-F) correspond to α/β topology, (G-H) correspond to all β topology. The structures are created using the package Visual Molecular Dynamics (VMD) [38] and are colored by an index along the chain from red (N-terminus) to blue (C-terminus).

Table S2: Quantities extracted from the density of states for the 9 proteins with the same size but different structural topologies.

Protein	ARR10-B	SF3a120	PBX1	AhrC	ferredoxin I	CI2	ITK/TSK	bucandin	bubble
PDB	1IRZ	2DT6	1DU6	2P5K	1DFD	1YPA	2RN8	1IJC	1UOY
N	64	64	64	63	64	64	64	63	64
Q	95	110	116	122	123	142	134	144	157
RCO	0.22	0.30	0.26	0.31	0.38	0.35	0.35	0.35	0.28
ACO	13.80	18.93	16.78	19.75	24.63	22.25	22.66	22.00	18.09
RMSD Criterion	0.49	0.21	0.44	0.20	0.20	0.15	0.29	0.16	0.26
E_n	-90.88	-105.77	-112.18	-117.60	-117.86	-137.89	-129.45	-139.89	-151.85
E_{GS}	-79.92	-86.99	-101.89	-98.25	-98.67	-113.86	-110.90	-123.44	-116.92
δE	180.66	203.34	207.23	210.81	211.80	228.17	218.16	226.59	240.83
ΔE	5.54	7.29	5.64	8.40	7.76	10.14	7.85	8.58	10.84
S_0 (a)	174.66	168.05	183.38	162.45	181.67	166.82	193.83	199.70	182.66
$\frac{\delta E}{\sqrt{2S_0}}$	9.67	11.09	10.82	11.70	11.11	12.49	11.08	11.34	12.60
$\frac{\delta E}{\Delta E}$	32.60	27.89	36.73	25.08	27.28	22.51	27.81	26.40	22.23
T_f (b)	0.92	0.92	1.04	0.99	0.97	1.03	1.04	1.05	1.06
T_g	0.30	0.40	0.29	0.47	0.41	0.55	0.40	0.43	0.57
$\frac{T_f}{T_g}$	3.12	2.31	3.54	2.13	2.39	1.86	2.61	2.46	1.88
Λ	1.74	1.52	1.91	1.39	1.43	1.23	1.41	1.32	1.16
T_0 (c)	0.70	0.60	0.89	0.80	0.80	0.89	0.85	0.96	0.72
$\frac{T_0}{T_g}$	2.36	1.50	3.03	1.71	1.96	1.61	2.13	2.24	1.26
T_χ (d)	0.70	0.89	0.89	0.96	0.95	1.02	1.00	1.03	1.04
$\frac{T_\chi}{T_g}$	2.36	2.24	3.02	2.06	2.33	1.83	2.51	2.40	1.83

(a) $S = \sum_E \ln(n(E, Q))$, $n(E, Q)$ is the density of states. S_0 is the configurational entropy at Q_{Min} .

(b) T_f is calculated from the peak of heat capacity curve as the collapsed temperature (T_θ) is same with folding temperature in structure based model [20, 22, 34, 35, 37].

(c) T_0 is the optimum temperature at which the folding of proteins will be the fastest.

(d) T_χ is the temperature where $\chi = 0.2$ and native folding population is 80%.

(#) Temperature is in the unit of energy scale by multiplying Boltzmann constant k . Length and energy scale are in the reduced units of Gromacs [3, 4].

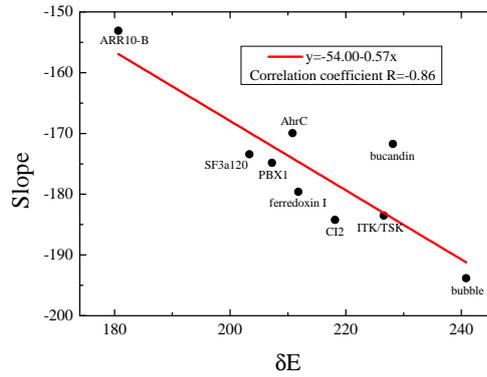


Figure S18: Correlation between the slope of energy landscape and the energy gap. Fits the average slope of energy landscape as a function of energy gap δE for 9 proteins

of same size with correlation coefficient -0.86

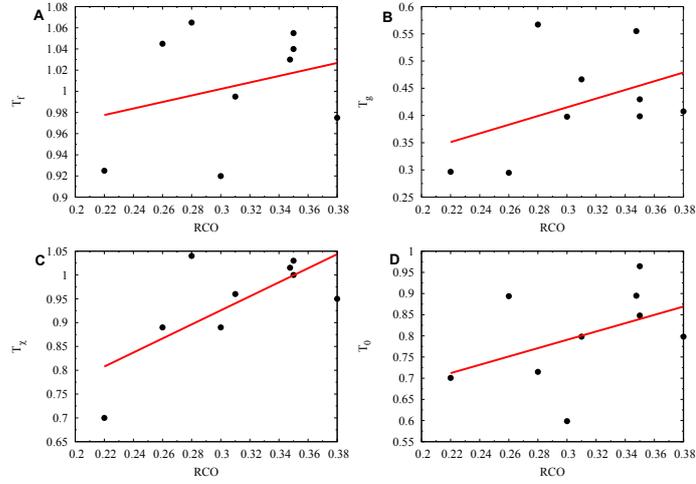


Figure S19: The typical temperatures varied with different topology. Fits of temperature as a function of relative contact order (RCO) for the 9 different proteins with same size. (A) Folding temperature T_f . The straight line is $y=0.91+0.31x$ and the correlation coefficient is 0.29. (B) Glass trapping transition temperature T_g . The straight line is $y=0.18+0.80x$ and the correlation coefficient is 0.42. (C) T_χ . The straight line is $y=0.48+1.47x$ and the correlation coefficient is 0.71. (D) Optimum temperature T_0 . The straight line is $y=0.50+0.98x$ and the correlation coefficient is 0.44.

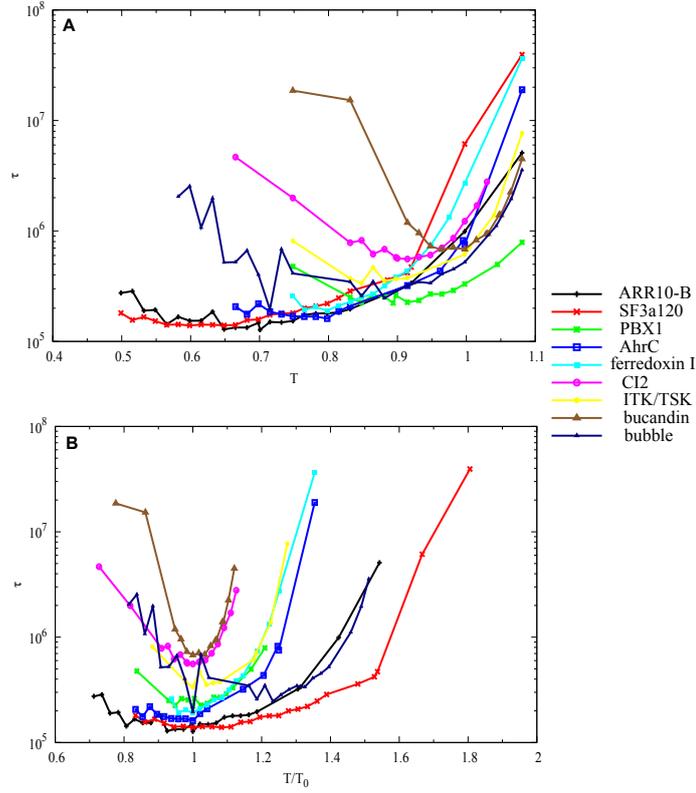


Figure S20: Folding time τ as a function of (A) absolute simulation temperature T and (B) normalized temperature T/T_0 for the 9 proteins with same size. The time is in the unit of MD step.

The folding times of different sequences of the same size with respect to temperatures are shown in Figure S20. The similar U shape dependence is observed as in Figure S14.

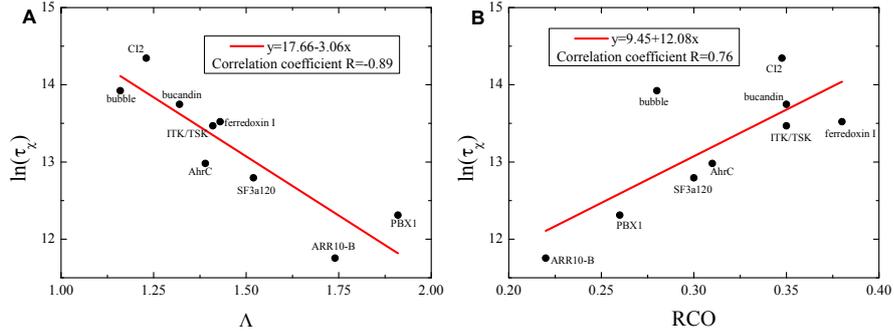


Figure S21: Folding time $\ln\tau_\chi$ at temperature T_χ as a function of (A) Λ and (B) relative contact order (RCO). (A) Fits $\ln\tau_\chi$ as a function of Λ with straight line and the correlation coefficient is -0.89. (B) Fits $\ln\tau_\chi$ as a function of RCO with straight line and the correlation coefficient is 0.76.

Figure S21 shows that Λ has a stronger correlation with the folding times than RCO does. Contact order has been considered as an important factor in determining

folding kinetics [40, 50, 78–81]. Compared with RCO, the energy landscape topography measure Λ , which includes both energetic and topological factors rather than entropy alone, is a more accurate predictor of the folding kinetics.

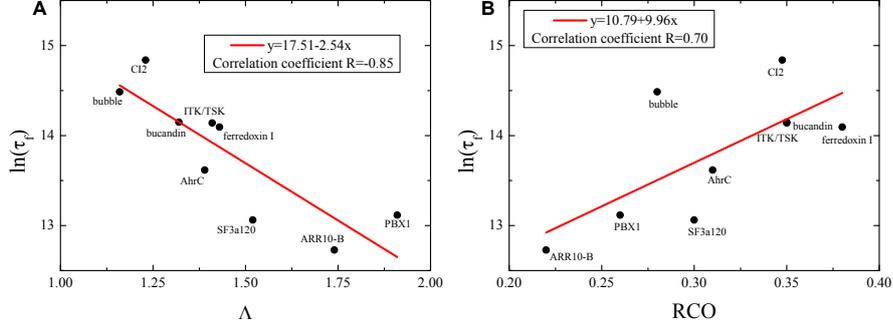


Figure S22: Folding time at folding temperature correlates with Λ and RCO. (A) Fit of $\ln(\tau_f)$ as a function of Λ with correlation coefficient -0.85. (B) Fit of $\ln(\tau_f)$ as a function of RCO with correlation coefficient 0.70.

Figure S22 shows that folding time at folding temperature is strongly correlated with Λ and RCO. Since Λ has a stronger correlation with the folding time than RCO does, it is an accurate predictor of the folding time at folding temperature.

