Supporting Information

Lin et al. 10.1073/pnas.1416300111

SI Text

Ensemble-convergent molecular-dynamics (MD) simulations were performed using the CHARMM suite of programs (1) and the CHARMM22/CMAP (2, 3) force field on ensembles of 25-residue-long polyalanine chains (A_{25}) in explicit water, and the results were analyzed with the aid of VMD (4). The macromolecules were solvated by 7,583 TIP3P (5) water molecules in a cubic box with periodic boundary conditions; the sides of the box equilibrated to 60 Å at $T = 278$ K. Long-range electrostatic interactions were computed using the particle mesh Ewald method (6, 7). Initially, the system was heated to $T_0 = 500$ K during 500 ps, and 287 conformational snapshots were randomly selected from the last 500 ps of the resulting 5-ns trajectory to construct a macromolecular ensemble of the unfolded state. The initial structure of each of the 287 members of the ensemble was energy minimized for 12,000 steps in aqueous solution and then heated to $T = 278$ K and preequilibrated for 100 ps. To study the folding of A_{25} in a regime strongly favoring α-helix formation, the final temperature was chosen to be $T = 278$ K; for each trajectory, the MD simulation time window was 49 ns long.

To assess the fraction of intact native contacts as well as both local and global structural changes throughout the ensemble as a function of time, structural changes were measured as follows. First, for all sets of independent trajectories, the fraction of each native (helical) contact remaining intact at time t was calculated. The fraction of intact hydrogen bonds was obtained for every native-contact pair and further averaged over the $n = 500$ independent trajectories to obtain the average decay of each native contact as a function of time. A hydrogen bond was defined to be 100% intact if the distance between the donated proton and the nitrogen or oxygen atom (the hydrogen acceptor) was less than 1.8 Å and the angle defined by N–H bond of the donor residue and O atom of the acceptor residue was at least 120°. In addition, the smoothness of the transition between fully intact and a fully broken hydrogen bonds was enforced using an exponential attenuation of the bond strength such that the hydrogen bond would be $1/e$-fold intact at a distance of 2.5 Å. We note that the above criteria are consistent with established conventions for geometry-based hydrogen bond determination (8). The same criteria were used when monitoring ensemble-wide nonnative hydrogen bonds such as formation/decay of β-hairpin–type structures during the course A_{25} folding.

Given that carrying out massively distributed ensemble-convergent MD simulations constitutes a challenge, even for relatively small polypeptides, developing a simple yet predictive analytical model of (un)folding would greatly aid the detailed exploration of the physicochemical processes involved. More importantly, analytical methods, even those limited to low structural resolution, provide the crucial causal connection between fundamental driving forces and the resulting collective behavior that may be difficult to obtain otherwise. In what follows, we provide a detailed description of the kinetic-intermediate structure (KIS) model of α-helix folding/unfolding.

The free-energy landscape $\Delta G(i, j)$, which defines the folding behavior of the polypeptide, was obtained by calculating the free energy for each (i, j)-state with respect to the unfolded state of the macromolecular ensemble:

$$\Delta G(i,j) = (L - i - j - 1)\Delta G_{\text{prop}}(T) + \Delta G_{\text{nuc}}(T). \qquad \textbf{[S1]}$$

Here, L is the number of native hydrogen-bonding contacts, $\Delta G_{\text{prop}}(T)$ is the free-energy change associated with forming a single backbone hydrogen bond to propagate the helix, and $\Delta G_{\text{nuc}}(T)$ is free-energy change corresponding to the formation the helix nucleus. The (temperature-dependent) free-energy differences can be obtained from the corresponding enthalpy and entropy changes: $\Delta G(T) = \Delta H - T\Delta S$. The interstate ("misfolding") barriers crossed during the course of $(i, j - 1) \leftarrow (i, j) \rightarrow (i - 1, j)$ transitions between individual microstates are given by the following:

$$\Delta G^{misfold}_{(i,j)\rightarrow(i-1,j)} = kT \ln\left[P_\beta(i;i,j)e^{\Delta H_\beta/kT} + 1 - P_\beta(i;i,j)\right], \qquad \textbf{[S2a]}$$

$$\Delta G^{misfold}_{(i,j)\rightarrow(i,j-1)} = kT \ln\left[P_\beta(L-j;i,j)e^{\Delta H_\beta/kT} + 1 - P_\beta(L-j;i,j)\right]. \qquad \textbf{[S2b]}$$

