Electronic Supplementary Material for Clementi and Plotkin (2004). The effects of non-native interactions on protein folding rates: Theory and Simulation. Protein Science 13(7).

APAPPENDIX A: ENTROPY OF A PARTIALLY COLLAPSED PROTEIN AS A FUNCTION OF THE NUMBER OF NATIVE AND NON-NATIVE CONTACTS

In terms of the packing fraction the total number of non-native contacts is

$$MA = M\eta(1-Q)\,, \tag{A.1}$$

where η is the packing fraction of non-native polymer surrounding the dense ($\eta = 1$) native core.

The mean-field configurational entropy of a self-avoiding polymer of n links with packing fraction η is given by [62, 63]

$$\frac{S_c^{\mathrm{SA}}(\eta)}{n} = \ln\frac{\nu}{\mathrm{e}} - \left(\frac{1-\eta}{\eta}\right)\ln(1-\eta) \tag{A.2}$$

The conformational entropy of the self-avoiding walk in terms of the fraction of non-native contacts A is given by

$$S_c^{\mathrm{SA}}(A) = S_c^{\mathrm{SA}}(\eta)\big|_{\eta=A/(1-Q)}\,. \tag{A.3}$$

Expressions (A.2) and (A.3) imply that the polymer chain in question will tend to have $\eta = 0$ and $A = 0$ since this maximizes the entropy. However a finite-length chain of n links tends to have a non-zero packing fraction given by

$$\eta(n) \approx \frac{na^3}{R_g(n)^3} \approx \frac{na^3}{\Delta R^3} \tag{A.4}$$

where a^3 is the volume per monomer and R_g is the radius of gyration of the chain. Up to factors of order unity the

RMS size of the polymer can be used as well. For chains obeying ideal statistics $\eta(n) \approx n^{-1/2}$. For self-avoiding chains in a good solvent, accounting for swelling gives $\eta(n) \approx n^{-4/5}$. However these expressions for the typical packing fraction are *inconsistent* with expression (A.2), which implicitly assumes an infinite chain limit. For finite-length chains, we seek an entropy function which is peaked at non-zero values of η.

The assumption of ideal chain statistics for protein segments is not as bad as it may at first seem, because disordered polymer segments interact with each other in addition to themselves. Polymers in a melt obey Gaussian statistics [64]. Swelling due to excluded volume is counterbalanced by compression due to the surrounding polymer medium if the protein is sufficiently large. However, for polymer loops dressing a native core, self-avoidance must be taken into account to fully treat the effects of non-native interactions.

We take the effects of self-avoidance, finite size, and "inter-loop" interactions into account by letting the number of walks with density η be the number of states at density η, $\exp S_c(\eta)$ above, times the probability that an ideal walk of ℓ steps has density η:

$$\Omega(\eta, \ell) = \mathrm{e}^{S_c(\eta,\ell)}\, p(\eta|\ell)\,. \tag{A.5}$$

For smaller values of ℓ, larger values of η are more probable. But at higher values of Q, smaller values of ℓ are more probable. Hence the non-native packing fraction tends to increase with folding. This is the effect we are quantifying here.

The number of states of the disordered polymer with packing fraction η, at degree of nativeness Q, is given by

$$\Omega(\eta, Q) = \prod_{\ell} \Omega(\eta, \ell)\, n(\ell|Q) = \prod_{\ell} \mathrm{e}^{S_c(\eta,\ell)}\, p(\eta|\ell)\, n(\ell|Q)\,. \tag{A.6}$$

This is the product over all lengths ℓ, of the number of states for a loop of length ℓ and packing fraction η, times the probability that the loop of finite length ℓ has packing fraction η, times the number of disordered loops of length ℓ at nativeness Q.