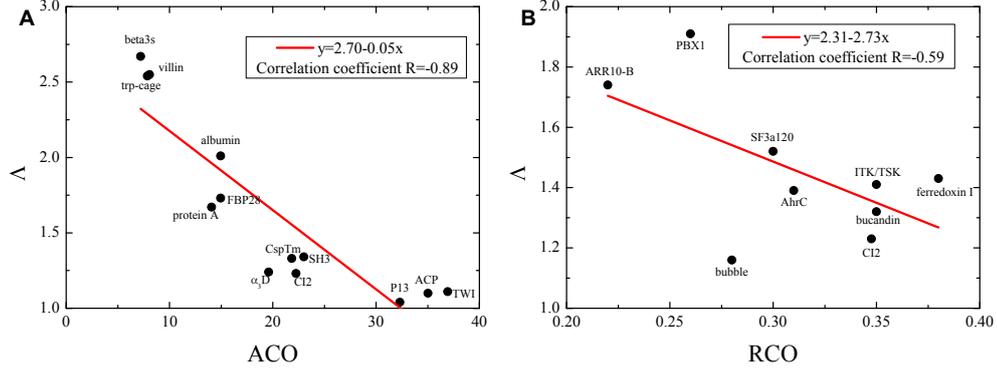


Figure S23: The relationship between contact order and Λ obtained from the topological roughness without energetic roughness. (A) Λ versus the absolute contact order (ACO) for 13 proteins with different sizes. (B) Λ versus the relative contact order (RCO) for 9 proteins with the same size but different structural topologies. The names of proteins are shown.

In order to see the relationship between the topology of the structure and Λ , we show absolute contact order (ACO) for proteins with different sizes and relative contact order (RCO) for proteins with the same size in Figure S23. The correlation between contact order and Λ implies that Λ contains the structural topological information of proteins. Λ includes not only entropic factor but also energetic contributions, such as the energy gap and roughness. It can also quantitatively describe the folding thermodynamics and kinetics when the protein folding is not fast and is not determined by topology alone.

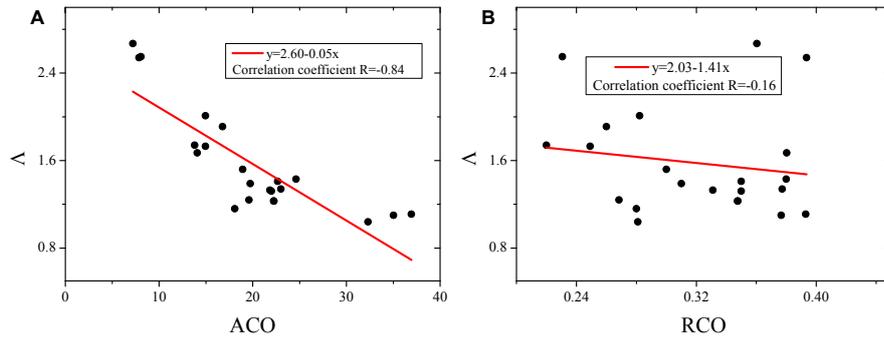


Figure S24: Correlation between contact order and Λ for total proteins with 13 different sizes and 9 same size. (A) Fits Λ as a function of absolute contact order (ACO) with correlation coefficient -0.84. (B) Fits Λ as a function of relative contact order (RCO) with correlation coefficient -0.16.

Figure S24 indicates that Λ has a strong correlation with ACO, which is correlated with protein size. The relationship between Λ and RCO is poor, and RCO shows little correlation with protein size.

2.1.3 Folding energy landscape

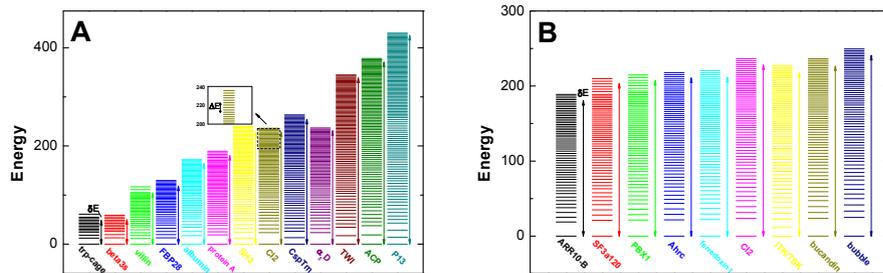


Figure S25: Folding energy landscape in zero dimension. The distribution of the energy levels (spectrum) for (A) 13 proteins with different sizes and (B) 9 proteins with the same size. The lowest (native) energy E_n set to 0 for visualization purposes. The stability gap δE is indicated by vertical arrows. Each energy level of the distribution represents the sum of a cluster of states, except for the native band. The inset is a magnification of the energy levels of CI2. The energetic roughness ΔE is also indicated by vertical arrows.

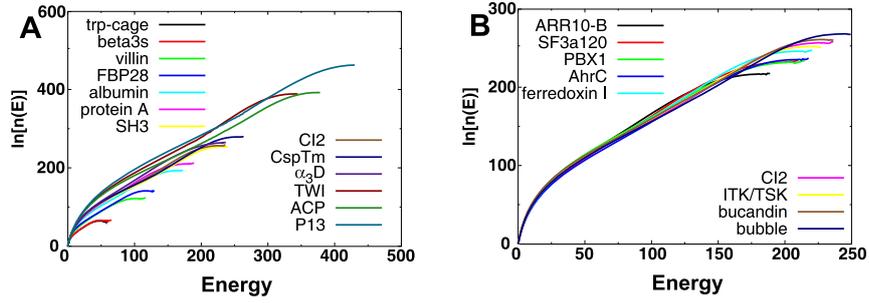


Figure S26: Folding energy landscape in one dimension. Logarithm of the density of states as a function of energy for (A) 13 proteins with different sizes and (B) 9 proteins with same size. The lowest (native) energy E_n set to 0 for visualization purposes.

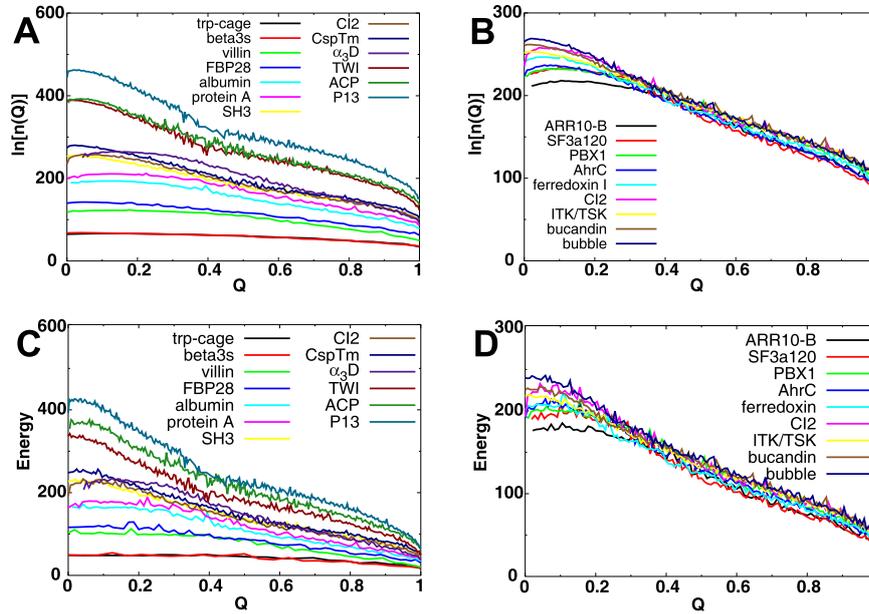


Figure S27: Folding energy landscape in one dimension. Logarithm of the density of states as a function of the fraction of native contacts Q for (A) 13 proteins of different sizes and (B) 9 proteins of same size. Average energy as a function of Q for (C) different-sized 13 proteins and (D) 9 proteins with same size.

2.1.4 Free energy profile

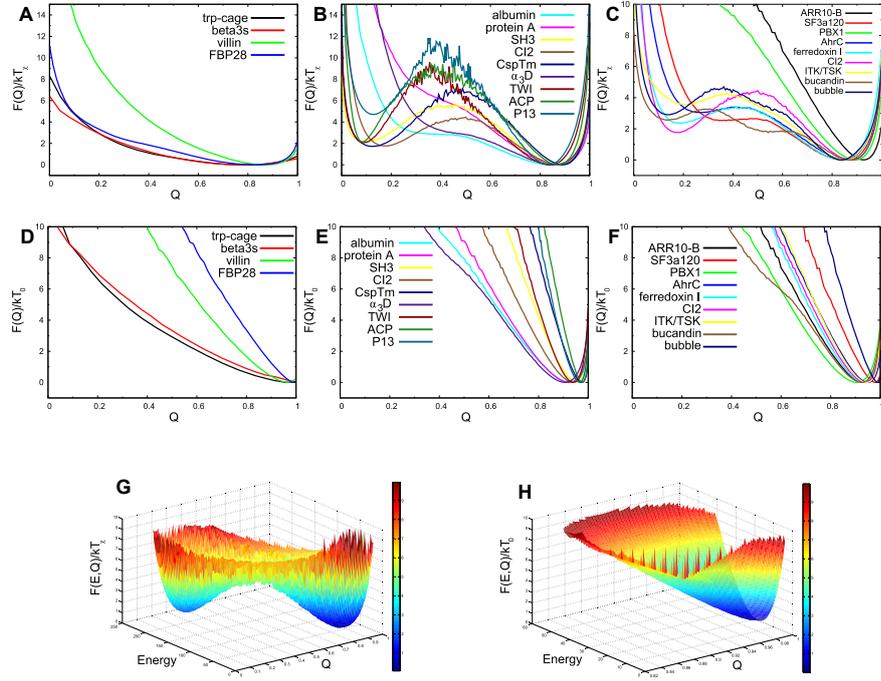


Figure S28: The free energy in one dimension as a function of fraction of native contacts Q and in two dimension as a function of Q and energy. Free energy profile at kinetic temperature T_x normalized by the corresponding kT_x as a function of Q for (A) one-state downhill folding proteins, (B) 2-state folding proteins with different sizes and (C) proteins with same size. Free energy profile at the optimum temperature T_0 when the folding kinetics is the fastest normalized by the corresponding kT_0 as a function of Q for (D) one-state downhill folding proteins, (E) 2-state folding proteins with different sizes and (F) proteins with same size. Two dimensional free energy profile for CI2 as a function of Q and energy at (G) T_x , (H) T_0 normalized by the corresponding kT . The white region is not probed by the protein.

2.2 Energetic roughness

The following figures and tables provide the details of all calculated quantities and detailed descriptions of each protein studied.

2.2.1 Protein: Villin headpiece subdomain (1YRF)

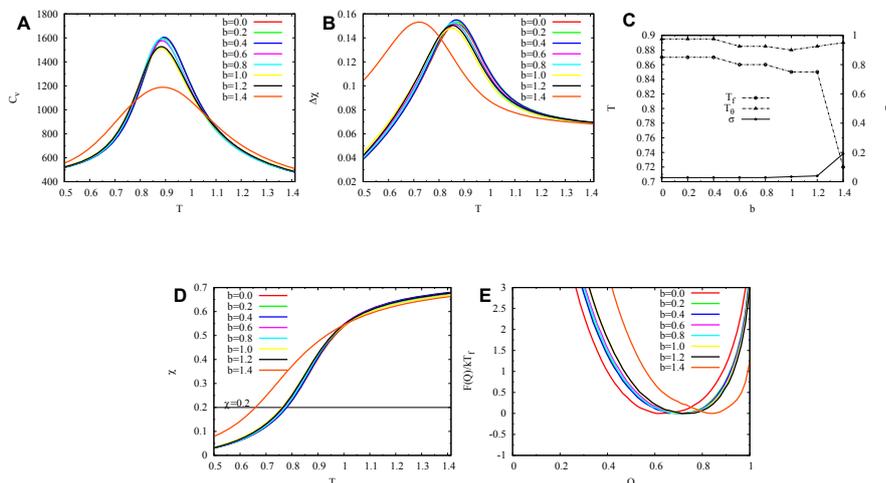


Figure S29: Thermodynamic results from REMD simulations. (A) Heat capacity as a function of temperature for different b . The peak of the curve corresponds to folding temperature T_θ . (B) The standard deviation of χ as a function of temperature for different b . The peak of the curve corresponds to folding temperature T_f . (C) Collapsed temperature T_θ , folding temperature T_f and thermodynamic parameter $\sigma = (T_\theta - T_f)/T_f$ as a function of b . (D) χ as a function of temperature for different b . The temperatures T_χ at which $\chi = 0.2$ are picked to perform the kinetic simulation. (E) Free energy profile along Q at the corresponding folding temperature for different b normalized by the corresponding kT_f .

