Supporting Information

Zhang and Ma 10.1073/pnas.1112143109

SI Text

Dihedral-Biased Tempering. In this section, we describe a dihedral-biased tempering technique that combines the virtues of a tempering method (1) and a dihedral bias (2–6). The tempering part accelerates a simulation by borrowing faster motion from higher temperatures. The dihedral bias improves sampling helical conformations and accelerates the dynamics in the dihedral space. The method preserves room-temperature properties, so the dihedral bias is added only on higher temperatures.

Tempering. Tempering is a sampling technique that circumvents a slow dynamics common in a regular room-temperature molecular dynamics (MD) by exploiting a faster motion at higher temperatures. Here, we employed a single-trajectory-based tempering method (1), which can readily handle large systems and thus is suitable for an explicit-solvent simulation that involves many water molecules.

The technical objective of tempering is to allow the temperature to vary in a certain range, but at the same time to make sure the statistical distribution at each individual temperature remains correct. To do so, we employ a generalized statistical ensemble composed of canonical ensembles from a temperature range. Typically, the temperature range starts from T_L, somewhere slightly below the room temperature $T_0 = 300$ K, to a much higher one ($T_H \approx 600$ K). We denoted the temperature by the reciprocal temperature $\beta = 1/(k_B T)$, with k_B being the Boltzmann constant, and similarly $\beta_0 = 1/(k_B T_0)$, $\beta_L = 1/(k_B T_L)$, $\beta_H = 1/(k_B T_H)$.

In the generalized ensemble, we maintain the desired statistical distribution by following a Langevin equation for β in each simulation step as

$$\frac{d(1/\beta)}{dt} = E - \tilde{E}(\beta) - \frac{\partial \log w(\beta)}{\partial \beta} + \frac{\sqrt{2}}{\beta}\xi, \qquad \text{[S1]}$$

where E is the current potential energy, $\tilde{E}(\beta)$ is the estimated average potential energy at β according to a runtime average, $w(\beta)$ is the overall temperature distribution density, and ξ is a Gaussian white noise that satisfies $\langle \xi(t) \cdot \xi(t') \rangle = \delta(t - t')$.

The only unknown in Eq. **S1** is $\tilde{E}(\beta)$. It is estimated from data collected in a window $(\beta - \Delta\beta, \beta + \Delta\beta)$ surrounding β using an integral identity; details can be found in ref. 1. We also used an adaptive updating scheme to quickly refresh data in the window. The scheme allows a fast random walk across the temperature space in the early stages of a simulation, and then gradually shifts to a more equilibrium sampling.

Tempering With a Bias Potential. On top of the tempering method, a bias potential $V(X)$, expressed as a function of molecular coordinates X, can be added to the generalized ensemble without disturbing room-temperature properties. This is achieved via a temperature-dependent Hamiltonian

$$H(\beta, X) = H_0(X) + \left(\frac{\beta_0}{\beta} - 1\right) V(X), \qquad \text{[S2]}$$

where $H_0(X)$ is the unbiased potential energy. Thus the magnitude of $V(X)$ is full at the β_H, but zero at $\beta = \beta_0$. The distribution of the generalized ensemble becomes

$$\begin{aligned}\rho(\beta, X) &= w(\beta)\frac{\exp[-\beta H(\beta, X)]}{Z(\beta)} \\ &= w(\beta)\frac{\exp[-\beta H_0(X) - (\beta_0 - \beta)V(X)]}{Z(\beta)}, \end{aligned} \qquad \text{[S3]}$$

where $Z(\beta) = \int \exp[-\beta H(\beta, X)]dX$ is the configuration space partition function. A regular tempering is recovered if $V(X) = 0$.

No matter what the bias potential is, the Hamiltonian at the room temperature is undisturbed, since $H(\beta_0, X) = H_0(X)$. The biased Hamiltonian also preserves a linear relation between the effective Hamiltonian $\beta H(\beta, X)$ and the temperature β; i.e.,

$$\beta H(\beta, X) = \beta_0 V(X) + [H_0(X) - V(X)]\beta.$$

Thus, room-temperature properties can be readily recovered via the multiple histogram method (7–9) as in a regular tempering. The only necessary modification is to substitute $H_0(X) - V(X)$ for $H_0(X)$ in using the method (also for E in Eq. S1) for

$$\langle H_0(X) - V(X) \rangle_\beta = -\partial \log \rho(\beta, X)/\partial\beta.$$

Dihedral Bias. In our case, the bias potential $V(X)$ applies only to the two backbone dihedrals of amino acid residues in a protein.

For each residue i, the potential handles the two dihedrals φ_i (C-N-C_a-C) and ψ_i(N-C_a-C-N) separately by two functions $V_\varphi(\varphi_i)$ and $V_\psi(\psi_i)$. Thus

$$V(X) = \sum_{a_i \neq G,P} V_\varphi[\varphi_i(X)] + V_\psi[\psi_i(X)],$$

where a_i denotes the one-letter amino acid name of residue i, $\varphi_i(X)$ and $\psi_i(X)$ are the two backbone dihedrals as functions of molecular coordinates X. After excluding all glycine (G) and proline (P) residues, the two functions V_φ and V_ψ apply to all residues without the distinction of residue type or index.

The problem now is to find $V_\varphi(\varphi)$ and $V_\psi(\psi)$ that properly bias towards the helical conformation. Since "the helical conformation" is hard to define precisely (especially at a high temperature), we directly learn it from the force field.

First, we allow both $V_\varphi(\varphi)$ and $V_\psi(\psi)$ to be linear combinations of a few dihedral modes, each of which roughly corresponds to a potential basin in the dihedral space towards a typical secondary structure. The coefficients of combination are user-adjustable; e.g., a larger coefficient of the helical mode increases the bias towards the helical conformation.

Second, we obtain the dihedral modes from decomposing the potentials of mean force (PMFs) along the two dihedrals. In this way, the modes are intrinsic to the force field and are usually smooth functions of the two dihedrals.

The PMFs were obtained by a special simulation on a helical protein α_3D (10) (PDB ID: 2A3D, also the first folding system) at the highest temperature $\beta_H = 0.2$ mol kJ^{-1} ($T_H \approx 600$ K) for 50 ns. The special simulation started from a fully extended chain (thus no native structure was required), and was independent of all folding simulations. In the simulation, we aimed at a flat distribution along each dihedral. To achieve this, we adaptively offset the projection of the force along each dihedral by its runtime-averaged value, or the mean force. After the simulation, the PMFs for φ and ψ were obtained by integrating the respective mean forces. In our implementation, the histogram along a dihe-

dral was not completely flat due to a small missing force component from constraints; the corresponding correction based on the histogram was thus applied. The approach allowed us to comprehensively sample the entire dihedral space, covering an otherwise rarely visited barrier region. However, since we only sampled extended conformations, the PMFs obtained were expected to deviate from the actual ones, which were dominated by the folded state with highly helical conformations.

We then decomposed the PMF (using φ as example) to three modes as

$$V^0_\varphi(\varphi) \approx V_\varphi(\varphi) = -\sum_{k=\mathrm{A,B,C}} M_k u_k(\varphi),$$

where $V^0_\varphi(\varphi)$ is the intrinsic PMF along φ; $V_\varphi(\varphi)$ is its best approximation by a superposition of modes; M_k is the amplitude of mode k ; $u_k(\varphi)$ specifies mode k as a function of φ. Note before the decomposition, we shift the PMF such that $V^0_\varphi(\varphi)$ is always negative, and thus every $u_k(x)$ is positive definite.