We now seek the probability distribution $p(\eta|N)$. Consider for the moment one dimensional random walks of

N steps, which we generalize to three dimensions below. The probability $p(\eta|N)$ is maximal at the value of η corresponding to a Gaussian distribution for the chain (i.e. $N^{-1/2}$ above). Again however, this alone does not account for self-avoidance, which is why $S_c(\eta, \ell)$ must be included later in the analysis. If we let the fraction of walks with variance $\lambda N a^2$ by given by $p(\lambda|N)$, the problem of finding $p(\eta|N)$ is equivalent to the problem of finding $p(\lambda|N)$. This is the probability a walk of N steps has an anomalous variance of $\lambda N a^2$, given that the most-probable distribution of walks $\overline{p}$ is given by

$$\overline{p}(x) = (2\pi N a^2)^{-1/2} \exp\left(-\frac{x^2}{2Na^2}\right) . \tag{A.7}$$

The probability $p(\lambda, N)$ can be written as a functional integral over all possible probability distributions, of the probability of a given distribution $P[p(x)]$, times a delta function which counts only those walks that have a given variance of $\lambda N a^2$:

$$p(\lambda|N) = \int \mathcal{D}p(x)\, P[p(x)]\, \delta\left(\lambda - \frac{1}{Na^2}\int dx\, x^2 p(x)\right) . \tag{A.8}$$

The calculation is performed in § B. The result for the probability distribution of anomalous variance λ is:

$$p(\lambda|N) = \sqrt{\frac{N}{6\pi}}\, \mathrm{e}^{-N(\lambda-1)^2/6} \tag{A.9}$$

We can see from equation (A.9) that the mean value of $\lambda = 1$, meaning that a walk of N steps has on average a variance Na^2. However there is variance $\delta\lambda^2 = 6/N$ in the distribution, so that some walks are either particularly diffuse or condensed statistically. The anomalous variance decreases monotonically with increasing N.

For a walk in three-dimensions, we define λ through the variance

$$\Delta \mathbf{R}^2 = \lambda N a^2 . \tag{A.10}$$

From the definition of η in equation (A.4), the parameter λ depends on η (and N) as

$$\lambda(\eta) = \eta^{-2/3} N^{-1/3} \tag{A.11}$$

The probability distribution of walks of density η is then given by

$$p(\eta|N) = p(\lambda(\eta)|N) \left| \frac{d\lambda}{d\eta} \right| \tag{A.12}$$

(the Jacobian is not particularly important here as it enters the entropy only logarithmically).

With the above definition in equation (A.10) for λ in three-dimensions, $p(\lambda|N)$ remains unchanged from the one-dimensional form in equation (A.9) (see Appendix B).

The conformational entropy for a chain of length ℓ having packing fraction η is obtained from equations (A.2), (A.5),(A.9), and (A.12):

$$S(\eta,\ell) = \ln\Omega(\eta,\ell) \approx S_c(\eta,\ell) - \frac{\ell}{6}\left[\left(\frac{\overline{\eta}}{\eta}\right)^{2/3} - 1\right]^2 \tag{A.13}$$

where $\overline{\eta} = \ell^{-1/2}$ gives the most probable value for the packing fraction for an ideal (non-self-avoiding) chain of length ℓ. For an interacting chain, enthalpy and entropy must both be considered in finding the most-probable packing fraction, which is obtained by minimizing the free energy with respect to η (see equations (12) and (13).

We still must find the dependence of loop length ℓ on the amount of native structure present. We proceed by making several approximations for the quantities in equation (A.6). The result is not sensitive to the exact values of these quantities. We approximate the product over loop lengths in equation (A.6) by taking a saddle-point value for ℓ, effectively letting all loops have the typical loop length $\overline{\ell}(Q)$. Then $n(\ell|Q) = \delta(\ell - \overline{\ell}(Q))n_{\mathrm{L}}(Q)$ where $n_{\mathrm{L}}(Q)$ is the total number of loops at Q. The typical loop length $\overline{\ell}(Q)$ is obtained from the total number of loops and the total number of disordered residues. We estimate the total number of disordered residues as a linear function of Q: $N(1-Q)$. This is a mean-field approximation. In capillarity models, the deviations from linearity scale as $N^{2/3}$, but are of order unity for a typical size protein (see Appendix C). We estimate the typical loop length $\overline{\ell}(Q)$ as the total number of disordered residues divided by the total number of loops:

$$\overline{\ell}(Q) \cong \frac{N(1-Q)}{n_{\rm L}(Q)} . \tag{A.14}$$

Generically for small native cores, the number of loops dressing the native core is proportional to the surface area of the core, which goes as the number of native residues NQ to the $2/3$ power. However for large native cores (a nearly folded protein), the unfolding nucleus consists of disordered protein, so that the number of constraints on loops within the core (the surface entropy cost) is proportional to the number of non-native residues $N(1-Q)$ to the $2/3$ power [4]. We linearly interpolate between these two regimes to obtain