Here, ΔH_β is the enthalpic barrier associated with breaking a backbone–backbone (β-hairpin–type) hydrogen bond and $P_\beta(m; i, j)$ is the probability of having to break such a nonnative hydrogen bond at position $1 \le m \le L$ of the state (i, j) to complete the helix elongation step. We note that $P_\beta(m; i, j)$ is given by the statistical weight of the subpopulation of the state (i, j), which is characterized by a nonnative hydrogen bond at position m, divided by the statistical weight of the state (i, j):

$$P_\beta(m;i,j) = \frac{\exp(-G(m;i,j)/kT)}{\exp(-G(i,j)/kT)}. \qquad \textbf{[S3]}$$

For A_{25}, we involved the experimental enthalpic and entropic values associated with nucleation and elongation of α-helices and β-hairpin–like contact formation in polyalanine, as well as the conformational entropy change associated with formation of a loop. For a polypeptide chain composed of a set of amino acid residues $\{k\}$, $k \in [1, k_{\max}]$, a (native) hydrogen-bond formation between the C=O group of one amino acid (residue k_0; $1 \le k_0 \le k_{\max} - 4$) and the N–H group of another amino acid located four residues upstream along the chain (residue $k_0 + 4$) gives rise to a single turn of an α-helix. In the present study, the KIS model, which was originally used to parameterize the unfolding of DNA/RNA hairpins in aqueous solution following a T-jump (9–11), is extended to describe the formation of α-helices.

For a given state (i, j), let N be the number of microstates in which the β-hairpin–type nonnative contacts are formed. The denominator of Eq. **S3** can then be expressed as the sum of statistical weights of the N microstates as follows:

$$\exp(-G(i,j)/kT) = \sum_{n=1}^{N} \exp\left(\frac{-d(n)\Delta G_\beta(T) - \Delta G_{\text{loop}}(n)}{kT}\right), \qquad \textbf{[S4]}$$

where $d(n)$ is the number of β-hairpin–type contacts, $\Delta G_\beta(T)$ is the free-energy change upon formation of a β-hairpin–type contact, and $\Delta G_{\text{loop}}(n)$ is mainly due to the entropy correction associated with loop formation in microstate n. We note that all N states in the sum can be computationally generated. The numerator of Eq. **S3** is calculated in the same way as the denominator except for the constraint that there is a nonnative contact at position m.

The experimentally obtained thermodynamics parameters used in the KIS model of folding of A_{25} (in units of kilocalories per mole) are given by the following: $\Delta G_{\text{prop}}(T) = -1.3 + 0.004T$ (12–14), $\Delta G_{\text{nuc}}(T) = -1.3 + 0.014T$ (15), $\Delta H_\beta = 4$, and $\Delta G_\beta(T) = -1 + 0.002T$ (16). We note that the barrier height ΔH_β is estimated

from the electrostatic stabilization associated with formation of a β-hairpin and that its magnitude is reduced to account for solvation effects when the enthalpic component of $\Delta G_{\beta}(T)$ is calculated. Finally, for a given misfolded microstate n, $\Delta G_{\mathrm{loop}}(n)$ is obtained from the entropy of a polymer loop of length l, which can be expressed as $A\ln(l/l_{\mathrm{ref}})$ (17), with the reference loop length $l_{\mathrm{ref}} = 10$, and the excluded volume prefactor $A = -2.4$ kcal/mol (18). We note that the loops characterized by $l < 3$ are disregarded within the framework of the model.