Each $u_k(\varphi)$ is a bell-shaped function, asymmetric for $\varphi > \varphi_c$ and $\varphi < \varphi_c$(φ_c being the center), and with up to 6th order moments on the exponent:

$$u_k(\varphi) = \exp[-a_2^{\pm}(\varphi-\varphi_c)^2 - a_4^{\pm}(\varphi-\varphi_c)^4 - a_6^{\pm}(\varphi-\varphi_c)^6],$$

where the superscripts + and – denote parameters for $\varphi > \varphi_c$ and $\varphi < \varphi_c$, respectively. The parameters $a_2^{\pm}$, $a_4^{\pm}$, $a_6^{\pm}$, and φ_c of different modes are determined via a least square fitting in minimizing the difference between $V^0_\varphi(\varphi)$ and $V_\varphi(\varphi)$

$$(\delta V_\varphi)^2 = \frac{1}{2\pi}\int_0^{2\pi} [V^0_\varphi(\varphi) - V_\varphi(\varphi)]^2 d\varphi.$$

The resulting parameters are listed in Table S6. As shown in Fig. S20, the decompositions authentically recovered the original PMFs: the average differences δV_φ and δV_ψ were 0.048 kJ mol^{-1} and 0.030 kJ mol^{-1}, respectively (representing about 0.2% relative error compared with the magnitudes of intrinsic PMFs).

For the dihedral φ, the three decomposed modes roughly corresponded to α-helix, β-strand, and left-handed helix conformations. However, the first two modes shared a large overlap in the dihedral space due to the broadening at the high temperature $T_H \approx 600$ K. For the dihedral ψ, the first two modes corresponded to α-helix, β-strand conformations, with the last one lying in the transition region.

For convenience, we used the above PMFs and decomposed dihedral modes in simulations of all proteins, although ideally one could perform the above procedure for every target protein.

Adjusting Dihedral Modes. With the dihedral modes in hand, the bias potential [using $V_\varphi(\varphi)$ for an example] is constructed as a superposition of modes. In the first step, we construct a rough $\tilde{V}_\varphi(\varphi)$ via three parameters, p_A, p_B and p_C, provided by the user, as

$$\tilde{V}_\varphi(\varphi) = \sum_{k=A,B,C} \min\{-k_\mathrm{B}T\log p_k, M_k\} \times u_k(\varphi). \qquad \textbf{[S4]}$$

Eq. **S4** means that if the expected number of visits to mode k ($k = A, B, C$) is to be multiplied by p_k fold, its magnitude in the bias potential is $\min\{-k_\mathrm{B}T\log p_k, M_k\}$. The min function annihilates the potential mode in case p_k is so small that $-k_\mathrm{B}T\log p_k$ exceeds the intrinsic magnitude M_k.

To promote mode k, we can either increase p_k or reduce $p_{k'}$s of other $k' \neq k$ modes. The latter produces a softer potential surface. $\tilde{V}_\varphi(\varphi)$ vanishes if all p_ks are 1.0, and the method is reduced to a regular tempering. On the other hand, if all p_ks are 0.0, $\tilde{V}_\varphi(\varphi)$ annihilates the intrinsic PMF $V_\varphi(\varphi)$ and recovers a flat histogram along either dihedral. A set of small p_ks reduces the magnitude of all modes, and thus encourages transitions among them. Note the populations of modes in a production simulation are expected to disagree with the p_ks given by the user (even if the simulation were run only at T_H instead of in the generalized ensemble) because the intrinsic PMFs were obtained from an extended state simulation instead of a native one and because we ignored correlations between the two dihedrals and those among different residues.

The second (optional) step is to polish $\tilde{V}_\varphi(\varphi)$ to reduce the free-energy barriers via the last parameter $p_{\min}$ ($p_{\min} \geq 1.0$). We first constructed an expected distribution $\rho_\varphi(\varphi) = \exp[-\beta V^0_\varphi(\varphi) - \beta\tilde{V}_\varphi(\varphi)]$. We then trimmed the potential such that the expected distribution was $\rho_\varphi(\varphi) \to \max\{\rho_\varphi(\varphi), \rho_{\min}p_{\min}\}$, where $\rho_{\min} = \min\rho_\varphi(\varphi)$. In this way, the highest free-energy barrier is lowered by $\log p_{\min}/\beta$.

Parameters p_A, p_B, p_C, and $p_{\min}$ of different trajectories are listed in Table S3.

Generalization to Other Types of Proteins. So far, both the tempering method and the decomposition of the dihedral modes have been general. However, an optimal bias potential could accelerate the folding.

In simulations presented in this study, we have assumed that the potential $V(X)$ is biased towards the helical conformation; i.e., the amplitude of the helical mode is set to be greater than those of the other modes. If extended structures are to be biased towards, then the relative amplitude should be inverted. Note, however, that generally the dynamics in the dihedral space is not the only contributing factor to protein folding. As shown in the trajectories, hydrophobic collapse and determination of the right geometry often take longer. Thus, more structural factors can be included in the bias potential.

In case a preferable $V(X)$ is lacking, a random trial strategy can be used. We can try several different bias potentials among independent trajectories [note the method is single-trajectory-based, thus the trial is relatively cheap in comparison of a parallel strategy (11–14)]. Naturally, only a helpful $V(X)$ would facilitate the folding.

We can monitor the average potential energy at the room temperature to see if a trajectory employs a helpful $V(X)$. As the native conformation is usually energetically favorable, a trajectory reaching the native conformation would have a lower internal energy. As the bias is absent at the room temperature, statistics from different trajectories can be compared there. Thus, as long as some of the bias potential promotes the folding, folding is accelerated in these trajectories before a complete equilibrium is reached.

For example, if in a set of trajectories, half of them adopt a $V(X)$ biased towards the helix conformation while the other half a $V(X)$ biased against it, then at least the former half can achieve a fast folding for helical proteins.

For other types of proteins, one could include in the bias potentials, in addition to the dihedral term, other structural factors; e.g., the strength of hydrogen bonds, the strength of the hydrophobic and electrostatic interactions, and some surface area terms. Different parameters within the potential can then be tried in different trajectories.

From a more technical viewpoint, we can also apply the bias potential through an additional sampling dimension at the roof temperature. In this way, even at higher temperatures one can switch between the biased and unbiased modes according to a Monte Carlo transition probability. The strategy would reduce the influence of the bias potential and produce a fraction of trajectories behaving like a regular tempering.

Performance of the Dihedral-Biased Tempering. ***Setup of the Testing Case.*** We used the first protein, α_3D (10), as a test case of the dihedral-biased tempering method. Starting from a fully extended chain, we simulated the protein using three different methods: a room-temperature (300 K) regular MD, a regular (unbiased) tempering, and a dihedral-biased tempering (this method). As a control, we also ran a regular 300 K MD starting from the native structure.

For the dihedral-biased tempering, we ran multiple independent trajectories, each of which used a different bias potential (parameters are listed in Table S3).