$$\begin{aligned} n_{\rm L}(Q) &\approx (1-Q)[NQ]^{2/3} + Q[N(1-Q)]^{2/3} + 1 \\ &\approx N^{2/3}[Q(1-Q)]^{2/3}\left\{Q^{1/3} + (1-Q)^{1/3}\right\} + 1 \\ &\approx N^{2/3}[Q(1-Q)]^{2/3} + 1 \end{aligned} \tag{A.15}$$

where the expression in curly brackets is approximated as unity since it varies between 1 and about 1.6 over the range $0 \leq Q \leq 1$. One loop must always be present so that $\overline{\ell}(Q)$ remains non-divergent, so we have explicitly added unity in equation (A.15). Equations (A.14) and (A.15) together give the typical disordered loop length at Q in the model. Equation (A.15) is consistent with previous statements that the number of loops dressing the folding nucleus scales as $N^{2/3}$ [65], however here the Q-dependence is made explicit. When $Q = 0$ or $Q = 1$, $n_{\rm L} = 1$, and by (A.14) $\overline{\ell}(0) = N$, and $\overline{\ell}(1) = 0$, so the limits behave sensibly.

The entropy of the disordered polymer at Q, $S(\eta, Q)$, is then given by $n_{\rm L}(Q)S(\eta, \overline{\ell}(Q))$, or using equations (A.2), (A.13), and (A.14),

$$\begin{aligned} S_c(Q,\eta) &= N(1-Q)\left\{\ln\frac{\nu}{\epsilon} - \left(\frac{1-\eta}{\eta}\right)\ln(1-\eta) - \frac{1}{6}\left[\left(\frac{\overline{\eta}(Q)}{\eta}\right)^{2/3} - 1\right]^2\right\} \\ &\equiv N(1-Q)\, s_{nn}(Q,\eta) \end{aligned} \tag{A.16}$$

where $\overline{\eta}(Q) = \overline{\ell}(Q)^{-1/2} = [n_{\mathrm{L}}(Q)/N(1-Q)]^{1/2}$. In equation (A.16) the quantity in curly brackets is the entropy per residue for the remaining disordered polymer at Q. Equation (A.16) scales extensively with chain length, which is a consequence of the mean-field approximation made above.

APPENDIX B: CALCULATION OF THE PROBABILITY DISTRIBUTION OF ANOMALOUS VARIANCE

We again write the probability $p(\lambda, N)$ as a functional integral over all possible probability distributions, of the probability of a given distribution $P[p(x)]$, times a delta function which counts only those walks that have a given variance of $\lambda N a^2$:

$$p(\lambda, N) = \int \mathcal{D}p(x)\, P[p(x)]\, \delta\left(\lambda - \frac{1}{Na^2}\int dx\, x^2 p(x)\right) . \tag{B.1}$$

To obtain $P[p(x)]$ we imagine dividing the x-axis up into bins of width dx, where each bin is labeled by i, has coordinate $x_i = idx$, and we let $\overline{p}(x_i)dx \equiv \overline{p}_i$. The probability after N trials or events, of a distribution of numbers $\{n_i\}$ across all the bins is a multinomial distribution of essentially infinitely many variables

$$p\{n_i\} = \frac{N!}{\ldots n_1!\, n_2! \ldots} \cdots \overline{p}_1^{n_1}\, \overline{p}_2^{n_2} \cdots \tag{B.2}$$

Expanding the log of $p\{n_i\}$ to second order, subject to the constraint that $\sum n_i = N$, and using Stirling's formula, gives

$$p\{n_i\} = \left(\prod_i 2\pi N \overline{p}_i (1-\overline{p}_i)\right)^{-1/2} \exp\left(-\sum_i \frac{\left(n_i - N\overline{p}_i\right)^2}{2N\overline{p}_i(1-\overline{p}_i)}\right) \tag{B.3}$$

This is the distribution in the limit of large N. We apply it with the understanding that when N is not so large the distribution is an approximate solution. The approximation is best where n_i is the largest, which is where the distribution is most appreciable.