For the dynamical simulations on the KIS landscape (both with and without the misfolding barriers), each Monte Carlo step corresponds to a trial time $t_{\mathrm{prop}} = 12\,\eta\,V_{\mathrm{a}}\,V_{\mathrm{b}}/\mathrm{kT}(V_{\mathrm{a}} + V_{\mathrm{b}})$, which is the (entropy-dominated) elongation time step that corresponds to dividing the helix nucleation (19) time step by the square of the probability ($P = 1/2$) of a set of backbone dihedral angles being in the helical basin because elongation requires fixing one set of backbone dihedral angles rather than three. Here, η is the (temperature-dependent) viscosity of water and $V_{\mathrm{a/b}}$ correspond to the effective van der Waals volumes of the peptide fragments that are conformationally diffusing on either side of the helical contact i or j on the KIS landscape. Each residue contributes 600 Å^3 to $V_{\mathrm{a/b}}$ until $V_{\mathrm{a/b}}$ saturates at the Kuhn length of three residues. For elongation at the end of the peptide at room temperature, t_{prop} is about 2 ns.

Time evolution of the helix content as obtained from the Monte Carlo dynamics simulations on the KIS landscape (Fig. S2*A*) was fitted to a biexponential form $H(t) = A_1(1 - 2^{-t/\tau 1}) + A_2(1 - 2^{-t/\tau 2})$, with $A_1 + A_2$ being equal to the final (equilibrium) helicity $H(\infty)$. We note that, for all of the simulated temperatures, $A_1 >> A_2$ and $\tau_1 >> \tau_2$. At late simulation times, i.e., when $(1/\tau_1)\,2^{-t/\tau 1} >> (1/\tau_2)\,2^{-t/\tau 2}$, the slope of $\mathrm{Log}[H(\infty) - H(t)]$ corresponds to $(1/\tau_1)$, whereas at short simulated times the slope is given by $[A_1(1/\tau_1) + A_2(1/\tau_2)]/(A_1 + A_2)$. Because $H(0) = 0$, the y intercept in Fig. 4*A* is the final helicity fraction at $t = \infty$, and the slope of the line at $t = \infty$ is the inverse of the slower timescale $1/\tau_1$. The faster temporal component τ_2 is associated with rapid zipping along the diagonal of the landscape, and the (order of magnitude slower) temporal component τ_1 stems from near-the-edge folding pathways, $i >> j$ and $j >> i$.

The kinetic flux through a state (i, j) on the KIS free-energy landscape is defined as a traffic through that state, i.e., it is calculated as $\min[\mathrm{in}(i, j), \mathrm{out}(i, j)]$, where $\mathrm{in}(i, j)$ and $\mathrm{out}(i, j)$ are the net inflow and net outflow of some 5,000 Monte Carlo trajectories associated with all of the states neighboring—or kinetically connected to—the state (i, j). Thus, e.g., the native state $(0, 0)$ on the KIS free-energy landscape has zero kinetic flux because $\mathrm{out}(0, 0) = 0$ (we note that a trajectory generates zero flux through a state if it arrives at the state and subsequently leaves it through the same pathway). A comprehensive picture of the evolution of the abovementioned populations with time, including Monte Carlo dynamics in the absence of the off-pathway interstate barriers on the KIS free-energy landscape, is provided in Movies S1–S4 (Fig. 3).