Convergence of Tempering. We first examined the convergence of the tempering part. We used trajectory 7 (Table S3) as an example. In Fig. S2*A* and *B*, we show the time series for the instantaneous potential energy and temperature, both of which underwent many round-trips in their respective range.

The convergence of tempering can be checked from the temperature histogram $H(\beta)$, which should be proportional to the intended temperature density $w(\beta)$. In Fig. S2*C*, we show that the ratio $H(\beta)/w(\beta)$ indeed approached a constant.

To illustrate the efficiency of tempering, we show in Fig. S2*D* the reconstructed potential energy histograms plotted at a temperature interval $\Delta\beta = 0.002$ from $\beta = 0.20$ to 0.42 (thus 111 histograms). A parallel tempering simulation would need about 100 copies running in parallel to cover the same temperature range with decent overlap between neighboring temperatures, and the computational cost would be much higher.

Folding Progress (RMSD). We used the backbone root-mean-square deviation (RMSD) from the experimentally determined (NMR) structure to monitor the progress of folding.

As shown in Fig. S3*A*, a typical RMSD at 300 K native simulation was 3.2 Å. In a regular MD starting from a fully extended conformation (a large spiral with RMSD≈19 Å), the RMSD lingered around 25 ∼ 32 Å, a typical value for the unfolded state.

In an unbiased tempering, the RMSD had a larger fluctuation from 6 Å to 25 Å, but a complete folding was unfortunately not achieved in 1 μs.

In trajectories using the dihedral-biased tempering, the RMSD readily dropped under 3.2 Å, indicating successful folding. Additionally, large-scale unfolding/refolding events occurred in both trajectories 2 and 7.

The comparison showed that the only the biased tempering method was able to quickly fold the protein and also sampled various unfolded states.

Folding Progress (Backbone Dihedrals). Similar to the previous case, we monitor the progress in the dihedral space by computing the fraction of correct backbone dihedrals (FCBD). We first define four backbone dihedrals conformations:

$\varphi < 0$ and $\psi \in [-100°, 80°)$.
$\varphi < 0$ and $\psi \in [-180°, -100°) \cup [80°, 180°)$.
$\varphi > 0$ and $\psi \in [-80°, 100°)$.
$\varphi > 0$ and $\psi \in [-180°, -80°) \cup [100°, 180°)$.

For each residue, a dihedral conformation is considered correct if it falls into the category as that in the native structure. FCBD was then computed as the fraction of residues with the correct dihedral conformations (excluding the two terminals).

As shown in Fig. S3*B*, after equilibration, the 300 K MD simulation from the native yielded a FCBD of about 80 ∼ 90%. The fraction could not, however, reach 100% possibly due to the flexibility of loop regions.

The 300 K regular MD from a fully extended chain yielded a low FCBD of about 15 ∼ 35%. The regular (unbiased) tempering allowed a larger fluctuation that spanned from 20% to 60%, but failed to reach the native level.

In the dihedral-biased tempering, a large FCBD was reached after the initial 1 ∼ 10 ns, long before a complete folding occurred. Thus, the bias helped the protein to quickly build up helical content, while the folding process was still limited by the formation of correct tertiary structure.

Convergence in the Dihedral Space. We further examined the convergence in the dihedral space. We wished to see if trajectories with different dihedral bias potentials reached the same dihedral distributions at T_0 as designed.

We computed the residue-averaged dihedral distributions at the room temperature $T_0 = 300$ K and a higher temperature 580 K, using the multiple histogram method (7–9). Since the Hamiltonian was unbiased only at 300 K, the dihedral distributions from simulations with different biases should agree with each other at 300 K, but not at 580 K.

There was, however, a practical difficulty for a direct comparison. Although different trajectories shared the same native conformation, they sampled different nonnative conformations, with each conformation manifesting slightly different dihedral distributions. Thus, the resulting distributions could differ due to simply a relatively slow convergence in sampling various nonnative conformations, not to the method itself. To avoid this distracting factor, we only included frames in the folded states by imposing the condition RMSD <6 Å, and filtered out all other frames. We emphasize, however, that the distributions computed here could only be used as a test of the method, not to represent the actual free-energy landscape due to the filtering.

As shown in Fig. S3*C–F* (for distributions of φ and ψ, respectively), the dihedral distributions at 300 K from three biased tempering trajectories agreed with each other, and also with the distributions from the native simulation at the same temperature. On the other hand, the distributions at 580 K differed significantly as expected.

Geometric Measures. We specify the mathematical definition of (*i*) the chiral distance of helix-packing, (*ii*) the distance between two helices, and (*iii*) the angle of two helices.

Representing a Helix as a Rod. We needed to model a helix by a representative rod. To do so, we first compute the eigenvalues of the moments of inertia $\mathbf{I} = \{I_{\alpha\beta}\}$ from all alpha carbon (C_α) atoms on the helix, where $I_{\alpha\beta} = \Sigma_i(r_{i\alpha} - r_{c\alpha})(r_{i\beta} - r_{c\beta})$ with $\alpha, \beta = x, y, z$; $\mathbf{r}_i = \{r_{ix}, r_{iy}, r_{iz}\}$ and $\mathbf{r}_c = \{r_{cx}, r_{cy}, r_{cz}\}$ are the coordinates of the ith C_α atom on the helix and the center of mass, respectively. Geometrically, the eigenvectors and eigenvalues of the matrix correspond to the directions and lengths of the principle axes of the representative ellipsoid. For the representative rod, we choose the eigenvector $\mathbf{v}_1$ ($|\mathbf{v}_1| = 1$) with the largest eigenvalue λ_1 as the direction, and the construct the two ends $\mathbf{b}_\pm$ by extending from the center of mass $\mathbf{r}_c$ oppositely along the direction as $\mathbf{b}_\pm = \mathbf{r}_c \pm \sqrt{3\lambda_1}\mathbf{v}_1$. It is easy to see that if atoms are densely and uniformly distributed on a thin rod of length l along x, then $I_{xx} = l^2/12$, and all other $I_{\alpha\beta}$'s are zeros. since $\lambda_1 = l/(2\sqrt{3})$, $\mathbf{r}_c \pm (l/2)\mathbf{v}_1$ recovers the ends of the rod as expected.

Chiral Distance for Packing Helices. We now define a chiral distance D for three helices A, B, and C. If helices A and B are parallel, D equals to the signed distance from helix C to the plane extended by A and B. If A and B are not parallel, the definition is generalized as follows: First, the middle helix B is modeled as a rod as in the above, with $\mathbf{b}_0$ and $\mathbf{b}_1$ being the two ends (connected to A and C, respectively). We can then compute, for any two points $\mathbf{a}$ and $\mathbf{c}$,

$$D(\mathbf{a}, \mathbf{b}_0, \mathbf{b}_1, \mathbf{c}) = \frac{(\mathbf{b}_0 - \mathbf{a}) \times (\mathbf{b}_1 - \mathbf{b}_0)}{\|(\mathbf{b}_0 - \mathbf{a}) \times (\mathbf{b}_1 - \mathbf{b}_0)\|} \cdot (\mathbf{c} - \mathbf{b}_1). \quad \text{[S5]}$$

Finally, we let **a** run through all C_α atoms on helix A and **c** through those on helix C, and compute the average as

$$D(\mathrm{A}, \mathrm{B}, \mathrm{C}) = \langle D(\mathbf{a}, \mathbf{b}_0, \mathbf{b}_1, \mathbf{c}) \rangle_{\mathbf{a} \text{ on helix A, } \mathbf{c} \text{ on helix C}}$$

$D(\mathrm{A}, \mathrm{B}, \mathrm{C})$ is positive if viewing along helix B from $\mathbf{b}_0$ to $\mathbf{b}_1$, helix A can be rotated clockwise to the position of helix C by an angle less than 180°, and is negative otherwise.