In Figure S29, we can clearly see the effect of the energetic roughness in the thermodynamics of protein folding. Figure S29A shows that the peak of the heat capacity decrease with increasing b , suggesting that the folding cooperativity is reduced. In Figure S29B, S29C, the folding temperature decreases as b increases and σ increases slightly as b increases, demonstrating that the energetic roughness can decrease the foldability of the proteins. In Figure S29D, T_χ decreases as b increases.

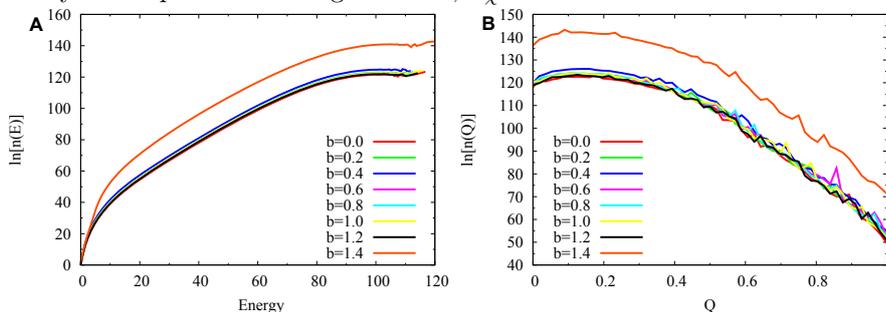


Figure S30: One dimensional density of states. (A) Log density of states as a function of energy for different b . (B) Log density of states as a function of fraction of native contacts Q for different b . The lowest energy E_n is relocated to 0 for a better visualization.

Table S3: Quantities extracted from the density of states for villin with different b .

b	0.2	0.4	0.6	0.8	1.0	1.2	1.4
E_N	-54.44	-54.64	-54.46	-54.49	-54.45	-54.41	-57.50
A_{max} ^(a)	37	37	38	33	35	38	39
Q_{NN} ^(b)	0.34	0.25	0.21	0.30	0.32	0.16	0.38
$\delta E(Q_{NN})$ ^(c)	101.97	100.34	97.56	97.66	110.45	100.75	96.20
$\Delta E(Q_{NN})_{Ene}$ ^(d)	1.22	2.43	3.70	4.60	5.92	7.40	8.74
$\Delta E(Q_{NN})$ ^(d)	3.08	3.73	4.66	5.40	6.56	7.92	9.19
$S(Q_{NN})$ ^(e)	119.76	124.30	122.66	120.67	122.03	122.97	135.55
$\frac{\delta E}{\sqrt{2S}}$	6.74	6.52	6.36	6.41	7.13	6.53	5.98
$\frac{\delta E}{\Delta E}$	31.58	25.70	20.02	17.13	16.43	12.17	10.47
T_f	0.87	0.87	0.86	0.86	0.85	0.85	0.72
T_g ^(f)	0.21	0.25	0.32	0.37	0.43	0.54	0.57
$\frac{T_f}{T_g}$	4.14	3.48	2.69	2.32	1.98	1.57	1.26
Λ	2.19	1.75	1.37	1.19	1.09	0.82	0.65
T_0	0.58	0.65	0.58	0.62	0.67	0.70	0.73
$\frac{T_0}{T_g}$	7.25	4.33	2.42	2.07	1.76	1.49	1.38
T_X	0.78	0.78	0.77	0.77	0.76	0.76	0.66
$\frac{T_X}{T_g}$	9.75	5.20	3.21	2.57	2.00	1.62	1.25
T_θ	0.89	0.89	0.88	0.88	0.88	0.88	0.89
σ	0.02	0.02	0.02	0.02	0.03	0.03	0.19

^(a) A_{max} is the maximum number of non-native contacts the protein can form in non-native states.

^(b) Q_{NN} is the fraction of native contacts in non-native states with the maximum number of non-native contacts.

^(c) $\delta E(Q_{NN})$ is the energy gap between native states and average of non-native states at Q_{NN} .

^(d) $\Delta E(Q_{NN})$ is the roughness of the energy landscape of non-native states. $\Delta E^2 = \Delta E_{Top}^2 + \Delta E_{Ene}^2$. ΔE_{Top} is topological roughness, obtained from the results of the plain structure based simulation. ΔE_{Ene} is energetic roughness, defined as $\sqrt{A_{max} b^2}$.

^(e) $S(Q_{NN})$ is the entropy of non-native states.

^(f) T_g is defined as $\sqrt{(\Delta E^2(Q_{NN})/2S(Q_{NN}))}$.

^(#) Temperature is in the unit of energy scale by multiplying Boltzmann constant k . Length and energy scale are in the reduced units of Gromacs [3, 4]. τ is in the unit of MD step.

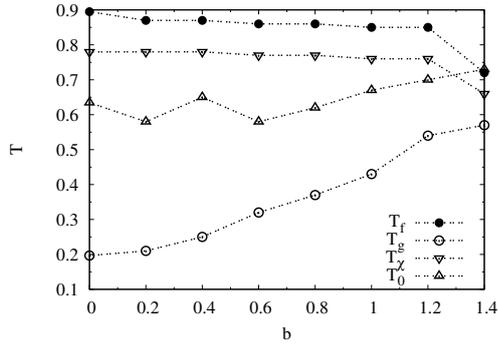


Figure S31: Folding temperature T_f , glass trapping transition temperature T_g , kinetic temperature T_χ and optimum temperature T_0 as a function of b .

We see in Figure S31 that T_f and T_χ slightly decrease as b increases, which is consistent with previous studies [31, 82]. In contrast, T_g and T_0 increase as b increases, since increasing energetic roughness increases the role of energetic traps.

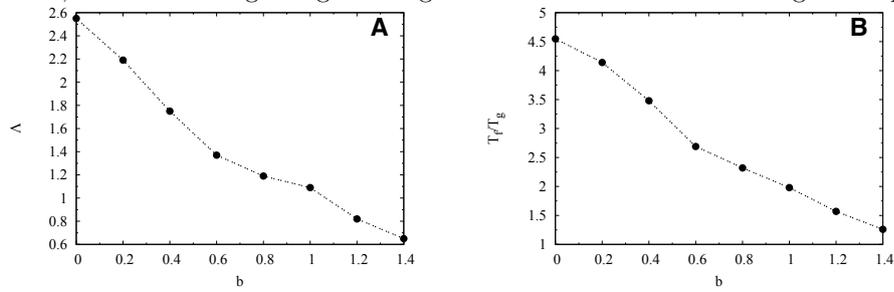


Figure S32: (A) Λ and (B) T_f/T_g as a function of b . Both Λ and T_f/T_g show a monotonically negative behavior with b .

In Figure S32, Λ and T_f/T_g decreases as the non-native interactions increases. This indicates that the thermodynamic stability drops as the non-native interactions increase, whereas, the landscape becomes more bumpy and less smooth as the strong of non-native interactions is increased.

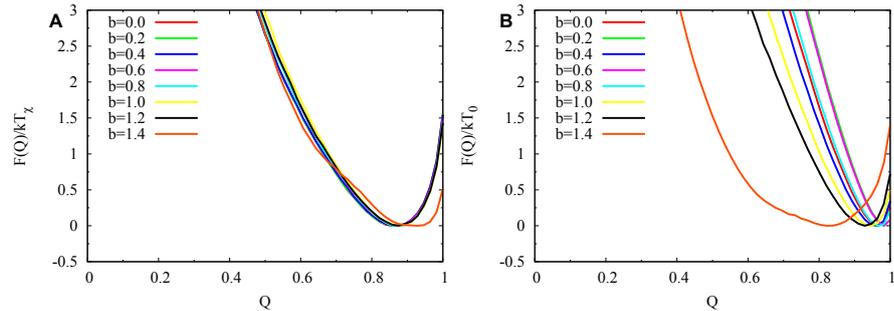


Figure S33: Free energy profile at the kinetic temperature T_χ and optimum temperature T_0 as a function of native contacts Q for different b . The free energy is normalized by the corresponding temperature kT .

We see in Figure S33 that either at T_χ or T_0 , the folding of villin is downhill in the same way as when no non-native interactions are considered.

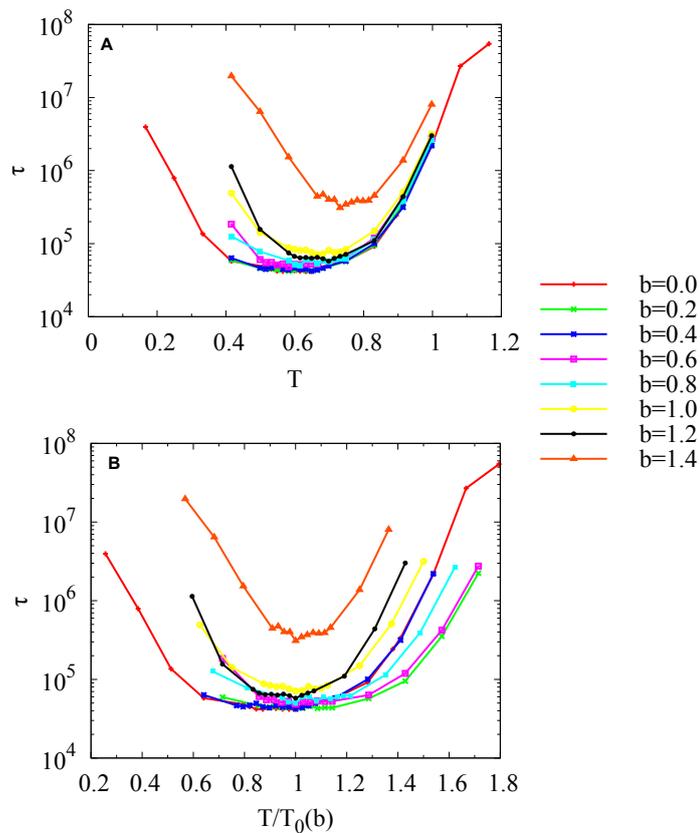


Figure S34: Folding time τ as a function of (A) absolute simulation temperature T and (B) normalized temperature T/T_0 for different b . Folding time is in the unit of MD step.

The folding time of different non-native interactions changes with temperature are shown in Figure S34. The behavior is similar as no non-native interactions are considered in Figure S14. The overall kinetic rate slows down as the non-native interactions increase.

