Supporting Information

Liu et al. 10.1073/pnas.1019500108

SI Methods

Coarse-Grained Self-Organized Polymer (SOP)-SC Model for Polypeptide Chains. In the SOP-SC representation the effective energy function is taken to be

$$E_P(\{r_i\}) = V_{\rm FENE} + V_{LJ}^{\rm NAT} + V^{\rm NEI} + V_{LJ}^{\rm NN}. \quad [S1]$$

The finite extensible nonlinear elastic potential (FENE), $V_{\rm FENE}$, which describes the chain connectivity between backbones and side-chains, is given by

$$\begin{aligned} V_{\rm FENE} &= V_{\rm FENE}^{\rm bb} + V_{\rm FENE}^{\rm bs} \\ &= -\sum_{i=1}^{N-1} \frac{k}{2} R_o^2 \log\left(1 - \frac{(r_{i,i+1_{\rm bb}} - r^o_{i,i+1_{\rm bb}})^2}{R_o^2}\right) \\ &\quad - \sum_{i=1}^{N} \frac{k}{2} R_o^2 \log\left(1 - \frac{(r_{i,i_{\rm bs}} - r^o_{i,i_{\rm bs}})^2}{R_o^2}\right). \end{aligned} \quad [S2]$$

The nonbonded interaction, $V_{LJ}^{\rm NAT}$ in Eq. **S1**, which accounts for the stability of the folded structures, is taken to be

$$\begin{aligned} V_{LJ}^{\rm NAT} &= V_{LJ_{\rm NAT}}^{\rm bb} + V_{LJ_{\rm NAT}}^{\rm ss} + V_{LJ_{\rm NAT}}^{\rm bs} \\ &= \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} e_{\rm bb}\left[\left(\frac{r^o_{i,j_{\rm bb}}}{r_{i,j_{\rm bb}}}\right)^{12} - 2\left(\frac{r^o_{i,j_{\rm bb}}}{r_{i,j_{\rm bb}}}\right)^{6}\right]\Delta_{ij}^{\rm bb} \\ &\quad + \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} e_{\rm ss}|\epsilon_{ij} - 0.7|\left[\left(\frac{r^o_{i,j_{\rm ss}}}{r_{i,j_{\rm ss}}}\right)^{12} - 2\left(\frac{r^o_{i,j_{\rm ss}}}{r_{i,j_{\rm ss}}}\right)^{6}\right]\Delta_{ij}^{\rm ss} \\ &\quad + \sum_{\substack{i=1,j=1\\ |i-j|\geq 3}}^{N} e_{\rm bs}\left[\left(\frac{r^o_{i,j_{\rm bs}}}{r_{i,j_{\rm bs}}}\right)^{12} - 2\left(\frac{r^o_{i,j_{\rm bs}}}{r_{i,j_{\rm bs}}}\right)^{6}\right]\Delta_{ij}^{\rm bs}. \end{aligned} \quad [S3]$$

If the distance between two noncovalently linked beads, $r_{ij}(|i-j| \geq 3)$ in the Protein Data Bank (PDB) structure is within a cutoff distance R_c, a native contact is formed and correspondingly $\Delta_{ij} = 1$; if $r_{ij} > R_c$, $\Delta_{ij} = 0$. The strengths of nonbonded interactions $e_{\rm bb}$, $e_{\rm ss}$, $e_{\rm bs}$ are assumed to be uniform. The Betancourt–Thirumalai (BT) (1) statistical potential matrix with elements ϵ_{ij}, is used to explicitly treat the dependence of side chain–side chain interactions on the side-chain type.

We used repulsive interactions to account for excluded volume effects between neighboring beads with strength e_l. The ranges of repulsion are $\sigma_{\rm bb}$, $\sigma_{i,j_{\rm ss}}$, $\sigma_{j_{\rm bs}}$ for backbone–backbone (bb), side chain-side chain (ss), and backbone–sidechain (bs) interactions respectively. The form of $V^{\rm NEI}$ is

$$\begin{aligned} V^{\rm NEI} &= V_{\rm NEI}^{\rm bb} + V_{\rm NEI}^{\rm ss} + V_{\rm NEI}^{\rm bs} \\ &= \sum_{i=1}^{N-2} e_l\left(\frac{\sigma_{\rm bb}}{r_{i,i+2_{\rm bb}}}\right)^6 + \sum_{i=1}^{N-1} e_l\left(\frac{\sigma_{i,i+1_{\rm ss}}}{r_{i,i+1_{\rm ss}}}\right)^6 \\ &\quad + \sum_{i=1}^{N-2} e_l\left(\frac{\sigma_{i,i+2_{\rm ss}}}{r_{i,i+2_{\rm ss}}}\right)^6 + \sum_{\substack{i=1,j=1\\ 0<|i-j|<3}}^{N} e_l\left(\frac{\sigma_{j_{\rm bs}}}{r_{i,j_{\rm bs}}}\right)^6. \end{aligned} \quad [S4]$$