Helix Distance. We compute the distance between two helices A and B as follows: First, for a C_α atom **a** on helix A, we computed the minimal distance $d_{\mathbf{a}}$ to any C_α atom on helix B. Similarly $d_{\mathbf{b}}$ is computed for a C_α atom **b** on helix B. The average of all $d_{\mathbf{a}}s$ and $d_{\mathbf{b}}s$ is then defined as the distance between helices A and B.

Helix Angle. For two helices A and B, their angle is computed from the angle of the two corresponding representative rods.

Hydrophobic Contact Number. We compute the number of contacting hydrophobic residues as the number of pairs of C_α atoms on different helices within a 12 Å distance cutoff.

Clustering Method. ***Requirements From Statistical Mechanics.*** For a simulation trajectory, the objective of the clustering method is to form a reasonable set of groups of frames with similar structures, or a cluster configuration. From the viewpoint of statistical mechanics, the cluster configuration is merely a discrete distribution. Hence, the clustering method must be consistent with the principles of statistical mechanics.

The above requirement sets two conditions. First, the cluster method must be deterministic and independent of the order of frames, for otherwise it would not be an ensemble property.

Second, each frame i should be able to carry the weight w_i from the multiple histogram method (7–9). It ensures that the cluster configuration reflects the properties at the room temperature; i.e., high-temperature noises should be filtered out.

The first condition can be satisfied by setting the cluster configuration as the one that minimizes some target function H. The second requires the multiple-histogram weight to enter H.

The Target Function. The following target function satisfy the above conditions

$$H = \frac{1}{2}\left(\sum_c w_c \bar{d}_c\right) + \frac{1}{2} N_c \mu W,$$

where the sum is performed over all clusters c s; $w_c = \sum_{i \in c} w_i$ is the total weight of all frames in cluster c; $\bar{d}_c = \sum_{i,j \in c} d_{ij} w_i w_j / (\sum_{i \in c} w_i)^2$ is the average distance among any two frames in cluster c (the distance d_{ij} was computed as the backbone RMSD of two protein conformations i and j after a mutual structural alignment); N_c is the number of clusters; μ is the energy cost of forming a new cluster, or the "chemical potential"; and W is the total weight of all frames in the trajectory.

To see why the above target function produces a sensible cluster configuration, we considered the process of adding a new frame j to a cluster c, with its weight w_j much smaller than the total weight $w_c = \sum_{i \in c} w_i$ of the cluster. The resulting change to the target function is

$$\delta H_{j \to c} = \left(\frac{\sum_i d_{ij} w_i}{w_c} - \frac{\bar{d}_c}{2}\right) \frac{w_j}{w_c}. \quad \text{[S6]}$$

On the right hand side of Eq. **S6**, the first term in the parentheses gives the average distance from the new frame j to an existing frame in the cluster c, while the second the half of the average distance among all existing frames in the cluster c, or a "radius." Thus, the target function decreases if the new frame is closer to existing frames than the current radius. Further, the new frame j only joins the cluster that minimizes $\delta H_{j \to c}$.

Minimization of the Target Function. Now we only have to minimize the target function. Due to many local minima in the space of cluster configurations, we used a global minimizer (15). It requires a statistical ensemble of clustering configurations, which we define as

$$Z = \sum \left(\prod_c n_c!\right) \exp(-\beta H),$$

where β is an reciprocal temperature (just for the clustering, not to be confused with the actual temperature in sampling), and n_c is the number of frames in cluster c. The factor $n_c!$ encourages clusters to change sizes as it varies drastically with the cluster size n_c.

The number of resulting clusters depends on μ. A larger μ increases the cost of forming a new cluster and thus leads to fewer clusters. We used $\mu = 0.7$ Å, which usually yielded a single cluster for a 300 K native simulations but a handful of clusters for a folding trajectory.

The method is relatively expensive, since it is based on a thorough sampling of clustering configurations and operates on a full RMSD matrix of all frame pairs. Thus we skip frames at a regular interval to limit the total number of frames within 2000.

Initial Structures. The initial structures for folding trajectories were all prepared according to the following protocol. For each protein, the initial structure was a large extended spiral such that it fit into the simulation box. The program we used for generating the initial spiral can be found in the website mentioned in the article text. The resulting configuration was then solved in a simulation box with water molecules subsequently added. A mild (usually less than 1000 steps) energy minimization was preformed to remove strong frictions between the protein and its surrounding water molecules. A short (2–10 picoseconds) room-temperature MD simulation was then performed on the energy-minimized structure at 300 K to raise the energy to a normal level. The resulting structure was used as the initial structure for all folding trajectories of the same protein.

For the four existing proteins, we show in Fig. S12 the initial structures after the above procedures. The water molecules are not included for clarity.

Representative Structures at the Room Temperature. To illustrate the nature of typical native and nonnative structures found in folding trajectories, we shall analyze trajectory 7 for α_3D and trajectory 2 for $\alpha_3 D_L$ in detail. In Fig. S13, we plotted the RMSD as a function of time with representative structures shown at roughly a 300 ~ 500 ns interval. The representative structures must carry a weight $w > 0.5$ after the histogram reweighting method applied to 300 K to ensure that they were relevant to the room temperature.

For the trajectory of α_3D, the N-terminal and middle helices were roughly formed while the C-terminal helix was not at 300 ns. A fully folded structured was reached at the 600 ns and maintained to 1000 ns. However, a subsequent unfolding changed its chirality to a left-handed one, as shown in the 1400 ns frame. The left-handed structure was unstable, as indicated by a completely disintegrated C-terminal helix. Nevertheless, the geometry sustained to the 1700-ns frame, in which the N-terminal helix was loosened as well. The protein refolded at the 2000-ns frame

and unfolded again at the 2400-ns frame, in which a pair of short β-strands was formed.

For the trajectory of α_3D_L, at 200 ns, only a small fraction of structure was helical, while a β-hairpin and several random coils were formed. At 500 ns, the protein was almost folded, while the packing of the C-terminal helix against the other two was still loose. After an unfolding around 700 ns, the structures at 900 ns and 1200 ns consisted mainly of random coils, with only a small fraction of the helical structures. At 1500 ns, however, the helices *N* and *C* were aligned antiparallel, leading to a structure similar to the configuration 3 of α_3W and LQLQ; see Fig. S7.

The above observations suggest that even in the nonnative conformations, there was still a fraction of residual helical content (cf. Fig. S3*B*). This was in agreement with a recent MD simulation on Anton (16). Nevertheless, nonhelical structures such as coils and β-strands were also common.

Analyses on the Root-Mean-Square Deviation (RMSD). We computed the distributions of the root-mean-square deviation (RMSD) from the native conformation for each protein (at 300 K after histogram reweighting). For the mutants, the putative structures were used.