In the continuum limit $p\{n_i\} \longrightarrow P[p(x)]$, so that equation (B.1) can be written as

$$p(\lambda, N) = \frac{1}{2\pi} \int dk\, \mathrm{e}^{-ik\lambda} \int \mathcal{D}p(x)\, \mathrm{e}^{\int dx\, \mathcal{L}(p,x,k)} \tag{B.4}$$

where we have Fourier transformed the delta function. The effective Lagrangian here is

$$\mathcal{L}(p,x,k) = -N \frac{(p(x) - \overline{p}(x))^2}{2\overline{p}(x)} + ik \frac{x^2}{Na^2} p(x) \tag{B.5}$$

where we have used the fact that the probability to be within a given slice of width dx is small.

The functional integral amounts to finding the extremum of the effective action in the exponent. The extremal probability $p^*(x) = \overline{p}(x) + ik \frac{x^2}{N^2a^2} \overline{p}(x)$ and the extremal action $S^*(k) = \int dx\, \mathcal{L}(p^*, x, k) = -\frac{3}{2N}k^2 + ik$. The integral over k is then a simple Gaussian integral, so the result for the probability of anomalous variance is

$$p(\lambda, N) = \sqrt{\frac{N}{6\pi}}\, \mathrm{e}^{-N(\lambda-1)^2/6} \tag{B.6}$$

For a walk in three-dimensions, there are three parameters characterizing anomalous variance in x, y, and z. Since e.g. steps in y are uncorrelated from those in x, the probability of finding parameters λ_x, λ_y, and λ_z is the product of three terms each of the form (B.6), but formally with $1/3$ the number of steps in each of the three dimensions:

$$\begin{aligned} p(\lambda_x, \lambda_y, \lambda_z, N) &= p(\lambda_x, N/3)\, p(\lambda_y, N/3)\, p(\lambda_z, N/3) \\ &= \left(\frac{N}{18\pi}\right)^{3/2} \mathrm{e}^{-\frac{N}{18}\left[(\lambda_x-1)^2 + (\lambda_y-1)^2 + (\lambda_z-1)^2\right]} \end{aligned} \tag{B.7}$$

The variance $\Delta\mathbf{R}^2$ is given by

$$\begin{aligned}\Delta\mathbf{R}^2 &= \Delta x^2+\Delta y^2+\Delta z^2 \\ &= \frac{Na^2}{3}\left(\lambda_x+\lambda_y+\lambda_z\right) \\ &\equiv \lambda N a^2 \end{aligned} \qquad \text{(B.8)}$$

so that we seek the probability distribution $p(\lambda, N)$ of $\lambda = (\lambda_x+\lambda_y+\lambda_z)/3$. This is given by

$$\begin{aligned} p(\lambda,N) &= \int d\lambda_x\, d\lambda_y\, d\lambda_z \left(\frac{N}{18\pi}\right)^{3/2} \mathrm{e}^{-\frac{N}{18}\left[(\lambda_x-1)^2+(\lambda_y-1)^2+(\lambda_z-1)^2\right]}\cdot\delta\left(\frac{\lambda_x+\lambda_y+\lambda_z}{3}-\lambda\right) \\ &= \int d\lambda_x\, d\lambda_y\, 3\left(\frac{N}{18\pi}\right)^{3/2} \mathrm{e}^{-\frac{N}{18}\left[(\lambda_x-1)^2+(\lambda_y-1)^2+(3\lambda-\lambda_x-\lambda_y-1)^2\right]} \\ &= \sqrt{\frac{N}{6\pi}}\,\mathrm{e}^{-N(\lambda-1)^2/6} \end{aligned} \qquad \text{(B.9)}$$

as in the one-dimensional case.

APPENDIX C: NUMBER OF DISORDERED RESIDUES FOR A GIVEN NUMBER OF NATIVE CONTACTS

We wish to find the number of disordered residues when a fraction Q of native contacts are present. Equivalently we can find the number of ordered (native) residues. In the capillarity model this is the number of residues N_{NUC} in the nucleus. The number of native interactions at Q can be written as the total number of residues N times the mean number of interactions per residue in the native structure z_{N}, times the fraction of possible native interactions Q. The number of native interactions in a capillarity nucleus is the number of interactions in a fully collapsed (Hamiltonian) walk [4], which has bulk and surface contributions, giving the equation

$$N z_{\mathrm{N}} Q = z_{\mathrm{B}}\left(N_{\mathrm{NUC}} - \sigma N_{\mathrm{NUC}}^{2/3}\right)\ , \qquad \text{(C.1)}$$

where z_B is the number of native interactions per residue in a nucleus of infinite size, and σ is the mean fraction of the z_B interactions lost at the surface. In the absence of roughening σ is a very weak function of N and is of order unity. For walks on a 3-D cubic lattice $\sigma = 1.5$.