To prevent interchain crossing, we choose $\sigma_{\rm bb} = a = 3.8$ Å (a is average distance between neighboring C_α atoms), $\sigma_{i,j_{\rm ss}} = f(\sigma_i + \sigma_j)(\sigma_i, \sigma_j$ are the van der Waals radii of the side chains and $f = 0.5$), $\sigma_{j_{\rm bs}} = f(a + \sigma_j)$.

The nonbonded nonnative interactions are given by

$$\begin{aligned} V_{LJ}^{\rm NN} &= V_{LJ_{\rm NN}}^{\rm bb} + V_{LJ_{\rm NN}}^{\rm ss} + V_{LJ_{\rm NN}}^{\rm bs} \\ &= \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} e_l\left(\frac{\sigma_{\rm bb}}{r_{i,j_{\rm bb}}}\right)^6(1-\Delta_{ij}^{\rm bb}) + \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} e_l\left(\frac{\sigma_{i,j_{\rm ss}}}{r_{i,j_{\rm ss}}}\right)^6(1-\Delta_{ij}^{\rm ss}) \\ &\quad + \sum_{\substack{i=1,j=1\\ |i-j|\geq 3}}^{N} e_l\left(\frac{\sigma_{j_{\rm bs}}}{r_{i,j_{\rm bs}}}\right)^6(1-\Delta_{ij}^{\rm bs}). \end{aligned} \quad [S5]$$

Besides the knowledge-based BT statistical potential, the SOP-SC energy function $E_P(\{r_i\})$ has seven parameters (see Table S1 for the values) of which R_o and k merely account for chain connectivity. In effect there are only five parameters. Analysis of PDB structures shows that $R_c = 8$ Å is a reasonable choice, which has also been used in several previous studies. We have shown previously by analyzing structures of folded proteins in the PDB that there are large favorable bs and bb contacts (2). The first and third terms account for these interactions, which in turn ensures that packing effects are appropriately described in the SOP-SC model. The van der Waals radii for the side chains are in Table S2.

Langevin and Brownian Dynamics Simulations. We assume that the folding dynamics is governed by the Langevin equation, which includes a damping term with a friction coefficient ζ and a Gaussian random force Γ. The equation of motion for a generalized coordinate r_i is $m\ddot{r}_i = -\zeta\dot{r}_i + F_c + \Gamma$ where m is the mass of a bead, $F_c = -\partial(E_P(\{r_i\}) + \Delta G(\{r_i\},[C]))/\partial r_i$, is the conformational force calculated using Eq. **3** in the main text, Γ is the random force with a white noise spectrum. The autocorrelation function for $\Gamma(t)$ in the discretized form is $\langle\Gamma(t)\Gamma(t+nh)\rangle = \frac{2\zeta k_B T}{h}\delta_{0,n}$ (3) where $\delta_{0,n}$ is the Kronecker delta function and $n = 0,1,2\ldots$. The value of the time step, h, depends on the friction coefficient ζ.

We use a low friction coefficient $\zeta = 0.05\ {\rm m}/\tau_L$ to obtain enhanced sampling (4), which allows us to accurately calculate the equilibrium properties that do not depend on ζ. Rapid convergence of thermodynamics is possible at a small ζ because the polypeptide chain makes frequent transitions between all accessible states (Fig. S1*A*). In the underdamped limit, which is inappropriate for describing the folding kinetics, we use the Verlet leap-frog algorithm to integrate the equation of motion. The velocity at time $t + h/2$ and the position at time $t + h$ of a bead are given by,

$$v_i(t+h/2) = \frac{2m - h\zeta}{2m + h\zeta}\cdot v_i(t-h/2) + \frac{2h}{2m+h\zeta}[F_c(t) + \Gamma(t)], \quad [S6]$$

$$r_i(t+h) = r_i(t) + h\cdot v_i(t+h/2). \quad [S7]$$

To simulate the kinetics of folding we use $\zeta = 50\ {\rm m}/\tau_L$, which approximately corresponds to the value in water and represents the overdamped limit (3). At the high ζ value, we use the Brownian dynamics algorithm (5), which allows us to integrate equations of motion using

$$r_i(t+h) = r_i(t) + \frac{h}{\zeta}(F_c(t) + \Gamma(t)). \quad [S8]$$

The typical time step used to integrate Eq. **S8** ranges from $0.01\tau_H$ to $0.1\tau_H$ depending on $[C]$, the denaturant concentration. The integration time step, h, in Eqs. **S6** and **S7** is $0.005\tau_L$.