1. Brooks BR, et al. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4(2):187–217.
2. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–3616.
3. Mackerell AD, Jr, Feig M, Brooks CL, 3rd (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25(11):1400–1415.
4. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–28.
5. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935.
6. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An NLog(N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089–10092.
7. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19): 8577–8593.
8. Torshin IY, Weber IT, Harrison RW (2002) Geometric criteria of hydrogen bonds in proteins and identification of "bifurcated" hydrogen bonds. *Protein Eng* 15(5): 359–363.
9. Lin MM, Meinhold L, Shorokhov D, Zewail AH (2008) Unfolding and melting of DNA (RNA) hairpins: The concept of structure-specific 2D dynamic landscapes. *Phys Chem Chem Phys* 10(29):4227–4239.
10. Lin MM, Shorokhov D, Zewail AH (2009) Structural ultrafast dynamics of macromolecules: Diffraction of free DNA and effect of hydration. *Phys Chem Chem Phys* 11(45):10619–10632.
11. Crow JM (2008) Computational model reveals how DNA and RNA fold into hairpins: Genetic code does the twist. *Chem Biol* 3(7):B50.
12. Scholtz JM, Qian H, Robbins VH, Baldwin RL (1993) The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry* 32(37):9668–9676.
13. Schellman JA (1955) The stability of hydrogen-bonded peptide structures in aqueous solution. *C R Trav Lab Carlsberg Chim* 29(14-15):230–259.
14. Yang AS, Honig B (1995) Free energy determinants of secondary structure formation: I. α-Helices. *J Mol Biol* 252(3):351–365.
15. Wang L, O'Connell T, Tropsha A, Hermans J (1995) Thermodynamic parameters for the helix-coil transition of oligopeptides: Molecular dynamics simulation with the peptide growth method. *Proc Natl Acad Sci USA* 92(24):10924–10928.
16. Yang AS, Honig B (1995) Free energy determinants of secondary structure formation: II. Antiparallel β-sheets. *J Mol Biol* 252(3):366–376.
17. Chan HS, Dill KA (1989) Intrachain loops in polymers: Effects of excluded volume. *J Chem Phys* 90(1):492–509.
18. Nagi AD, Regan L (1997) An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des* 2(1):67–75.
19. Lin MM, Mohammed OF, Jas GS, Zewail AH (2011) Speed limit of protein folding evidenced in secondary structure dynamics. *Proc Natl Acad Sci USA* 108(40):16622–16627.