α_3D and α_3D_L. The RMSD distributions for α_3D were computed from all folding trajectories, and the results are shown in Fig. S14*1A*. The distribution can be roughly divided at RMSD = 8.0 Å; the left and right are for the "folded" (RMSD < 8.0 Å) and unfolded states (RMSD > 8.0 Å), respectively.

In six out of seven trajectories, the folded state itself had two separated peaks (the exception was trajectory 4, in which the first peak was missing). The representative structures at the two peaks from each trajectory were superimposed in Fig. S11*A*.

For each trajectory the representative structure was computed as follows: First, the frames within a RMSD range (for the first peak: 2.8 Å to 3.0 Å, the second: 4.7 Å to 5.0 Å) were collected. Then the frame with the lowest average distance from all other frames in the collection was selected as the representative structure. In the averaging, frames carried the multiple histogram weights to 300 K. The procedure was used to produce similar figures for other proteins. We shall thus only indicate the RMSD filtering ranges, which were also approximately indicated by the color strips in Figs. S11 to S15. Note the filtering ranges were artificial and did not represent physical clusters.

From Table S3, we observed that the cluster centers of the most folded trajectories had a RMSD of 2.5 Å to 2.7 Å, which was close to the value of the first peak. As discussed in VII.F, the first peak represented the geometric center, while the second peak represented some peripheral states surrounding it.

However, it does not mean that conformations in second peak were completely void of structure. In fact, they resulted from a relative sliding of helices along the helical axis. From Fig. S11*A*, it was clear that the structures in the second peak were still compact and resembled those in the first peak. We also note that the direction of the sliding was also not random; e.g., the C-terminal helix tended to take a higher (according to the orientation of the figure) position in the second peak than in the first peak. Thus, we speculate that the conformations at the second peak might correspond to a transition state that facilitated the folding (17).

For α_3D_L, the RMSD distributions manifested a similar two-peak feature, as shown in Fig. S11*B* (for the representative structures, the RMSD range for the first peak was 2.4 Å to 2.8 Å; the range for the second peak was 4.3 Å to 4.7 Å). Thus, the relative sliding between helices might be common in the transition states.

α_3W. Similar to the previous case, the RMSD distributions of α_3W also had two peaks in the folded state, as shown in Fig. S12. The representative structures for the two peaks were computed from the RMSD ranges (2.1 Å, 2.5 Å) and (4.2 Å, 5.0 Å), respectively.

The first peak was close to the geometric center of the folded state, as it shared a similar value with the RMSD of cluster centers, 1.7 Å ~ 2.4 Å, from trajectories 1–4 (see Table S3). Note trajectories 5 and 6 were less reliable as they adopted a stronger bias towards helical conformations.

The relative sliding of helices still contributed to the difference between the structures of the first and the second peaks (although other local adjustments also occurred). The sliding was again not completely random, e.g., Trp34 was consistently lower in the second peak.

Unlike the α_3D case, the second peak of the distribution was more populated than the first one. Although insufficient sampling could be the reason, we also suspect that the shift of population could result from a more symmetric sequence of α_3W than that of α_3D. Thus, the second peak became more entropically favorable, as there were more energetically inexpensive ways of sliding helices and other adjustments.

LQLQ, LALQ, and LALA. We computed the RMSD distributions of the symmetrical three-helix bundles LQLQ, LALQ, and LALA. The experimentally determined (left-handed) conformation of α_3W was used as a template to produce the reference conformations in the three cases. As shown in Fig. S13, the distributions of LQLQ and LALQ were similar. In either case, there was a peak for the left-handed conformation around RMSD = 2 Å, as well as a peak for the right-handed one around RMSD = 9 Å. In the case of LALA, the left-handed conformation was not sampled. However, the right-handed conformation had a broadened peak covering both the bend and straight conformations (cf. Fig. S9).

Fap1-NR$_\alpha$. The RMSD distribution Fap1-NR$_\alpha$ (Fig. S14) showed that the native conformation did not occupy the largest cluster. However, in trajectory 1, the native resided in the second largest cluster, whose population was about 45% of that of the largest one. This could result from a combination of insufficient sampling and force field error. Due to the larger size in comparison with the previous protein, simulations should also take longer to converge. Additionally, since the protein was a natural domain, the native basin was possibly narrower than those in the designed proteins, further increasing the difficulty.

The main problem of folding the protein was that the long C-terminal helix was hard to maintain at the room temperature even through the sampling method explicitly used a helical bias at higher temperatures. To see this, we observed that the peak at RMSD of 6.4 Å corresponded to a conformation with a kinked C-terminal helix around the short GNNT sequence (Gly68, Asn69, Asn70, and Thr71), which was reasonably "mistaken" as a loop in the trajectories. The highest peak occurred at a RMSD around 12.5 Å, corresponding to a conformation of a more globular shape. Interestingly, the GNNT segment also adopted a loop conformation instead of a helical one there.

S-836. In Fig. S15, we show that in the RMSD distribution of S-836, the native peak had almost equal height to any of the three competing conformations. However, the native peak was still the most populated one, as it was broader. The native conformation occupied the largest cluster in the first four trajectories, and the population ratios to those in the next largest clusters were 2.1, 3.8, 1.1, and 1.9, respectively (the last trajectory never sampled the native conformation). As shown in Fig. S10, the energetic difference between the native and RR and RL conformations were relatively small (in comparison with the α_3D or α_3W cases), and hence the native did not appear to be as predominant.

Since this was the largest protein, we expect that longer simulations were needed to discern more accurately the small free-

energy difference between these conformations, and to compute more precisely the relative population ratios.

RMSD of the Cluster Center vs. Average RMSD of a Cluster. We observed that the RMSD of the central structure of the native cluster was considerably smaller than the average RMSD of the cluster. By the native cluster we mean all frames with RMSD < 6.0 Å, which include frames in both the first peak of the RMSD distribution and possibly a nearby intermediate one (cf. the α_3D or α_3W case). This RMSD-based definition was roughly equivalent to the output of our clustering method, which only recognizes large clusters.

For example, in the α_3D case, the native cluster had two peaks, one around 2.8 Å, the other 4.8 Å. Thus, the average RMSD should be somewhere in between. However, the RMSDs of cluster centers were 2.5 Å ~ 2.7 Å for 6 out of the 7 trajectories (Table S3).

In the α_3W case, the native cluster also had two peaks, one around 2.3 Å, the other around 4.6 Å with a larger population. Thus, the average RMSD should be between the two figures and closer to the latter 4.6 Å. But the RMSDs of the cluster centers were 1.7 Å ~ 2.2 Å in the first trajectories, again below the first figure.

In the S-836 case, the broad native peak spanned from 4 Å to 6 Å but the RMSDs of the cluster centers were 2.5 Å ~ 4.3 Å for the four trajectories that reached the native state.

We propose an explanation by a model illustrated in Fig. S16*A*. Structures in the native cluster can be characterized by fluctuations around the minimal free-energy configuration $\mathbf{x}_0$, which is roughly the experimental structure. Thus, in the configuration space, they are distributed over a volume roughly enclosed by an n-dimensional (n is the number of degrees of freedom) hypersphere surrounding $\mathbf{x}_0$, with the radius being a multiple of the average RMSD of cluster. On the other hand, the cluster center should be identified as some frame near $\mathbf{x}_0$ to minimize the average distance from all other frames (see *Clustering Method*). Thus its RMSD would be considerably less than the average RMSD.