Time Scales. The natural unit of time for overdamped condition at the simulation temperature T_s is

$$\tau_H \approx \frac{\zeta_H a^2}{k_B T_s} = \frac{(\zeta_H \tau_L/m)e_l}{k_B T_s}\tau_L. \quad \textbf{[S9]}$$

To convert simulation time to real time, we chose $e_l =$ 1 kcal/mol, average mass $m = 1.8 \times 10^{-22}$ g (3), $a = 4$ Å, which makes $\tau_L = 2$ ps. For $\zeta_H = 50\ \mathrm{m}/\tau_L$ we obtain $\tau_H =$ 148 ps.

Folding and Unfolding Kinetics. The folding (unfolding) kinetics is monitored using fraction of molecules, $P_u(t)$, that remains unfolded (folded) at time t,

$$P_u(t) = 1 - \int_0^t P_{\mathrm{fp}}(s)ds. \quad \textbf{[S10]}$$

The distribution of first passage times, $P_{\mathrm{fp}}(s)$, is

$$P_{\mathrm{fp}}(s) = \frac{1}{M}\sum_{i=1}^{M}\delta(s - \tau_{1i}), \quad \textbf{[S11]}$$

where τ_{1i} is the first passage time for the ith trajectory, when the protein reaches for the first time either the native basin of attraction (NBA) in the folding process or the unfolded basin of attraction (UBA) upon unfolding for the first time, and M is the total number of trajectories. The folding time, $\tau_f([C])$, (or the unfolding time, $\tau_u([C])$) is associated with the mean first passage time

$$\tau_{\mathrm{MFPT}} = \int_0^\infty tP_{\mathrm{fp}}(t)dt = \int_0^\infty P_u(t)dt. \quad \textbf{[S12]}$$

For some concentrations $P_u(t)$ can be fit using a single exponential and in some cases biexponential provides a better fit. We extracted the folding times using $P_u(t)$ fits and Eq. **S12**. The folding rate is $k_f([C]) = (\tau_f([C]))^{-1}$, and the unfolding rate is $k_u([C]) = (\tau_u([C]))^{-1}$.

Dependence of $\ln k_{\mathrm{obs}}$ on $[C]$. The rate constant for unfolding, $k_u([C])$, is found to increase with increasing $[C]$ according to

$$\ln k_u([C]) = \ln k_u^{H_2O} + m_u[C]/RT, \quad \textbf{[S13]}$$

where $k_u^{H_2O} \equiv k_u([C] = 0)$ is the rate constant for unfolding in water ($[C] = 0$) and m_u, a constant, is the slope of the unfolding arm in the chevron plot. From the linear dependence of $\Delta G_{NU}([C])$ on $[C]$ and Eq. **S13** it follows that the rate constant for folding $k_f([C])$ with $k_f^{H_2O} \equiv k_f([C] = 0)$ must be (6)

$$\ln k_f([C]) = \ln k_f^{H_2O} - m_f[C]/RT, \quad \textbf{[S14]}$$

because $m = m_f + m_u$, and $K_U = k_f([C])/k_u([C])$, m_f is the slope of the folding arm in the chevron plot, $K_U = e^{-\Delta G_{NU}/RT}$, is the equilibrium constant of unfolding. The relaxation rate $k_{\mathrm{obs}} = k_f([C]) + k_u([C])$ is fit using (6)

$$\ln k_{\mathrm{obs}} = \ln[k_f^{H_2O}e^{-m_f[C]/RT} + k_u^{H_2O}e^{m_u[C]/RT}]. \quad \textbf{[S15]}$$

Because $k_u^{H_2O}$ is small and hence hard to obtain directly using simulations, we fit $\ln k_{\mathrm{obs}}$ using 3 parameters m_f, m_u, $k_u^{H_2O}$. Because of the extrapolation the error in our estimates for $k_u^{H_2O}$ is likely to be large.