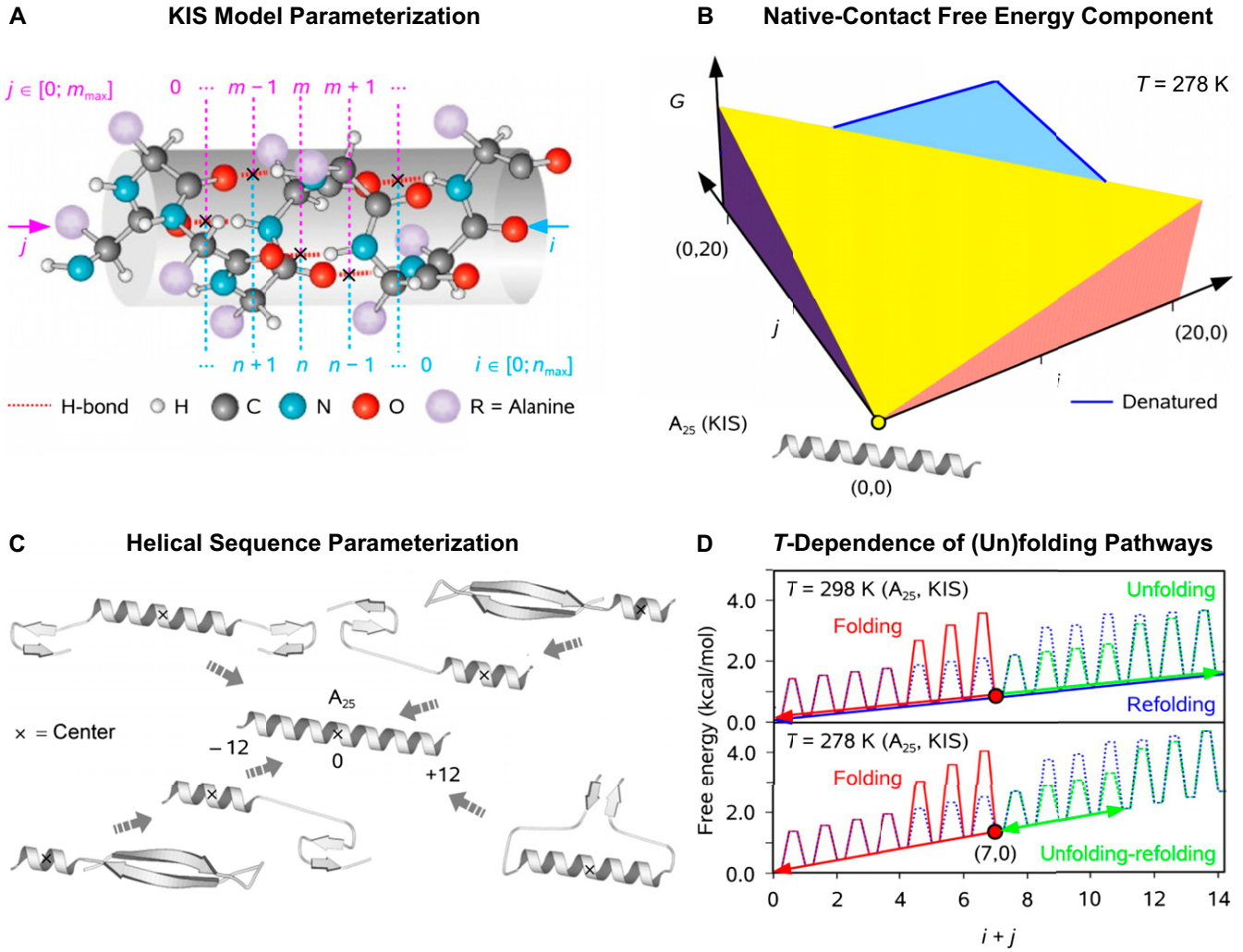


Fig. S1. Structure of the canonical α-helix with the KIS order parameters (*i*, *j*) (*A*). Native-hydrogen bond contact free-energy landscape of A_{25} as obtained from KIS model excluding off-pathway interstate barriers (*B*). Elongation of α-helical sequences is mostly hindered by formation of long, single-end–capping hairpins ("left + right"), with the share of end-to-end contacts being negligible. Individual residues *k* are numbered from –24 to +24 with the center of each helical stretch in the ensemble fixed at the origin by definition (*C*). Increasing availability of the $i \sim j$ folding locus at higher temperatures caused by (entropy-driven) tilting of the underlying native-contact free-energy landscape leads to the onset of rapid zipping along the diagonal of KIS landscape (*D*).