The model predicts that the RMSD from the native will positively correlate with the average distance from other frames in the native cluster. As shown in Figs. S16*B–E*, this was indeed the case for the four studied proteins. A frame of a low RMSD usually also had a low average distance from other frames, and vice versa.

Back to the two-peak distributions observed in α_3D and α_3W, the model suggests that the second peak was an aggregate of conformations surrounding the native structure with similar RMSDs but different fluctuations. Thus, even if the second peak had a higher population, it did not affect the location of the native. The model is useful in cluster analyses because it means that we can include structures in the second peak into the native cluster, but still identify the native one with precision.

Note the model does not require that structures in the second peak to be completely random, but only its systematic deviation from the native to be small compared with its random deviation.

Side-Chain Flexibility. We investigated the side-chain flexibility of the folded proteins. We focused on the side-chain dihedrals χ_1 (C-C_α-C_β-C_γ). We assigned a number q from 1 to 3 for the three rotamer conformations: 1 for $|\chi_1| > 120°$ (where the dihedral C-C_α-C_β-C_γ is in its *trans* conformation), 2 for $-120° < \chi_1 < 0°$, and 3 for $0° < \chi_1 < 120°$ (where the other dihedral N-C_α-C_β-C_γ is in its *trans* conformation).

The collective χ_1 conformation of a specific residue group, referred to as a rotamer combination (RC), is specified by a corresponding set of rotameric numbers, $\mathbf{q} = \{q_k\}$. We now define its rotameric entropy S as

$$S(\mathbf{q})/k_B = -\sum_{\mathbf{q}} p(\mathbf{q}) \log p(\mathbf{q}) \qquad \text{[S7]}$$

where the sum is carried over all possible RCs or $\mathbf{q}$s, with $p(\mathbf{q})$ being the observed occurring frequency of a particular $\mathbf{q}$, and k_B is the Boltzmann constant.

The effective number of RCs can be computed from S as $N = e^{S/k_B}$. If all protein rotamers were frozen, $S = 0$ and $N = 1$. Generally, $S > 0$ and $N > 1$ due to the thermal fluctuation. Since the entropy is extensive, we further define $\bar{s} = S(\mathbf{q})/N_{\text{res.}}$ and $\bar{n} = e^{\bar{s}/k_B}$ for the respective quantities per residue, where $N_{\text{res.}}$ is the number of residues in the group. Clearly, $1 \le \bar{n} \le 3$ as each χ_1 has only three rotamers.

The collective measures, S, N, $\bar{s}$, and $\bar{n}$, were computed for two residue groups, (*i*) residues involved in the hydrophobic cores, and (*ii*) peripheral hydrophilic ones. For each protein, we first collected data from both the 300 K-reweighted folding trajectories and the native regular MD, then limited ourselves to the folded state by only including frames with RMSD < 6 Å.

As shown in Table S5, even residues involved in the hydrophobic core usually had hundreds to thousands of RCs. However, compared with hydrophilic residues, the figures were smaller, suggesting certain mutual restriction and collectivity. It also shows that Fap1-NR_α had fewer RCs in comparison with the other three proteins, suggesting that the natural protein might have a tighter hydrophobic core than the designed ones.

In Fig. S5, we show that the dihedral conformations for individual residue of five proteins. For each residue, the height of a color bar represents the frequency of the corresponding χ_1. Note the frequency here is averaged over all other residues, and thus does not take collectivity into account. Nevertheless, we show that a residue could on average adopt two or three rotamers, agreeing with the results in Table S5. However, bulky residues (e.g., Phe and Trp) were often able to maintain a single χ_1 rotamer, so were Asp and Asn.

The Packing-Chirality *D* Distributions. The packing-chirality distributions of all simulation trajectories of α_3D, α_3W, LALA, and Fap1-NR_α are shown in Fig. S17. The chirality distributions of other proteins are shown in Figs. 3, 4, and 5 in the main text, and in Fig. S8. The distribution computed from the trajectory 4 of α_3D showed considerable difference from the other trajectories. This was related to the fact that this trajectory did not reach the native state with the same accuracy as the others (as shown in Table S3, the minimal RMSD 2.85 Å was much higher than those from the other trajectories). The outlying distribution from trajectory 5 of α_3W was due to a similar reason. We did not sample the left-handed conformation of LALA, and thus there was only the right-handed peak. For Fap1-NR_α, although there was not a sharp peak corresponding to a distinct folded state as in the other cases, the distribution was tilted to the left-handed side, where the native conformation belonged.

1. Zhang C, Ma J (2010) Enhanced sampling and applications in protein folding in explicit solvent. *J Chem Phys* 132:244101.
2. Zhou Y, Karplus M (1999) Folding of a model three-helix bundle protein: A thermodynamic and kinetic analysis. *J Mol Biol* 293:917–951.
3. Zhu Z, Tuckerman ME, Samuelson SO, Martyna GJ (2002) Using novel variable transformations to enhance conformational sampling in molecular dynamics. *Phys Rev Lett* 88:100201.
4. Jang S, Shin S, Pak Y (2003) Replica-exchange method using the generalized effective potential. *Phys Rev Lett* 91:058305.
5. Kannan S, Zacharias M (2007) Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. *Proteins* 66:697–706.
6. Kannan S, Zacharias M (2009) Folding simulations of Trp-cage mini protein in explicit solvent using biasing potential replica-exchange molecular dynamics simulations. *Proteins* 76:448–460.
7. Ferrenberg AM, Swendsen RH (1988) New Monte-Carlo technique for studying phase-transitions. *Phys Rev Lett* 61:2635–2638.
8. Ferrenberg AM & Swendsen RH (1989) Optimized Monte-Carlo data-analysis. *Phys Rev Lett* 63:1195–1198.

9. Chodera JD, Swope WC, Pitera JW, Seok C, Dill KA (2007) Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J Chem Theory Comput* 3:26–41.
10. Walsh ST, Cheng H, Bryson JW, Roder H, DeGrado WF (1999) Solution structure and dynamics of a de novo designed three-helix bundle protein. *Proc Natl Acad Sci USA* 96:5486–5491.
11. Swendsen RH, Wang JS (1986) Replica Monte-Carlo simulation of spin-glasses. *Phys Rev Lett* 57:2607–2609.
12. Geyer CJ (1991) *Proceedings of the 23rd Symposium on the Interface* (American Statistical Association, New York).
13. Hukushima K, Nemoto K (1996) Exchange Monte Carlo method and application to spin glass simulations. *J Physical Soc Japan* 65:1604–1608.
14. Hansmann UHE (1997) Parallel tempering algorithm for conformational studies of biological molecules. *Chem Phys Lett* 281:140–150.
15. Zhang C, Ma JP (2007) Simulation via direct computation of partition functions. *Phys Rev E* 76:036708.
16. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334:517–520.
17. Matouschek A, Kellis JT, Serrano L, Fersht AR (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340:122–126.