2.2.2 Protein: Albumin binding domain (1PRB)

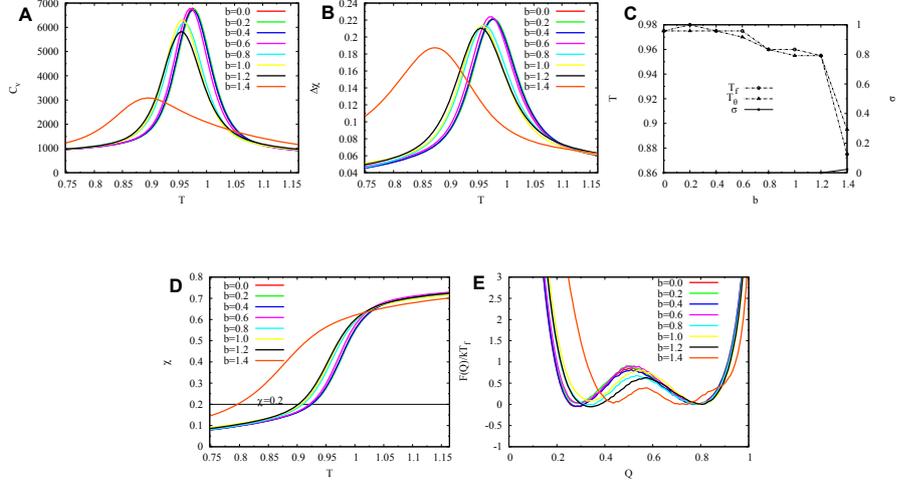


Figure S35: Thermodynamic results from REMD simulations. (A) Heat capacity as a function of temperature for different b . The peak of the curve corresponds to folding temperature T_θ . (B) The standard deviation of χ as a function of temperature for different b . The peak of the curve corresponds to folding temperature T_f . (C) Collapsed temperature T_θ , folding temperature T_f and thermodynamic parameter $\sigma = (T_\theta - T_f)/T_f$ as a function of b . (D) χ as a function of temperature for different b . The temperatures T_χ at which $\chi = 0.2$ are picked to do the kinetic simulation. (E) Free energy profile along Q at the corresponding folding temperature for different b normalized by the corresponding kT_f .

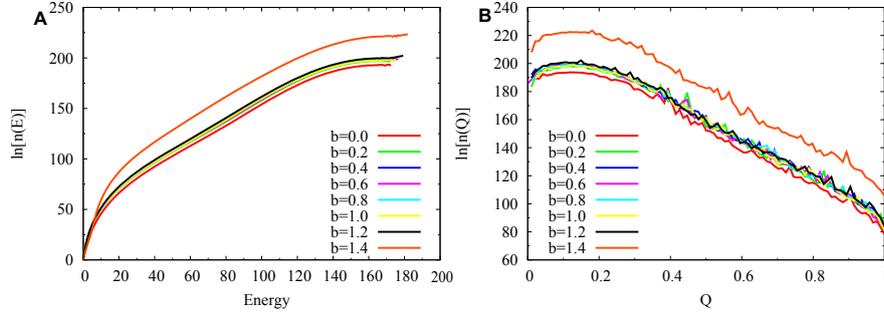


Figure S36: One dimensional density of states. (A) Log density of states as a function of energy for different b . (B) Log density of states as a function of fraction of native contacts Q for different b . The lowest energy E_n is relocated to 0 for a better visualization.

Table S4: Quantities extracted from the density of states for albumin with different b .

b	0.2	0.4	0.6	0.8	1.0	1.2	1.4
E_N	-90.85	-91.13	-91.00	-91.11	-91.01	-91.45	-96.17
A_{max} (a)	73	82	79	73	74	72	80
Q_{NN} (b)	0.19	0.18	0.16	0.27	0.16	0.18	0.26
$\delta E(Q_{NN})$ (c)	166.16	159.18	163.78	154.21	166.60	164.18	169.18
$\Delta E(Q_{NN})_{Ene}$ (d)	1.71	3.62	5.33	6.84	8.60	10.18	12.52
$\Delta E(Q_{NN})$ (d)	4.82	5.78	6.98	8.19	9.71	11.14	13.31
$S(Q_{NN})$ (e)	179.96	173.29	177.24	169.26	192.48	186.56	217.48
$\frac{\delta E}{\sqrt{2S}}$	8.57	8.26	8.47	8.14	8.37	8.36	8.11
$\frac{\delta E}{\Delta E}$	33.70	26.58	22.83	18.30	16.90	14.51	12.71
T_f	0.98	0.97	0.97	0.96	0.96	0.95	0.88
T_g (f)	0.25	0.31	0.37	0.45	0.49	0.58	0.64
$\frac{T_f}{T_g}$	3.92	3.13	2.62	2.13	1.96	1.64	1.38
Λ	1.78	1.43	1.21	0.99	0.86	0.75	0.61
T_0	0.80	0.88	0.80	0.86	0.88	0.90	0.91
$\frac{T_0}{T_g}$	8.89	4.89	2.96	2.46	2.05	1.76	1.52
T_χ	0.93	0.92	0.92	0.91	0.91	0.90	0.80
$\frac{T_\chi}{T_g}$	10.33	5.11	3.41	2.60	2.12	1.76	1.33
T_θ	0.97	0.97	0.97	0.96	0.95	0.95	0.89
σ	-0.01	0.00	0.00	0.00	-0.01	0.00	0.01

(a) A_{max} is the maximum number of non-native contacts the protein can form in non-native states.

(b) Q_{NN} is the fraction of native contacts in non-native states with the maximum number of non-native contacts.

(c) $\delta E(Q_{NN})$ is the energy gap between native states and average of non-native states at Q_{NN} .

(d) $\Delta E(Q_{NN})$ is the roughness of the energy landscape of non-native states. $\Delta E^2 = \Delta E_{Top}^2 + \Delta E_{Ene}^2$. ΔE_{Top} is topological roughness, obtained from the results of the plain structure based simulation. ΔE_{Ene} is energetic roughness, defined as $\sqrt{A_{max}b^2}$.

(e) $S(Q_{NN})$ is the entropy of non-native states.

(f) T_g is defined as $\sqrt{(\Delta E^2(Q_{NN})/2S(Q_{NN}))}$.

(#) Temperature is in the unit of energy scale by multiplying Boltzmann constant k . Length and energy scale are in the reduced units of Gromacs [3, 4]. τ is in the unit of MD step.

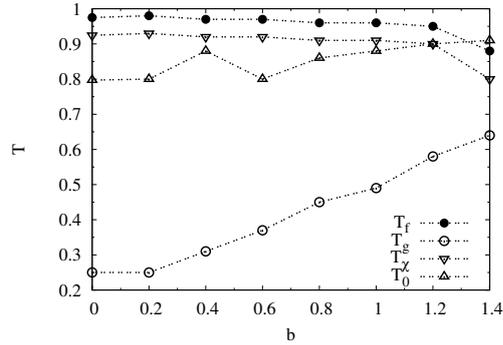


Figure S37: Folding temperature T_f , glass trapping transition temperature T_g , kinetic temperature T_χ and optimum temperature T_0 as a function of b .

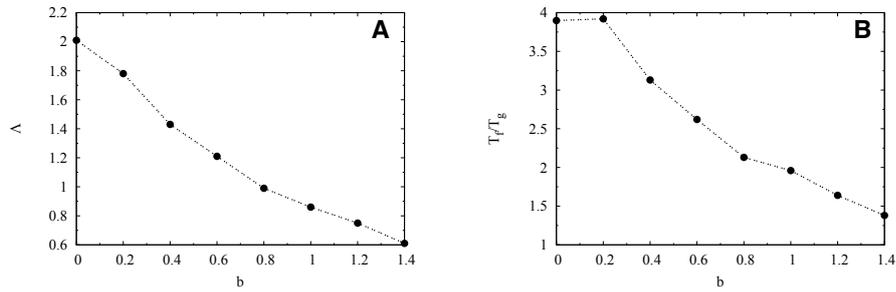


Figure S38: (A) Λ and (B) T_f/T_g as a function of b . Both Λ and T_f/T_g show a monotonically negative behavior with b .

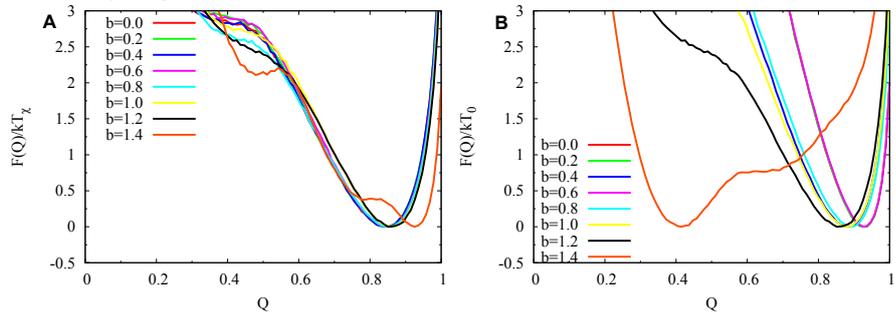


Figure S39: Free energy profile at the kinetic temperature T_χ and optimum temperature T_0 as a function of native contacts Q for different b . The free energy is normalized by the corresponding temperature kT .

We see in Figure S39 that except for $b=1.4$, either at T_χ or T_0 , the folding of albumin is a one-state downhill folder, which is similar to the results obtained without non-native interactions. When $b=1.4$, the free energy at T_0 is biased to non-native states, it is because at this moment the energetic roughness may be more significant as compared with the barrier height. The low temperature will make the protein trap easily and slow the folding rate.

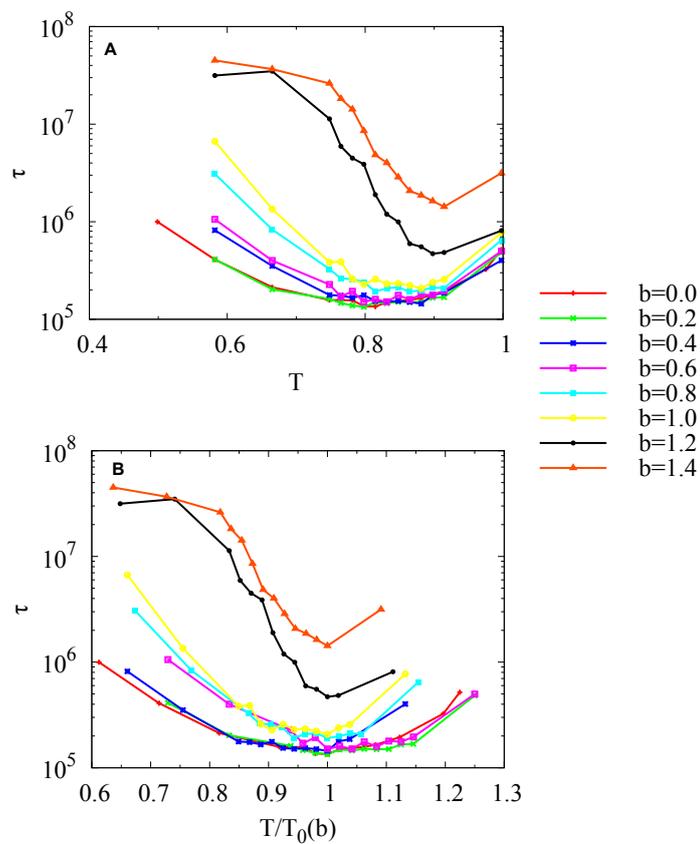


Figure S40: Folding time τ as a function of (A) absolute simulation temperature T and (B) normalized temperature T/T_0 for different b . Folding time is in the unit of MD step.

2.2.3 Protein: chymotrypsin inhibitor 2 CI2 (1YPA)

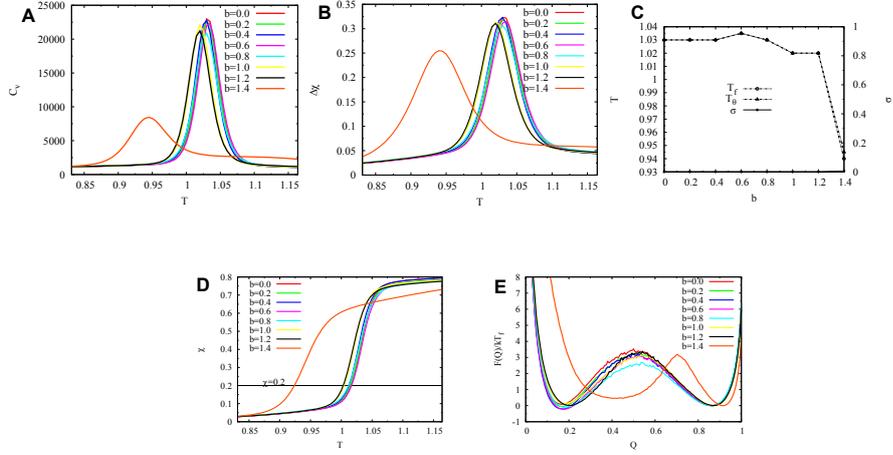


Figure S41: Thermodynamic results from REMD simulations. (A) Heat capacity as a function of temperature for different b . The peak of the curve corresponds to folding temperature T_θ . (B) The standard deviation of χ as a function of temperature for different b . The peak of the curve corresponds to folding temperature T_f . (C) Collapsed temperature T_θ , folding temperature T_f and thermodynamic parameter $\sigma = (T_\theta - T_f)/T_f$ as a function of b . The temperatures T_χ at which $\chi = 0.2$ are picked to do the kinetic simulation. (E) Free energy profile along Q at the corresponding folding temperature for different b normalized by the corresponding kT_f .