Structural Analysis. The structural overlap function

$$\chi = 1 - \frac{N_k}{N_T}, \quad \textbf{[S16]}$$

between a conformation k and the native state

$$N_k = \sum_{i=1}^{N-3}\sum_{j=i+3}^{N}\Theta(\delta - |r_{i,j_{\mathrm{bb}}} - r^o_{i,j_{\mathrm{bb}}}|) + \sum_{i=1}^{N-3}\sum_{j=i+3}^{N}\Theta(\delta - |r_{i,j_{\mathrm{ss}}} - r^o_{i,j_{\mathrm{ss}}}|) + \sum_{\substack{i=1,j=1\\ |i-j|\geq 3}}^{N}\Theta(\delta - |r_{i,j_{\mathrm{bs}}} - r^o_{i,j_{\mathrm{bs}}}|), \quad \textbf{[S17]}$$

is used to monitor the folding reaction; r_{ij} is the distance between interaction centers i and j, r^o_{ij} is the value in the native conformation, and $\Theta(x)$ is the Heavyside function, $\delta = 2$ Å. The number of interacting pairs that are within $\delta = 2$ Å in the kth conformation is N_k, and N_T, which is obtained by setting all the $\Theta(x)$ in Eq. **S17** to unity, is equal to 5,724 for src SH_3 in the folded state. In computing χ we include only residues that are separated by greater than 2 along the sequence. The fraction, $f_{\mathrm{NBA}}([C],T)$ of molecules in the NBA is given by

$$f_{\mathrm{NBA}}([C],T) = \frac{1}{Z([C],T)}\sum_{k=1}^{R}\sum_{t=1}^{n_k} \times \frac{\delta(\chi_c - \chi(k,t,[0]))e^{-\beta(E_P(k,t,[0])+\Delta G(k,t,[C]))}}{\sum_{m=1}^{R} n_m e^{f_m - \beta_m E_P(k,t,[0])}}, \quad \textbf{[S18]}$$

where $Z([C],T)$ is the partition function, R is the number of independent simulations, n_k is the number of conformations from the kth simulation, χ_c is cutoff value for NBA (see Fig. S1), $\delta(\chi_c - \chi(k,t,[0])) = 1$ if $\chi(k,t,[0]) \leq \chi_c$. $\beta = 1/k_B T$, k_B is Boltzmann's constant, and T is the temperature. In Eq. **S18**, n_m and f_m correspond to the number of conformations and free energy in the mth simulation, respectively.

The ensemble of conformations with $\chi \leq \chi_c$ belong to the NBA. To determine we calculated the distribution $P(\chi)$, at the melting temperature $T_m([C])$. At $T_m([C] = 0)$ the order parameter makes frequent transitions from low χ (NBA) to high χ (UBA) (Fig. S1*A*). From the observed bimodal distribution $P(\chi)$ (Fig. S1*B*) at $T_m([C] = 0)$ we surmise that $\chi_c = 0.65$. Conformations with $\chi > \chi_c$ corresponds to UBA. Although χ_c is in principle $[C]$-dependent we find that in src-SH_3 domain χ_c is nearly independent of $[C]$.

1. Betancourt M, Thirumalai D (1999) Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci* 8:361–369.
2. Dima R, Thirumalai D (2004) Asymmetry in the shapes of folded and denatured states of proteins. *J Phys Chem B* 108:6564–6570.
3. Veitshans T, Klimov D, Thirumalai D (1997) Protein folding kinetics: Timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold Des* 2:1–22.
4. Honeycutt JD, Thirumalai D (1992) The nature of folded states of globular-proteins. *Biopolymers* 32:695–709.
5. Ermak DL, Mccammon JA (1978) Brownian dynamics with hydrodynamic interactions. *J Chem Phys* 69:1352–1360.
6. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor-2 1. Evidence for a 2-state transition. *Biochemistry* 30:10428–10435.

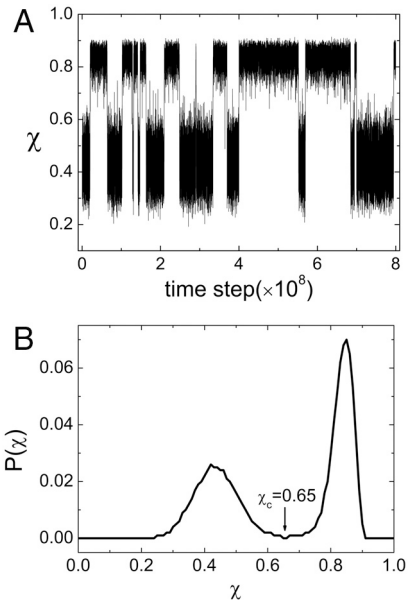

Fig. S1. Sampling of the order parameter χ. (*A*) Dependence of χ as a function of time steps generated at low ζ at the melting temperature with $[C] = 0$. (*B*) Distribution $P(\chi)$ of χ obtained using eight trajectories.