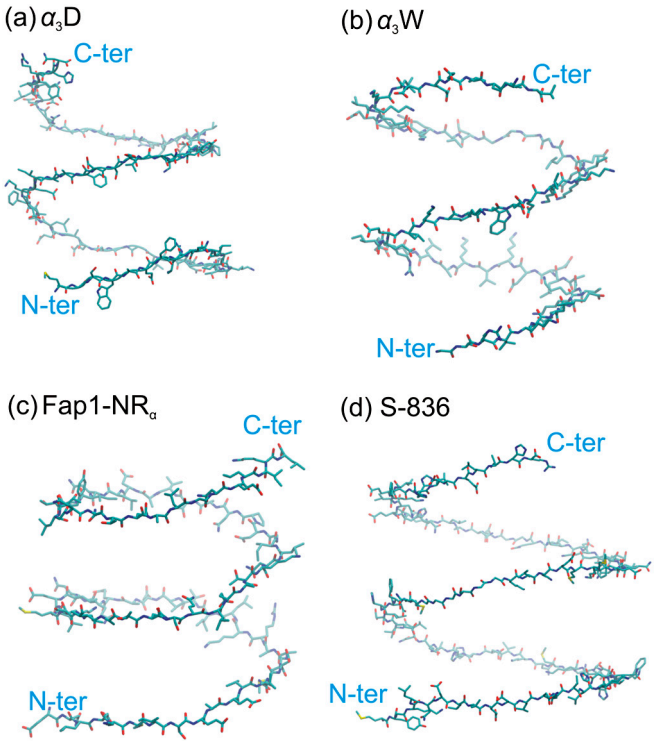

Fig. S1. Initial extended conformations of (*A*) α_3D, (*B*) α_3W, (*C*) Fap1-NR$_\alpha$, and (*D*) S-836. Water molecules are removed for clarity.

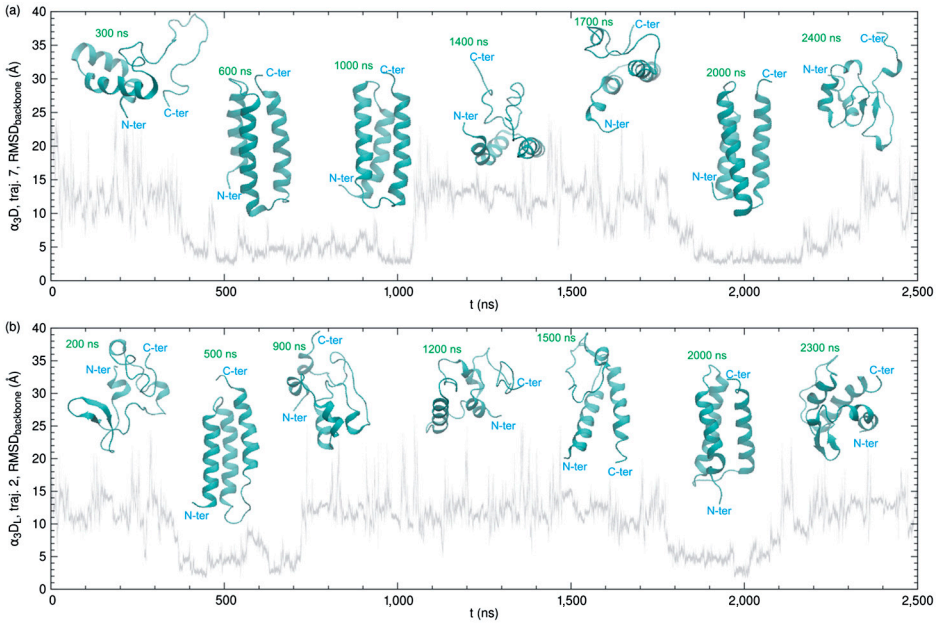

Fig. S2. Representative structures (reweighted to 300 K) along (*A*) trajectory 7 of α_3D and (*B*) the trajectory 2 of the mutant α_3D$_L$.

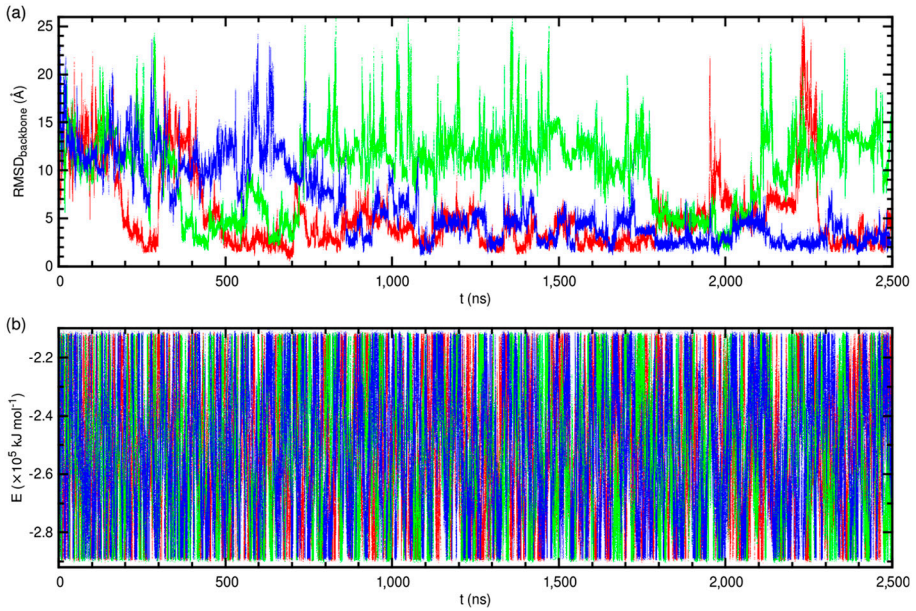

Fig. S3. Three trajectories of α_3D_L. (*A*) Root-mean-square deviation from the putative native structure and (*B*) instantaneous potential energy versus time *t*. Red, green, and blue are for trajectories 1, 2, and 3, respectively.

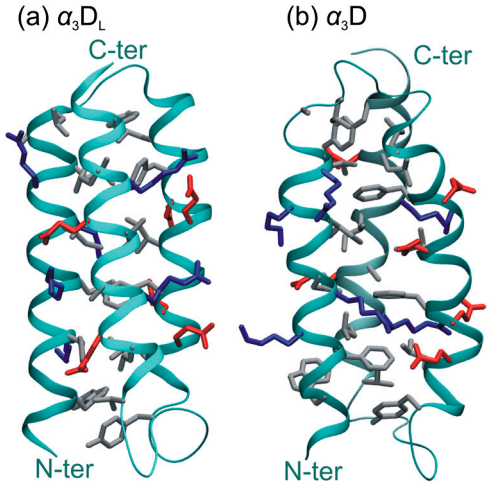

Fig. S4. The putative native structure of α_3D_L (*A*) versus the native one of α_3D (*B*). Hydrophobic, positively charged, and negatively charged residues are shown in gray, blue, and red, respectively.