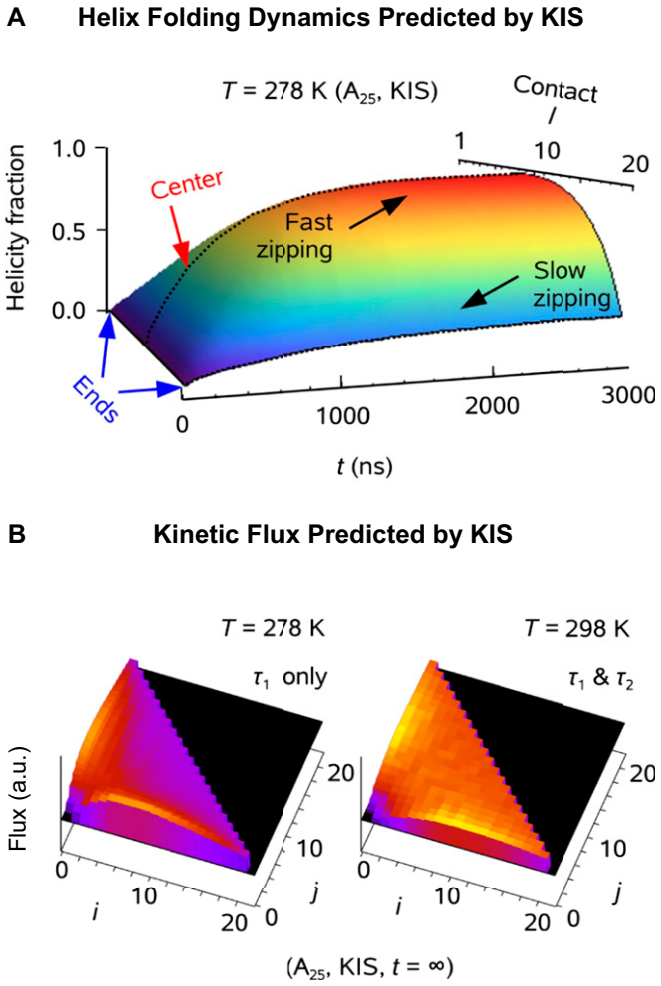
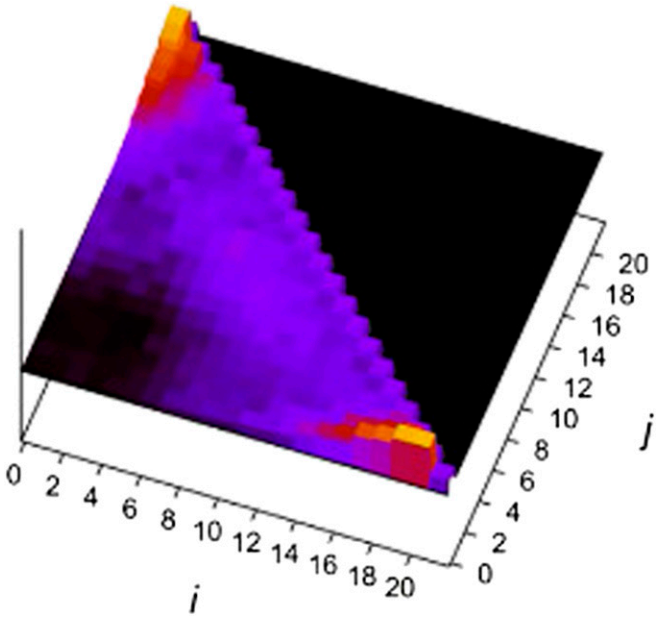
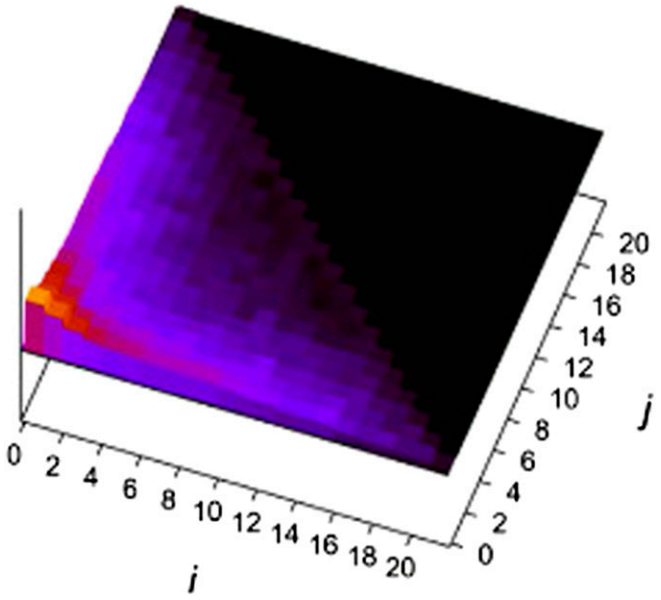


Fig. S2. Helix folding dynamics of A_{25} as obtained from the KIS free-energy landscape at $T = 278$ K. Helical nuclei that form close to the middle of the sequence tend to elongate more rapidly, whereas misfolded structures that follow after near-the-end helix nucleation significantly hinder the folding process (*A*). As evidenced in the "kinetic flux" patterns characteristic of A_{25} at $T = 278$ K and 298 K resulting from Monte Carlo dynamics simulated on the KIS landscape (*B*), the increasing availability of the $i \sim j$ folding locus at higher temperatures caused by (entropy-driven) tilting of the underlying native-hydrogen bond contact free-energy landscape leads to the onset of rapid zipping along the diagonal of the landscape.