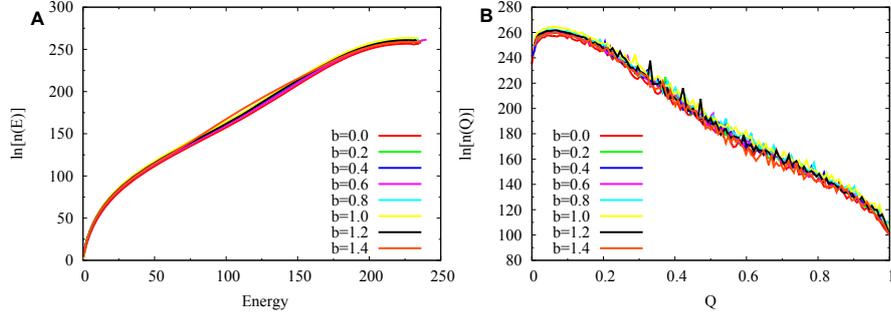


Figure S42: One dimensional density of states. (A) Log density of states as a function of energy for different b . (B) Log density of states as a function of fraction of native contacts Q for different b . The lowest energy E_n is relocated to 0 for a better visualization.

Table S5: Quantities extracted from the density of states for CI2 with different b .

b	0.2	0.4	0.6	0.8	1.0	1.2	1.4
E_N	-137.56	-137.50	-138.42	-138.37	-138.93	-138.11	-138.04
A_{max} (a)	72	72	74	69	74	71	70
Q_{NN} (b)	0.15	0.09	0.15	0.08	0.18	0.22	0.09
$\delta E(Q_{NN})$ (c)	217.67	223.31	219.89	223.80	208.08	201.01	220.49
$\Delta E(Q_{NN})_{Ene}$ (d)	1.70	3.39	5.16	6.65	8.60	10.11	11.71
$\Delta E(Q_{NN})$ (d)	10.28	10.69	11.38	12.12	13.30	14.32	15.49
$S(Q_{NN})$ (e)	179.93	188.93	177.28	186.30	175.16	170.62	183.16
$\frac{\delta E}{\sqrt{2S}}$	10.12	10.30	10.48	10.31	9.74	9.51	10.32
$\frac{\delta E}{\Delta E}$	18.67	18.73	17.34	16.41	13.71	12.26	12.75
T_f	1.03	1.03	1.03	1.03	1.02	1.02	0.94
T_g (f)	0.54	0.55	0.60	0.63	0.71	0.78	0.81
$\frac{T_f}{T_g}$	1.91	1.87	1.72	1.63	1.44	1.31	1.16
Λ	0.98	0.96	0.92	0.85	0.73	0.66	0.67
T_0	0.95	0.93	0.93	0.96	0.93	0.93	0.90
$\frac{T_0}{T_g}$	11.88	6.20	4.04	3.31	2.45	2.02	1.73
T_χ	1.01	1.01	1.02	1.01	1.00	1.00	0.92
$\frac{T_\chi}{T_g}$	12.62	6.73	4.43	3.48	2.63	2.17	1.77
T_θ	1.03	1.03	1.03	1.03	1.02	1.02	0.94
σ	0.00	0.00	0.00	0.00	0.00	0.00	0.00

- (a) A_{max} is the maximum number of non-native contacts the protein can form in non-native states.
(b) Q_{NN} is the fraction of native contacts in non-native states with the maximum number of non-native contacts.
(c) $\delta E(Q_{NN})$ is the energy gap between native states and average of non-native states at Q_{NN} .
(d) $\Delta E(Q_{NN})$ is the roughness of the energy landscape of non-native states. $\Delta E^2 = \Delta E_{Top}^2 + \Delta E_{Ene}^2$.
 ΔE_{Top} is topological roughness, obtained from the results of the plain structure based simulation.
 ΔE_{Ene} is energetic roughness, defined as $\sqrt{A_{max}b^2}$.
(e) $S(Q_{NN})$ is the entropy of non-native states.
(f) T_g is defined as $\sqrt{(\Delta E^2(Q_{NN})/2S(Q_{NN}))}$.
(#) Temperature is in the unit of energy scale by multiplying Boltzmann constant k . Length and energy scale are in the reduced units of Gromacs [3, 4]. τ is in the unit of MD step.

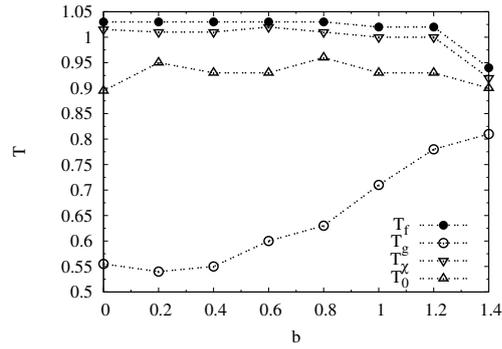


Figure S43: Folding temperature T_f , glass trapping transition temperature T_g , kinetic temperature T_χ and optimum temperature T_0 as a function of b .

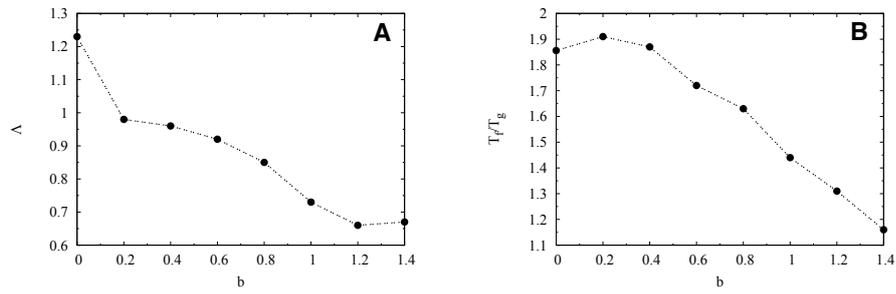


Figure S44: (A) Λ and (B) T_f/T_g as a function of b . Both Λ and T_f/T_g show a monotonically negative behavior with b .

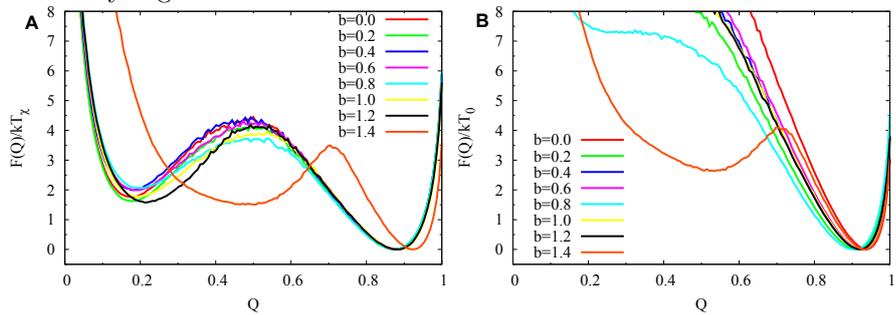


Figure S45: Free energy profile at the kinetic temperature T_χ and optimum temperature T_0 as a function of native contacts Q for different b . The free energy is normalized by the corresponding temperature kT .

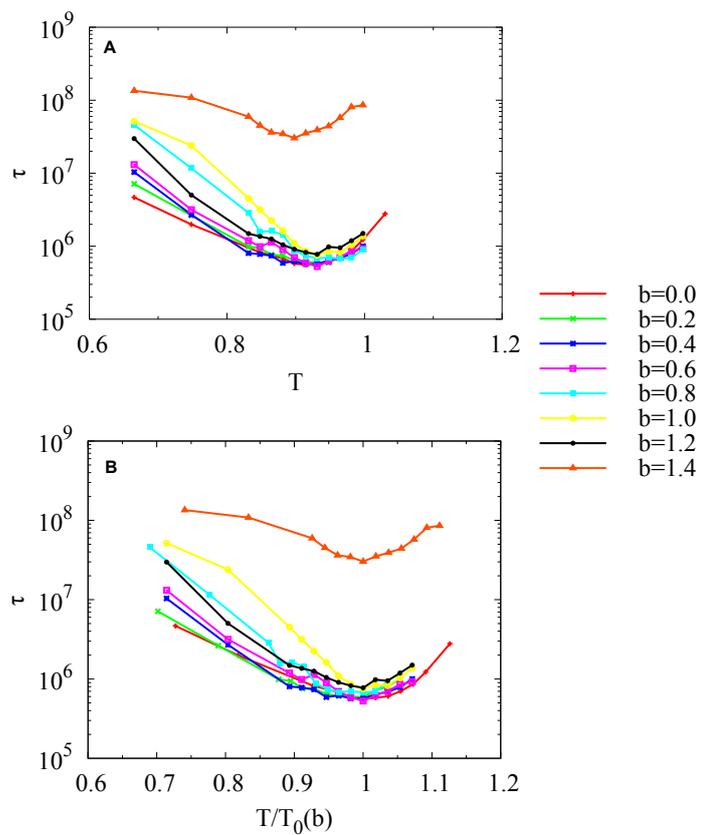


Figure S46: Folding time τ as a function of (A) absolute simulation temperature T and (B) normalized temperature T/T_0 for different b . Folding time is in the unit of MD step.

2.2.4 Protein: ARR10-B (1IRZ)

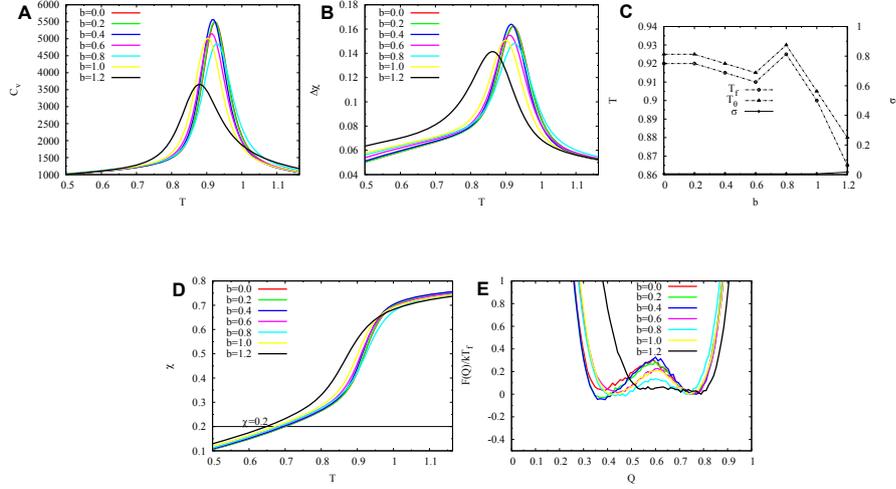


Figure S47: Thermodynamic results from REMD simulations. (A) Heat capacity as a function of temperature for different b . The peak of the curve corresponds to folding temperature T_θ . (B) The standard deviation of χ as a function of temperature for different b . The peak of the curve corresponds to folding temperature T_f . (C) Collapsed temperature T_θ , folding temperature T_f and thermodynamic parameter $\sigma = (T_\theta - T_f)/T_f$ as a function of b . (D) χ as a function of temperature for different b . The temperatures T_χ at which $\chi = 0.2$ are picked to do the kinetic simulation. (E) Free energy profile along Q at the corresponding folding temperature for different b normalized by the corresponding kT_f .