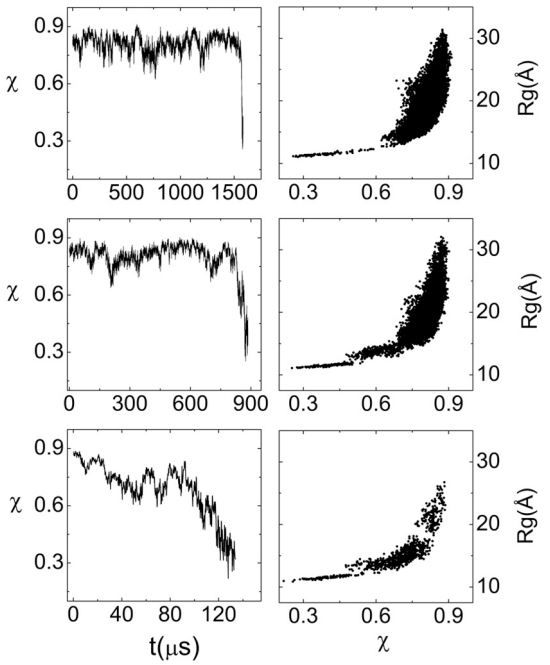

Fig. S2. Heterogeneity in the folding trajectories with $[C] = 0$. The left column shows χ (defined in (Eq. **4**) of the main text) for three trajectories. The top panel shows that χ remains in the UBA and reaches the NBA in the final stages. The trajectory in the middle panel is a state with $\chi \approx 0.6$ prior to transition to the folded conformation. In the last panel χ decreases nearly continuouly throughout the folding process. Projection of the fifty folding trajectories onto R_g and χ, are shown on the right. In all cases compaction and acquisition of structure coincide although there is dispersion in the range of R_g sampled.

Table S1. Parameters used in the SOP-SC model

Parameters	Values
R_o	2 Å
k	20 kcal/(mol· Å^2)
R_c	8 Å
e_{bb}	0.55 kcal/mol
e_{bs}	0.4 kcal/mol
e_{ss}	0.3 kcal/mol
e_l	1 kcal/mol

Table S2. Van der Waals radii for amino acids based on measured partial molar volumes

Residue	Radius (Å)
Gly	2.25
Ala	2.52
Val	2.93
Leu	3.09
Ile	3.09
Met	3.09
Phe	3.18
Pro	2.78
Ser	2.59
Thr	2.81
Asn	2.84
Gln	3.01
Tyr	3.23
Trp	3.39
Asp	2.79
Glu	2.96
Hse*	3.04
Hsd	3.04
Lys	3.18
Arg	3.28
Cys	2.74

*Hse—Neutral histidine, proton on NE2 atom. Hsd—Neutral histidine, proton on ND1 atom.

Table S3. Values of m_k and b_k in the transfer energy ($\delta g(k,[C])=m_k[C]+b_k$) of all residues and backbone (BB) for GdmCl

Residue	m_k (cal/mol/M)	b_k (cal/mol)
Gly	0.00	0.00
Ala	−7.20	−2.28
Val	−41.77	−23.41
Leu	−75.99	−41.42
Ile	−68.10	−37.93
Met	−85.86	−42.76
Phe	−124.57	−61.12
Pro	−50.86	−27.76
Ser	−18.75	−31.25
Thr	−18.75	−31.25
Asn	−102.03	−65.73
Gln	−56.90	−57.07
Tyr	−123.19	−78.71
Trp	−196.25	−138.75
Asp	−102.03	−65.73
Glu	−56.90	−57.07
His	−65.00	−85.00
Hsd	−65.00	−85.00
Lys	−67.95	0.00
Arg	42.34	0.00
BB	−39.21	−31.86

Table S4. Solvent accessibility of the backbone and side-chain groups of residue *k* in the tripeptide Gly-*k*-Gly ($\alpha_{k,\text{Gly-}k\text{-Gly}}$)

	$\alpha_{k,\text{Gly-}k\text{-Gly}}$, Å^2	
k	Backbone	Side chain
Gly	84.9817	0.000
Ala	62.5336	108.259
Val	53.8055	147.128
Leu	50.2648	164.683
Ile	50.2648	164.683
Met	50.2648	164.683
Phe	48.3400	174.605
Pro	56.8578	132.713
Ser	60.9227	114.940
Thr	56.2249	135.721
Asn	55.6039	138.647
Gln	52.0904	155.429
Tyr	47.3327	179.916
Trp	43.7780	198.715
Asp	56.6455	133.722
Glu	53.0339	150.786
Hse	51.3509	159.160
Hsd	51.3509	159.160
Lys	48.3400	174.605
Arg	46.1820	185.982
Cys	57.7220	128.640