In our problem we know the number of native interactions, Nz_N. We can find z_B by solving (C.1) when $N_\mathrm{NUC} = N$:

$$z_\mathrm{B} = \frac{z_\mathrm{N}}{1-\sigma N^{-1/3}} . \tag{C.2}$$

The number of native residues N_NUC in a capillarity model as a function of Q is then given by the solution of

$$N_\mathrm{NUC} - \sigma N_\mathrm{NUC}^{2/3} = \left(N - \sigma N^{2/3}\right) Q . \tag{C.3}$$

Equation (C.3) is a cubic equation in $N_\mathrm{NUC}^{1/3}$, with solution of the form

$$N_\mathrm{NUC}(Q) = \left[\frac{1}{3}\left(\sigma + \frac{\sigma^2}{A^{1/3}} + A^{1/3}\right)\right]^3 \tag{C.4}$$

where

$$\begin{aligned} A &= (B + \sqrt{B^2 - 4\sigma^6})/2 \\ B &= 2\sigma^3 + 27NQ - 27N^{2/3}Q\sigma . \end{aligned}$$

Along with the average loop length, the total number of disordered residues determines the number of loops at Q. A plot of the total number of disordered residues for both the capillarity model and the linear approximation is shown in figure 11. One can see from the figure that a linear approximation for the number of disordered residues is a good one.

APPENDIX D: SIMULATION MODEL AND METHOD

We introduce non-native interactions to an otherwise energetically unfrustrated C_α model of SH3 domain of *src tyrosine–protein kinease* (src-SH3). The energetically unfrustrated model is obtained by applying a Gō-like Hamiltonian [66] to an off–lattice minimalist representation of the src-SH3 native structure (pdb-code 1fmk, segment 84-140). We have previously shown that this topology-based model is able to correctly reproduce the folding mechanism of small, fast-folding proteins [25, 26]. A standard Gō–like Hamiltonian takes into account only native interactions, and each of these interactions contributes to the energy with the same weight. Protein residues are represented as single beads centered in their C–α positions. Adjacent beads are strung together into a polymer chain by means of bond and angle interactions. The geometry of the native state is encoded in the dihedral angle potential and a non–local potential. The Gō-like energy of a protein in a configuration Γ (with native state Γ_N) is given by the expression:

$$E(\Gamma,\Gamma_N)_{\text{G}\bar{\text{o}}} = \sum_{bonds} K_r (r-r_N)^2 + \sum_{angles} K_\theta (\theta-\theta_N)^2 + \tag{D.1}$$

$$+ \sum_{dihedral} K_\phi^{(n)} [1+\cos(n\times(\phi-\phi_0))] + \tag{D.2}$$

$$+ \sum_{i<j-3} \left\{ \epsilon_1(i,j) \left[6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} - 5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} \right] + \epsilon_2(i,j)\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} \right\} \tag{D.3}$$

where r and r_N represent the distances between two subsequent residues in, respectively, the configuration Γ and the native state Γ_N. Analogously, θ (θ_N), and ϕ (ϕ_0), represent the angles formed by three subsequent residues, and the dihedral angles defined by four subsequent residues, in the configuration Γ (Γ_N). The dihedral potential consists of a sum of two terms for every four adjacent C_α atoms, one with period $n = 1$ and one with $n = 3$. The last term in equation (D.3) contains the non–local native interactions and a short range repulsive term for non–native pairs (i.e. $\epsilon_1(i,j) = constant < 0$ and $\epsilon_2(i,j) = 0$ if i–j is a native pair, while $\epsilon_1(i,j) = 0$ and $\epsilon_2(i,j) = constant > 0$ if i–j is a non–native pair). The parameter σ_{ij} is taken equal to i–j native distance for native interactions, while $\sigma_{ij} = 4$A for non-native pairs. Parameters K_r, K_θ, K_ϕ, ϵ weight the relative strength of each kind of interaction entering in the energy and they are taken to be $K_r = 100\epsilon$, $K_\theta = 20\epsilon$, $K_\phi^{(1)} = \epsilon$ and $K_\phi^{(3)} = 0.5\epsilon$.