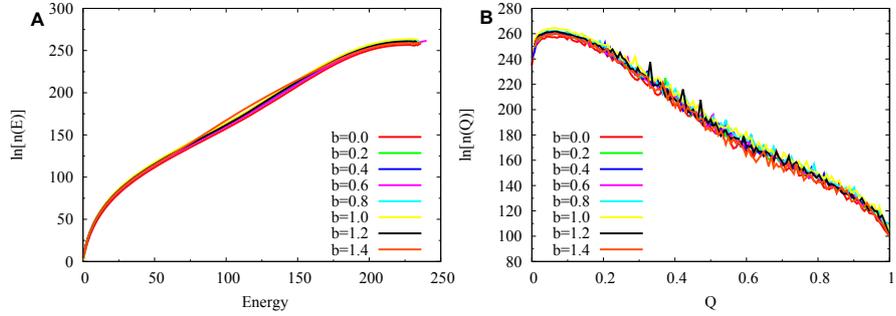


Figure S48: One dimensional density of states. (A) Log density of states as a function of energy for different b . (B) Log density of states as a function of fraction of native contacts Q for different b . The lowest energy E_n is relocated to 0 for a better visualization.

It seems that as input of the energetic roughness increases, the density of states become rougher and more bumpy as expected.

Table S6: Quantities extracted from the density of states for ARR10-B with different b .

b	0.2	0.4	0.6	0.8	1.0	1.2
E_N	-90.91	-90.79	-91.62	-94.56	-91.95	-92.05
A_{max} (a)	39	37	47	44	37	38
Q_{NN} (b)	0.31	0.31	0.19	0.27	0.27	0.22
$\delta E(Q_{NN})$ (c)	163.87	166.88	183.03	171.97	169.79	192.94
$\Delta E(Q_{NN})_{Ene}$ (d)	1.25	2.43	4.11	5.31	6.08	7.40
$\Delta E(Q_{NN})$ (d)	5.67	5.97	6.63	7.15	7.98	8.92
$S(Q_{NN})$ (e)	182.29	182.38	210.84	226.60	226.06	224.55
$\frac{\delta E}{\sqrt{2S}}$	8.81	8.86	8.36	8.51	8.14	8.47
$\frac{\Delta E}{\Delta E}$	29.66	28.34	25.88	25.31	21.68	20.11
T_f	0.92	0.92	0.91	0.93	0.90	0.86
T_g (f)	0.30	0.31	0.32	0.34	0.38	0.42
$\frac{T_f}{T_g}$	3.07	2.97	2.84	2.74	2.37	2.05
Λ	1.55	1.48	1.26	1.19	1.02	0.95
T_0	0.67	0.72	0.73	0.76	0.70	0.72
$\frac{T_0}{T_g}$	11.17	6.55	2.94	3.62	2.59	2.18
T_χ	0.70	0.70	0.68	0.68	0.67	0.65
$\frac{T_\chi}{T_g}$	11.67	6.36	4.00	3.24	2.48	1.97
T_θ	0.92	0.92	0.91	0.93	0.90	0.88
σ	0.00	0.00	0.00	0.00	0.00	0.02

- (a) A_{max} is the maximum number of non-native contacts the protein can form in non-native states.
(b) Q_{NN} is the fraction of native contacts in non-native states with the maximum number of non-native contacts.
(c) $\delta E(Q_{NN})$ is the energy gap between native states and average of non-native states at Q_{NN} .
(d) $\Delta E(Q_{NN})$ is the roughness of the energy landscape of non-native states. $\Delta E^2 = \Delta E_{Top}^2 + \Delta E_{Ene}^2$.
 ΔE_{Top} is topological roughness, obtained from the results of the plain structure based simulation.
 ΔE_{Ene} is energetic roughness, defined as $\sqrt{A_{max} b^2}$.
(e) $S(Q_{NN})$ is the entropy of non-native states.
(f) T_g is defined as $\sqrt{(\Delta E^2(Q_{NN})/2S(Q_{NN}))}$.
(#) Temperature is in the unit of energy scale by multiplying Boltzmann constant k . Length and energy scale are in the reduced units of Gromacs [3, 4]. τ is in the unit of MD step.

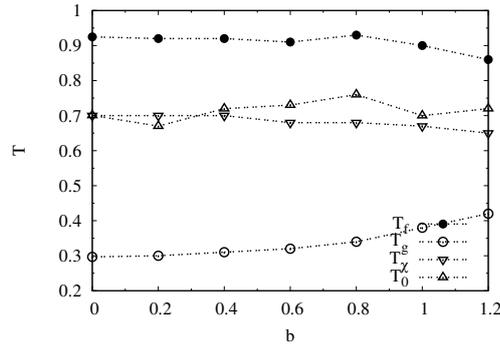


Figure S49: Folding temperature T_f , glass trapping transition temperature T_g , kinetic temperature T_χ and optimum temperature T_0 as a function of b .

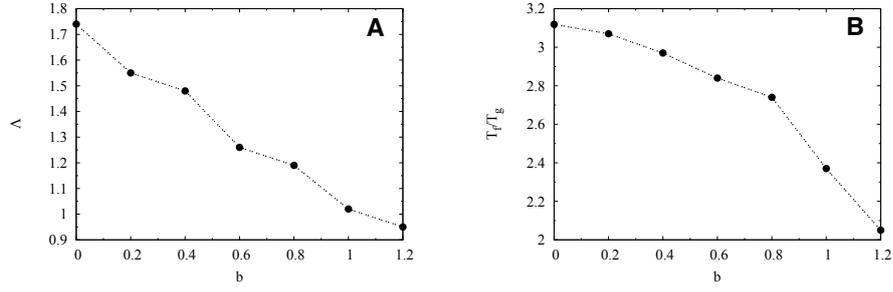


Figure S50: (A) Λ and (B) T_f/T_g as a function of b . Both Λ and T_f/T_g show a monotonically negative behavior with b .

In Figure S50, Λ and T_f/T_g decreases as the non-native interactions increases. This means the thermodynamic stability drops with the non-native interactions increase while landscape becomes more bumpy and less smooth as non-native interactions increase.

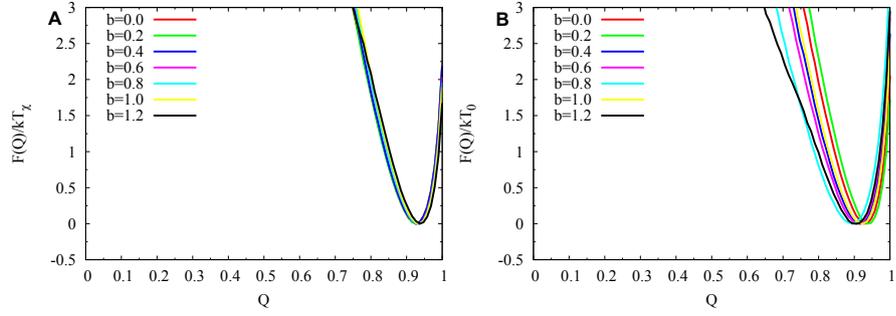


Figure S51: Free energy profile at the kinetic temperature T_χ and optimum temperature T_0 as a function of native contacts Q for different b . The free energy is normalized by the corresponding temperature kT .

We see in Figure S51 that either at T_χ or T_0 , the folding of ARR10-B is one-state downhill, similar to the situation when no non-native interactions.

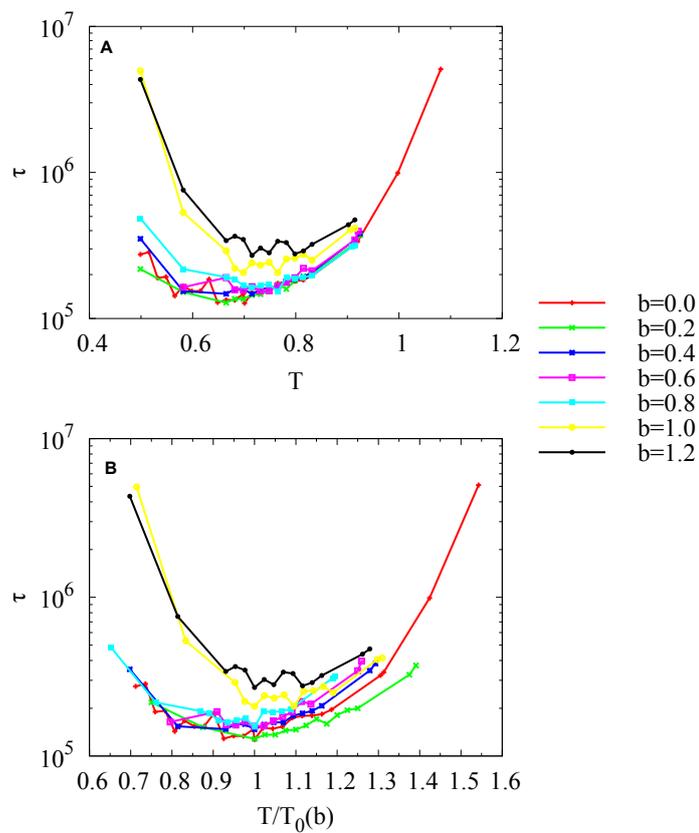


Figure S52: Folding time τ as a function of (A) absolute simulation temperature T and (B) normalized temperature T/T_0 for different b . Folding time is in the unit of MD step.

2.2.5 Protein: Bubble protein (1UOY)

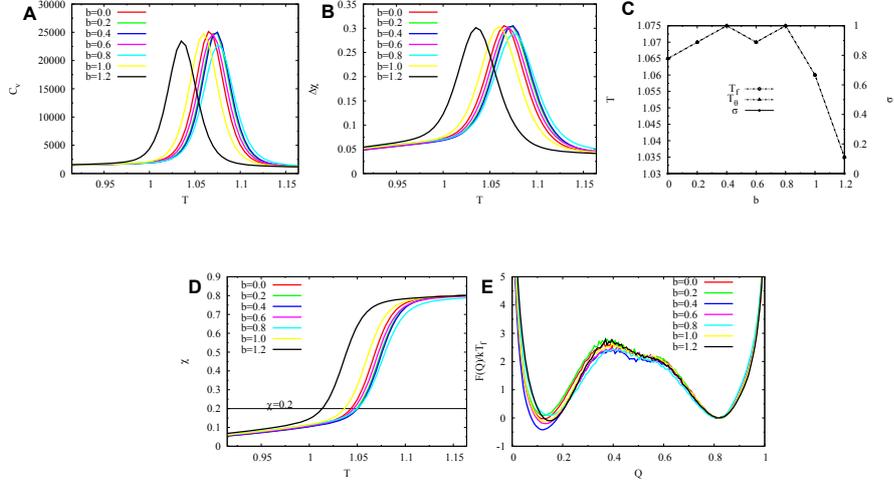


Figure S53: Thermodynamic results from REMD simulations. (A) Heat capacity as a function of temperature for different b . The peak of the curve corresponds to folding temperature T_θ . (B) The standard deviation of χ as a function of temperature for different b . The peak of the curve corresponds to folding temperature T_f . (C) Collapsed temperature T_θ , folding temperature T_f and thermodynamic parameter $\sigma = (T_\theta - T_f)/T_f$ as a function of b . (D) χ as a function of temperature for different b . The temperatures T_χ at which $\chi = 0.2$ are picked to do the kinetic simulation. (E) Free energy profile along Q at the corresponding folding temperature for different b normalized by the corresponding kT_f .