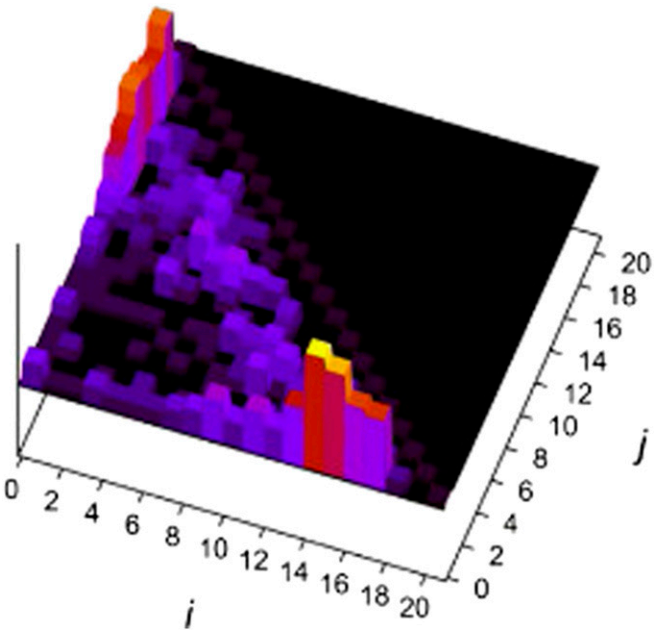
Movie S1. Temporal evolution of macromolecular ensemble populations of A_{25} at $T = 278$ K as obtained for $t = 50$ ns from Monte Carlo dynamics on the KIS landscape including off-pathway interstate barriers.

Movie S1



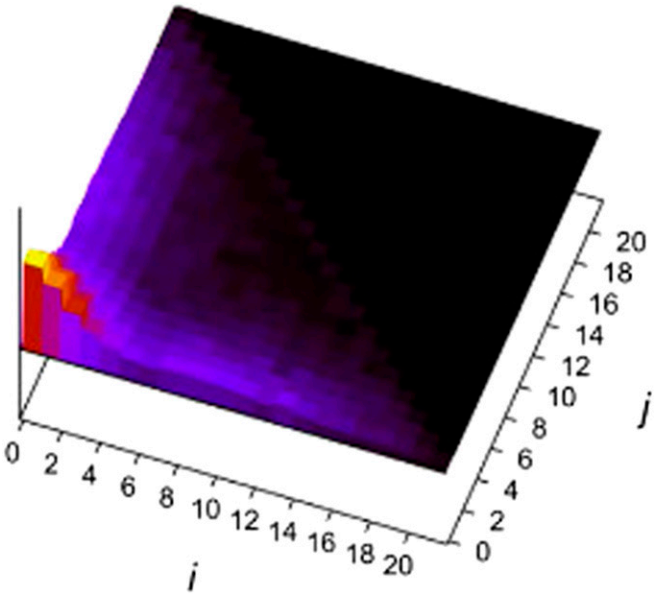
Movie S2. Temporal evolution of macromolecular ensemble populations of A_{25} at $T = 278$ K as obtained for $t = 50$ ns from Monte Carlo dynamics on the KIS landscape excluding off-pathway interstate barriers.

Movie S2



Movie S3. Temporal evolution of macromolecular ensemble populations of A_{25} at $T = 278$ K on the KIS parameter space for $t = 50$ ns, obtained from ensemble-converged MD simulations based on $n = 287$ independent trajectories.

Movie S3



Movie S4. Temporal evolution of macromolecular ensemble populations of A_{25} at $T = 278$ K as obtained for $t = 600$ ns from Monte Carlo dynamics on the KIS landscape including off-pathway interstate barriers. The frame rate is sped up by a factor of 5.

Movie S4