We introduce a progressively increasing perturbation to the Gō–like Hamiltonian by replacing the short range repulsive term in equation (D.3) with attractive or repulsive pairwise interactions $V_{nn}(r_{i,j})$ in the form:

$$V_{nn}(r_{ij}) = \begin{cases} \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} + \eta_{i,j}\left[1 - \frac{1}{2}\left(\frac{r_{i,j}}{r_N}\right)^{20}\right] & \text{if } r_{i,j} < r_N, \\ \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} + \frac{\eta_{i,j}}{2}\left(\frac{r_N}{r_{i,j}}\right)^{20} & \text{if } r_{i,j} > r_N. \end{cases} \tag{D.4}$$

Figure 12 shows non-native interactions for different values of the interaction strength η. The strength $\eta_{i,j}$ for each non-native pair (i,j) is extracted randomly from a Gaussian distribution with mean ϵ_{NN} and variance b^2. The parameter $\sigma_{i,j}$ in expression D.4 is kept equal to 4A for all non-native interactions, in order to recover the plain Gō like Hamiltonian (equation D.3) in the limit $b \to 0$, $\epsilon_{NN} \to 0$. The parameter r_N is set to $r_N = \frac{4}{3}\sigma_{i,j}$. The selected values for $\sigma_{i,j}$ and r_N allow non-native contacts to form in the range of $r_{i,j} \sim 4-5$A . The total energy of a configuration Γ (with a native state Γ_N), corresponding to a non-native perturbation strength b, is thus:

$$E(\Gamma,\Gamma_N)_b = E(\Gamma,\Gamma_N)_{\bar{Go}} + \sum_{non\text{–}native(i,j)} V_{nn}(r_{i,j},\{\eta_b\}), \tag{D.5}$$

where $\{\eta_b\}$ is a set of quenched variable randomly distributed as described above. The case of $b = 0$, $\epsilon_{NN} = 0$ corresponds to the unperturbed Gō-like representation of the protein, as it has been studied in refs. [25, 26], and we use it as reference case for comparing the folding rates and folding mechanism. Sequences with different amount of non-native energy are defined by progressively increasing the parameter b in the interval $[0,2]\epsilon$ while keeping $\epsilon_{NN} = 0$, or by varying the parameter ϵ_{NN} in the interval $[-1,1]\epsilon$.

The native contact map of a protein is obtained by using the approach described in ref. [67]. Native contacts between pairs of residues (i,j) with $j \leq i+3$ are discarded from the native map as any three and four subsequent residues are already interacting in the angle and dihedral terms. A contact between two residues (i,j) (native or non-native) is considered formed if the distance between the C_α's is shorter than γ times their equilibrium distance σ_{ij} (where σ_{ij} = native distance for a native pair, and σ_{ij} = 4A for a non-native pair). It has been shown [68] that the results are not strongly dependent on the choice made for the cut–off distance γ. We have chosen $\gamma = 1.2$ as in refs. [25, 26]. We have used constant temperature Molecular Dynamics (MD) for simulating the kinetics and thermodynamics of the protein models. We employed the simulation package AMBER (Version 6) [69] and Berendsen algorithm for coupling the system to an external bath [70].

For each Hamiltonian (obtained for different values of the parameter b), several constant temperature simulations were combined using the WHAM algorithm [71, 72] to generate a specific heat profile versus temperature and a free energy $F(Q)$ as a function of the folding reaction coordinates Q and A. In order to compute folding rates, several (typically 250) simulations are performed at the estimated folding temperature for each different sequence. The folding time τ is then defined as the average time interval between two subsequent unfolding and folding events over this set of simulations. The time length of a typical simulation is about 5×10^6 MD time steps. In this time range 2 to 5 folding events are normally observed for the unperturbed Gō-like protein model.

The errors (reported as error bars in the plots) on the estimates of thermodynamic quantities and folding rates are obtained by computing these quantities from several (more than 100) uncorrelated sets of simulations and then considering the dispersion of values obtained for the same quantity.