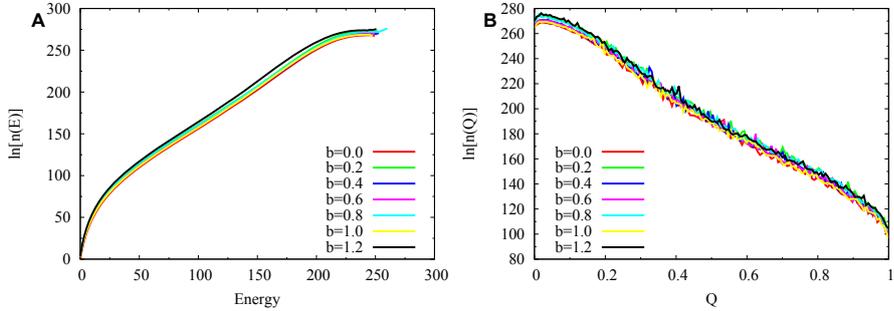


Figure S54: One dimensional density of states. (A) Log density of states as a function of energy for different b . (B) Log density of states as a function of fraction of native contacts Q for different b . The lowest energy E_n is relocated to 0 for a better visualization.

Table S7: Quantities extracted from the density of states for bubble with different b .

b	0.2	0.4	0.6	0.8	1.0	1.2
E_N	-152.88	-152.46	-152.21	-154.38	-150.53	-148.51
A_{max} (a)	37	38	38	46	42	37
Q_{NN} (b)	0.13	0.14	0.02	0.06	0.08	0.04
$\delta E(Q_{NN})$ (c)	230.28	224.31	241.32	237.05	234.40	241.60
$\Delta E(Q_{NN})_{Ene}$ (d)	1.22	2.47	3.70	5.43	6.48	7.30
$\Delta E(Q_{NN})$ (d)	10.91	11.10	11.45	11.86	12.43	13.07
$S(Q_{NN})$ (e)	177.68	177.95	186.41	192.46	183.73	187.62
$\frac{\delta E}{\sqrt{2S}}$	10.42	10.56	10.68	10.89	10.30	10.75
$\frac{\delta E}{\Delta E}$	18.01	17.95	18.00	18.02	15.88	15.94
T_f	1.07	1.07	1.07	1.07	1.06	1.03
T_g (f)	0.58	0.59	0.59	0.60	0.65	0.67
$\frac{T_f}{T_g}$	1.84	1.81	1.81	1.78	1.63	1.54
Λ	0.96	0.95	0.93	0.92	0.83	0.82
T_0	0.80	0.68	0.83	0.85	0.76	0.93
$\frac{T_0}{T_g}$	16.00	6.80	5.19	4.05	2.81	3.00
T_χ	1.05	1.05	1.05	1.05	1.04	1.01
$\frac{T_\chi}{T_g}$	21.00	10.50	6.56	5.00	3.85	3.26
T_θ	1.07	1.07	1.07	1.07	1.06	1.03
σ	0.00	0.00	0.00	0.00	0.00	0.00

(a) A_{max} is the maximum number of non-native contacts the protein can form in non-native states.

(b) Q_{NN} is the fraction of native contacts in non-native states with the maximum number of non-native contacts.

(c) $\delta E(Q_{NN})$ is the energy gap between native states and average of non-native states at Q_{NN} .

(d) $\Delta E(Q_{NN})$ is the roughness of the energy landscape of non-native states. $\Delta E^2 = \Delta E_{Top}^2 + \Delta E_{Ene}^2$. ΔE_{Top} is topological roughness, obtained from the results of the plain structure based simulation.

ΔE_{Ene} is energetic roughness, defined as $\sqrt{A_{max}b^2}$.

(e) $S(Q_{NN})$ is the entropy of non-native states.

(f) T_g is defined as $\sqrt{(\Delta E^2(Q_{NN})/2S(Q_{NN}))}$.

(#) Temperature is in the unit of energy scale by multiplying Boltzmann constant k . Length and energy scale are in the reduced units of Gromacs [3, 4]. τ is in the unit of MD step.

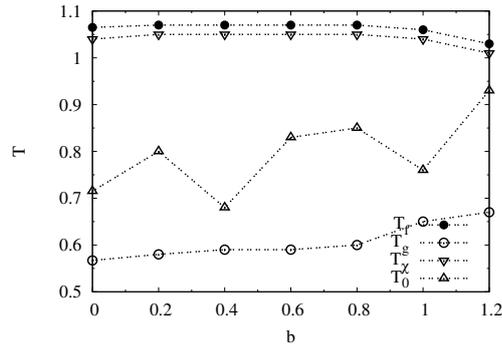


Figure S55: Folding temperature T_f , glass trapping transition temperature T_g , kinetic temperature T_χ and optimum temperature T_0 as a function of b .

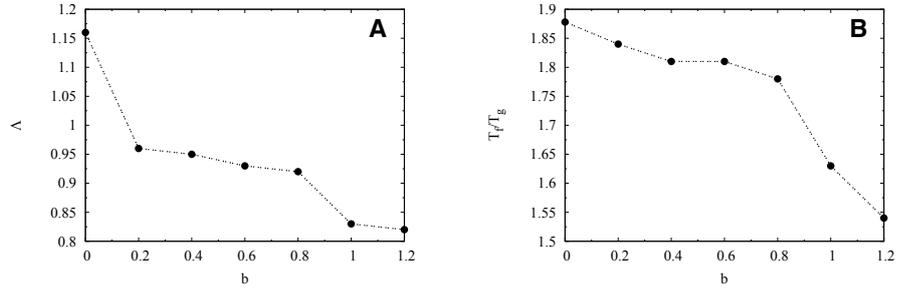


Figure S56: (A) Λ and (B) T_f/T_g as a function of b . Both Λ and T_f/T_g show a monotonically negative behavior with b .

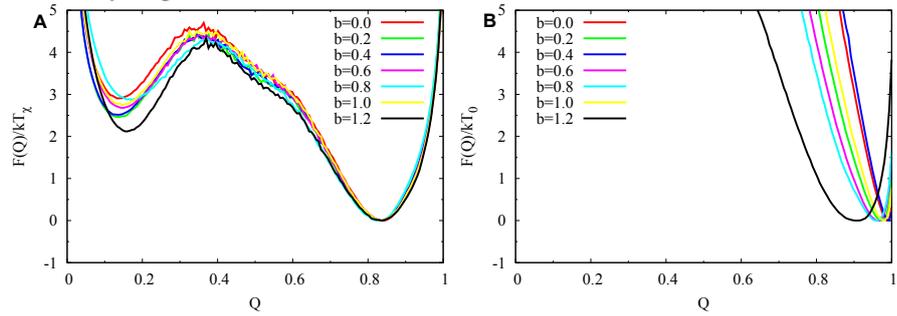


Figure S57: Free energy profile at the kinetic temperature T_χ and optimum temperature T_0 as a function of native contacts Q for different b . The free energy is normalized by the corresponding temperature kT .

In Figure S57, we see that at temperature T_χ , the free energy is biased to the native states with 2-state folding and at temperature T_0 , the folding behavior becomes downhill.

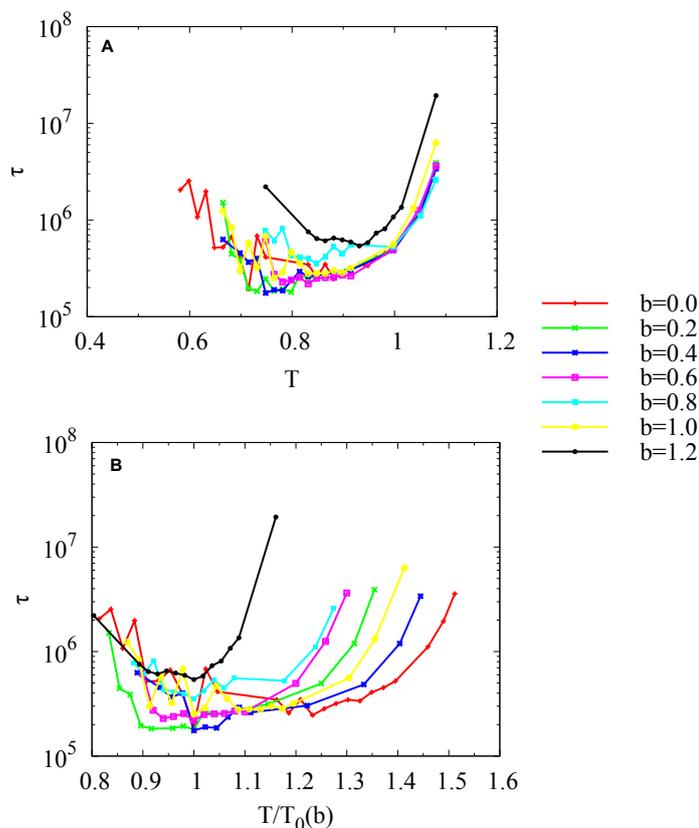


Figure S58: Folding time τ as a function of (A) absolute simulation temperature T and (B) normalized temperature T/T_0 for different b . Folding time is in the unit of MD step.

The folding time of different non-native interactions changes with temperature are shown in Figure S58. The behavior is similar to the situation with no non-native interactions as in Figure S20. The overall folding times are longer with increasing non-native interactions.

References

- [1] C. Clementi, H. Nymeyer, and J. N. Onuchic. 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*, 298(5):937–953, 2000.
- [2] V. Sobolev, A. Sorokine, J. Prilusky, E. E. Abola, and M. Edelman. Automated analysis of interatomic contacts in proteins. *Bioinformatics*, 15(4):327–332, 1999.
- [3] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*, 4(3):435–447, 2008.
- [4] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic. Smog@ctbp:

- simplified deployment of structure-based models in gromacs. *Nucleic Acids Res*, 38:W657–W661, 2010.
- [5] Y. Okamoto and Y. Sugita. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett*, 314(1-2):141–151, 1999.
 - [6] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules .1. the method. *J Comput Chem*, 13(8):1011–1021, 1992.
 - [7] M. Andrec, A. K. Felts, E. Gallicchio, and R. M. Levy. Protein folding pathways from replica exchange simulations and a kinetic network model. *Proc Natl Acad Sci USA*, 102(19):6801–6806, 2005.
 - [8] D. van der Spoel and M. M. Seibert. Protein folding kinetics and thermodynamics from atomistic simulations. *Phys Rev Lett*, 96(23), 2006.
 - [9] S. C. Yang, J. N. Onuchic, A. E. Garcia, and H. Levine. Folding time predictions from all-atom replica exchange simulations. *J Mol Biol*, 372(3):756–763, 2007.
 - [10] N. V. Buchete and G. Hummer. Peptide folding kinetics from replica exchange molecular dynamics. *Phys Rev E*, 77(3), 2008.
 - [11] S. Muff and A. Caffisch. Etna: Equilibrium transitions network and arrhenius equation for extracting folding kinetics from remd simulations. *J Phys Chem B*, 113(10):3218–3226, 2009.
 - [12] B. Derrida. Random-energy model - an exactly solvable model of disordered-systems. *Phys Rev B*, 24(5):2613–2626, 1981.
 - [13] E. I. Shakhnovich and A. M. Gutin. Engineering of stable and fast-folding sequences of model proteins. *Proc Natl Acad Sci USA*, 90(15):7195–7199, 1993.
 - [14] S. S. Plotkin and J. N. Onuchic. Understanding protein folding with energy landscape theory - part i: Basic concepts. *Q Rev Biophys*, 35(2):111–167, 2002.
 - [15] S. S. Plotkin and J. N. Onuchic. Understanding protein folding with energy landscape theory - part ii: Quantitative aspects. *Q Rev Biophys*, 35(3):205–286, 2002.
 - [16] A. Sali, E. Shakhnovich, and M. Karplus. How does a protein fold. *Nature*, 369(6477):248–251, 1994.
 - [17] A. Sali, E. Shakhnovich, and M. Karplus. Kinetics of protein-folding - a lattice model study of the requirements for folding to the native-state. *J Mol Biol*, 235(5):1614–1636, 1994.
 - [18] E. I. Shakhnovich and A. M. Gutin. Implications of thermodynamics of protein folding for evolution of primary sequences. *Nature*, 346(6286):773–775, 1990.
 - [19] E. I. Shakhnovich. Proteins with selected sequences fold into unique native conformation. *Phys Rev Lett*, 72(24):3907–3910, 1994.
 - [20] D. K. Klimov and D. Thirumalai. Factors governing the foldability of proteins. *Proteins*, 26(4):411–441, 1996.

- [21] D. K. Klimov and D. Thirumalai. Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J Chem Phys*, 109(10):4119–4125, 1998.
- [22] D. K. Klimov and D. Thirumalai. Criterion that determines the foldability of proteins. *Phys Rev Lett*, 76(21):4070–4073, 1996.
- [23] A. R. Dinner, V. Abkevich, E. Shakhnovich, and M. Karplus. Factors that affect the folding ability of proteins. *Proteins*, 35(1):34–40, 1999.
- [24] M. Karplus, A. R. Dinner, and E. Verosub. Use of a quantitative structure-property relationship to design larger model proteins that fold rapidly. *Protein Eng*, 12(11):909–917, 1999.
- [25] R. Melin, H. Li, N. S. Wingreen, and C. Tang. Designability, thermodynamic stability, and dynamics in protein folding: A lattice model study. *J Chem Phys*, 110(2):1252–1262, 1999.
- [26] J. D. Bryngelson and P. G. Wolynes. Spin-glasses and the statistical-mechanics of protein folding. *Proc Natl Acad Sci USA*, 84(21):7524–7528, 1987.
- [27] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich. Free-energy landscape for protein-folding kinetics - intermediates, traps, and multiple pathways in theory and lattice model simulations. *J Chem Phys*, 101(7):6052–6062, 1994.
- [28] J. Wang and G. M. Verkhivker. Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Phys Rev Lett*, 90(18), 2003.
- [29] J. N. Onuchic, Z. LutheySchulten, and P. G. Wolynes. Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem*, 48:545–600, 1997.
- [30] J. Wang, C. Lee, and G. Stell. The cooperative nature of hydrophobic forces and protein folding kinetics. *Chem Phys*, 316(1-3):53–60, 2005.
- [31] C. Clementi and S. S. Plotkin. The effects of nonnative interactions on protein folding rates: Theory and simulation. *Protein Sci*, 13(7):1750–1766, 2004.
- [32] C. Clementi, P. Das, and S. Matysiak. Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes. *Proc Natl Acad Sci USA*, 102(29):10141–10146, 2005.
- [33] D. Thirumalai and D. K. Klimov. Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr Opin Struct Biol*, 9(2):197–207, 1999.
- [34] D. Thirumalai. From minimal models to real proteins - time scales for protein-folding kinetics. *J Phys I*, 5(11):1457–1467, 1995.
- [35] C. J. Camacho and D. Thirumalai. Kinetics and thermodynamics of folding in model proteins. *Proc Natl Acad Sci USA*, 90(13):6369–6372, 1993.
- [36] J. D. Honeycutt and D. Thirumalai. The nature of folded states of globular-proteins. *Biopolymers*, 32(6):695–709, 1992.
- [37] L. Angelani and G. Ruocco. Saddles of the energy landscape and folding of model proteins. *Europhys Lett*, 87(1), 2009.

- [38] W. Humphrey, A. Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *J Mol Graph*, 14(1):33–38, 1996.
- [39] P. G. Wolynes. Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proc Natl Acad Sci USA*, 94(12):6170–6175, 1997.
- [40] S. S. Plotkin, J. Wang, and P. G. Wolynes. Statistical mechanics of correlated energy landscape models for random heteropolymers and proteins. *Physica D*, 107(2-4):322–325, 1997.
- [41] G. Stell, C. L. Lee, and J. Wang. First-passage time distribution and non-markovian diffusion dynamics of protein folding. *J Chem Phys*, 118(2):959–968, 2003.
- [42] Y. Q. Zhou, C. Zhang, G. Stell, and J. Wang. Temperature dependence of the distribution of the first passage time: Results from discontinuous molecular dynamics simulations of an all-atom model of the second beta-hairpin fragment of protein g. *J Am Chem Soc*, 125(20):6300–6305, 2003.
- [43] C. Clementi, L. L. Chavez, and J. N. Onuchic. Quantifying the roughness on the free energy landscape: Entropic bottlenecks and protein folding rates. *J Am Chem Soc*, 126(27):8426–8432, 2004.
- [44] D. P. Raleigh, B. Kuhlman, D. L. Luisi, and P. A. Evans. Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the n-terminal domain of the protein 19. *J Mol Biol*, 284(5):1661–1670, 1998.
- [45] H. Kaya and H. S. Chan. Energetic components of cooperative protein folding. *Phys Rev Lett*, 85(22):4823–4826, 2000.
- [46] V. B. P. Leite, J. N. Onuchic, G. Stell, and J. Wang. Probing the kinetics of single molecule protein folding. *Biophys J*, 87(6):3633–3641, 2004.
- [47] J. Wang. The complex kinetics of protein folding in wide temperature ranges. *Biophys J*, 87(4):2164–2171, 2004.
- [48] J. Wang. Diffusion and single molecule dynamics on biomolecular interface binding energy landscape. *Chem Phys Lett*, 418(4-6):544–548, 2006.
- [49] G. Stell, C. L. Lee, C. T. Lin, and J. Wang. Diffusion dynamics, moments, and distribution of first-passage time on the protein-folding energy landscape, with applications to single molecules. *Phys Rev E*, 67(4), 2003.
- [50] A. V. Finkelstein and A. Y. Badretdinov. Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. *Fold Des*, 2(2):115–121, 1997.
- [51] M. Cieplak, T. X. Hoang, and M. S. Li. Scaling of folding properties in simple models of proteins. *Phys Rev Lett*, 83(8):1684–1687, 1999.
- [52] F. Seno, C. Micheletti, A. Maritan A, J.R. Banavar (1998) Variational approach to protein design and extraction of interaction potentials. *Phys Rev Lett* 81:2172–2175.

- [53] Chan HS, Kaya H (2002) Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: How applicable is the transition state picture to folding and unfolding? *J Mol Biol* 315:899–909.
- [54] Eaton WA, Lipman EA, Schuler B, Bakajin O (2003) Single-molecule measurement of protein folding kinetics. *Science* 301:1233–1235.
- [55] Gruebele M, Nguyen H, Jager M, Moretto A, Kelly JW (2003) Tuning the free-energy landscape of a ww domain by temperature, mutation, and truncation. *Proc Natl Acad Sci USA* 100:3948–3953.
- [56] Frauenfelder H, Parak F, Young RD (1988) Conformational substates in proteins. *Annu Rev Biophys Biophys Chem* 17:451–479.
- [57] Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* 254:1598–1603.
- [58] Xie XS, Yang H (2002) Probing single-molecule dynamics photon by photon. *J Chem Phys* 117:10965–10979.
- [59] Socci ND, Onuchic JN (1995) Kinetic and thermodynamic analysis of proteinlike heteropolymers - monte-carlo histogram technique. *J Chem Phys* 103:4732–4744.
- [60] Hardin C, Eastwood MP, Prentiss MC, Luthey-Schulten Z, Wolynes PG (2003) Associative memory hamiltonians for structure prediction without homology: alpha/beta proteins. *Proc Natl Acad Sci USA* 100:1679–1684.
- [61] Bryngelson JD, Wolynes PG (1989) Intermediates and barrier crossing in a random energy-model (with applications to protein folding). *J Phys Chem* 93:6902–6915.
- [62] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein-folding - a synthesis. *Proteins* 21:167–195.
- [63] Wang J, Onuchic J, Wolynes P (1996) Statistics of kinetic pathways on biased rough energy landscapes with applications to protein folding. *Phys Rev Lett* 76:4861–4864.
- [64] Itzhaki LS, Otzen DE, Fersht AR (1995) The structure of the transition-state for folding of chymotrypsin inhibitor-2 analyzed by protein engineering methods - evidence for a nucleation-condensation mechanism for protein-folding. *J Mol Biol* 254:260–288.
- [65] Gruebele M, Sabelko J, Ervin J (1999) Observation of strange kinetics in protein folding. *Proc Natl Acad Sci USA* 96:6031–6036.
- [66] Eaton WA, Schuler B, Lipman EA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747.
- [67] Chan HS, Dill KA (1994) Solvation - effects of molecular-size and shape. *J Chem Phys* 101:7007–7026.
- [68] Gutin AM, Abkevich VI, Shakhnovich EI (1996) Chain length scaling of protein folding time. *Phys Rev Lett* 77:5433–5436.

- [69] M. Cieplak and T. X. Hoang. Scaling of folding properties in go models of proteins. *J Biol Phys*, 26(4):273–294, 2000.
- [70] S. Takada and N. Koga. Roles of native topology and chain-length scaling in protein folding: A simulation study with a go-like model. *J Mol Biol*, 313(1):171–180, 2001.
- [71] M. S. Li, D. K. Klimov, and D. Thirumalai. Dependence of folding rates on protein length. *J Phys Chem B*, 106(33):8302–8305, 2002.
- [72] A. V. Finkelstein, D. N. Ivankov, S. O. Garbuzynskiy, E. Alm, K. W. Plaxco, and D. Baker. Contact order revisited: Influence of protein size on the folding rate. *Protein Sci*, 12(9):2057–2062, 2003.
- [73] O. V. Galzitskaya, S. O. Garbuzynskiy, D. N. Ivankov, and A. V. Finkelstein. Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics. *Proteins*, 51(2):162–166, 2003.
- [74] M. S. Li, D. K. Klimov, and D. Thirumalai. Thermal denaturation and folding rates of single domain proteins: size matters. *Polymer*, 45(2):573–579, 2004.
- [75] V. Munoz, A. N. Naganathan, and J. M. Sanchez-Ruiz. Direct measurement of barrier heights in protein folding. *J Am Chem Soc*, 127(51):17970–17971, 2005.
- [76] A. N. Naganathan and V. Munoz. Scaling of folding times with protein size. *J Am Chem Soc*, 127(2):480–481, 2005.
- [77] M. Kouza, M. S. Li, E. P. O’Brien, C. K. Hu, and D. Thirumalai. Effect of finite size on cooperativity and rates of protein folding. *J Phys Chem A*, 110:671–676, 2006.
- [78] S. S. Plotkin, J. Wang, and P. G. Wolynes. Statistical mechanics of a correlated energy landscape model for protein folding funnels. *J Chem Phys*, 106(7):2932–2948, 1997.
- [79] D. Baker, K. W. Plaxco, and K. T. Simons. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol*, 277(4):985–994, 1998.
- [80] C. M. Dobson, F. Chiti, N. Taddei, P. M. White, M. Bucciantini, F. Magherini, and M. Stefani. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat Struct Biol*, 6(11):1005–1009, 1999.
- [81] A. R. Dinner and M. Karplus. The roles of stability and contact order in determining protein folding rates. *Nat Struct Biol*, 8(1):21–22, 2001.
- [82] S. S. Plotkin. Speeding protein folding beyond the go model: How a little frustration sometimes helps. *Proteins: Struct Funct Bioinform*, 45(4):337–345, 2001.