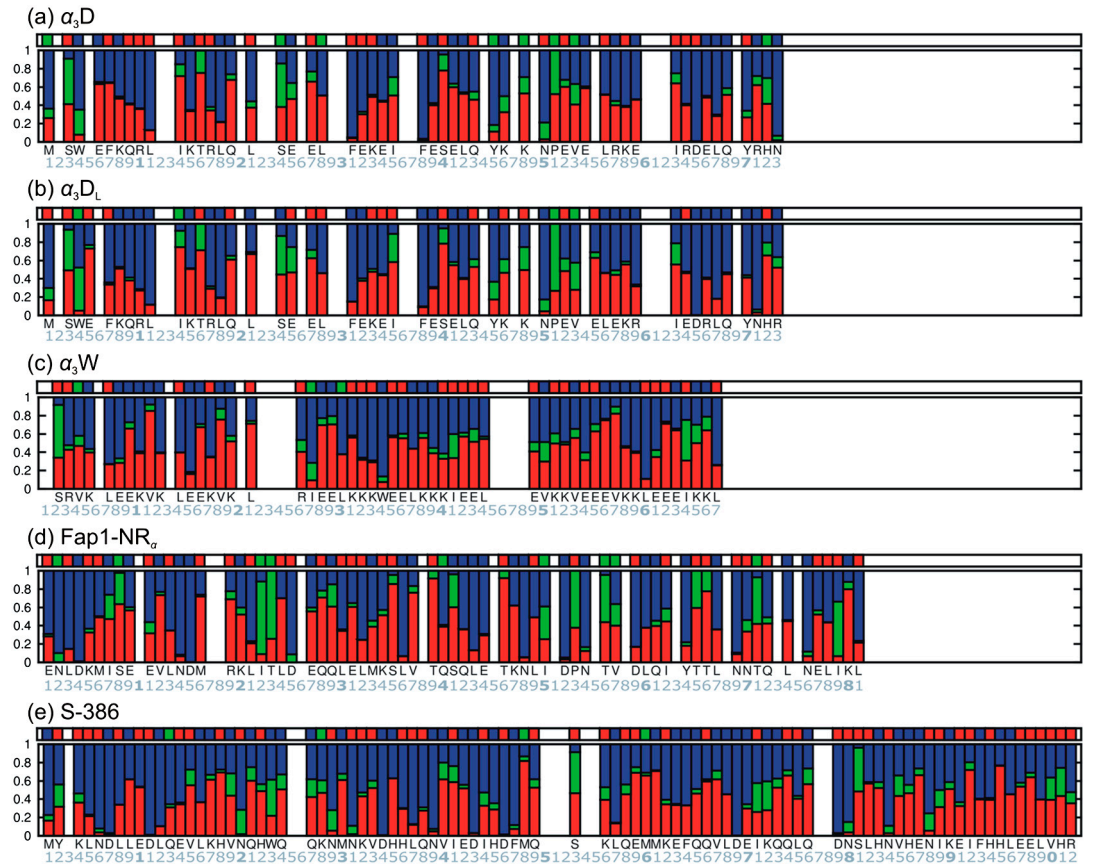

Fig. 5. The χ_1 rotamer profiles (computed from folding trajectories, reweighted to 300 K) of (*A*) α_3D, (*B*) α_3D$_L$, (*C*) α_3W, (*D*) Fap1-NR$_\alpha$, and (*E*) S-836. Gly and Ala residues are excluded for lack of C_γ atoms. Colors of the three rotamers, red: $|\chi_1| > 120°$, green: $-120° < \chi_1 < 0°$, and blue: $0° < \chi_1 < 120°$. In each panel, the top bar represents the χ_1 rotamer of each residue in the experimentally determined conformation.

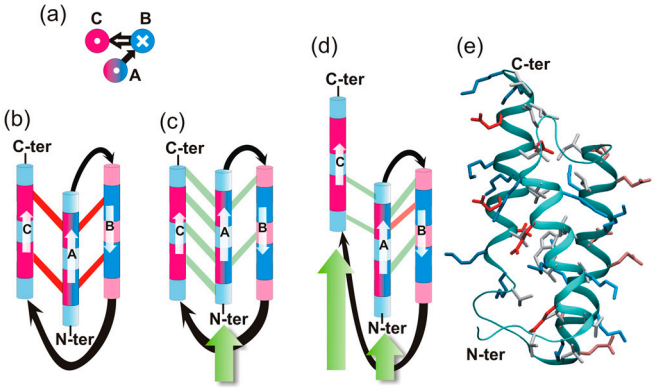

Fig. S6. Right-handed conformation of α_3W. (*A*) Top view of the three helices. (*B*) and (*C*) Schematic plots of the unadjusted and best adjusted conformations, respectively. In helices B or C, the *b* and *c* residues and *e*, *f*, and *g* residues carry opposite charges (the lowercase letters a–g are from the helix heptad). The unadjusted conformation involves strong electrostatic repulsion among *e*, *g* residues, as shown in red lines in *B*. However, if helix A slides along the helical axis by one helix turn (along the green arrow in panel *B*), such that the *e*, *f*, and *g* residues on helix A are roughly on the same level of the b and c residues on helices C and B (instead of opposing their counterparts with the same charges), the direct electrostatic repulsion can be replaced by a weak attraction, as illustrated by green lines in *C*. Similar arrangements involving more helices are also possible. (*D*) and (*E*) Schematic and ribbon representations of the a conformation observed from a trajectory.

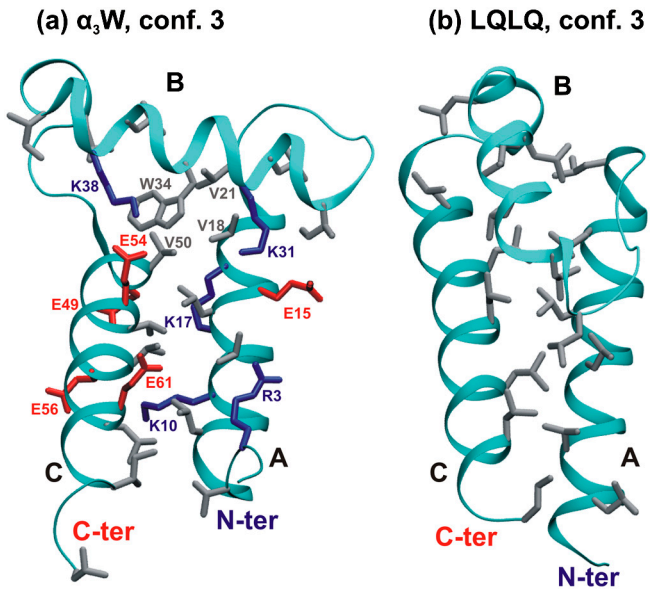

Fig. S7. (*A*) Conformation 3 of α_3W. Helices A and C are closely aligned in antiparallel through both hydrophobic residues and charged residues (Arg3-Glu61, Lys10-Glu56, Lys17-Glu49), leading to a closer distance 8.4 Å. Additionally, Trp34 on helix B is packed against the hydrophobic residues on helices A (Val18 and Leu21) and C (Val50). Helix B is further secured by two additional salt bridges: Lys31-Glu15 for helices B and A, and Lys38-Glu54 for helices B and C. (*B*) Conformation 3 of LQLQ. It breaks an interesting symmetry: helix B is aligned antiparallel (and more closely) to helix C instead of helix A. As both helices A and C are slightly bent according to the coiled-coil topology, the curvature makes helix C to tilt away from the hydrophobic face around the loop region between helices B and C. Thus stretching the B–C loop is more difficult than stretching the A–B loop, and hence a closer association of helices B and C. The conformation resembles a "helix hairpin," e.g., PDB ID: 2K1E (1).

1 Ma D, et al. (2008) NMR studies of a channel protein without membranes: Structure and dynamics of water-solubilized KcsA. *Proc Natl Acad Sci USA* 105:16537–16542.

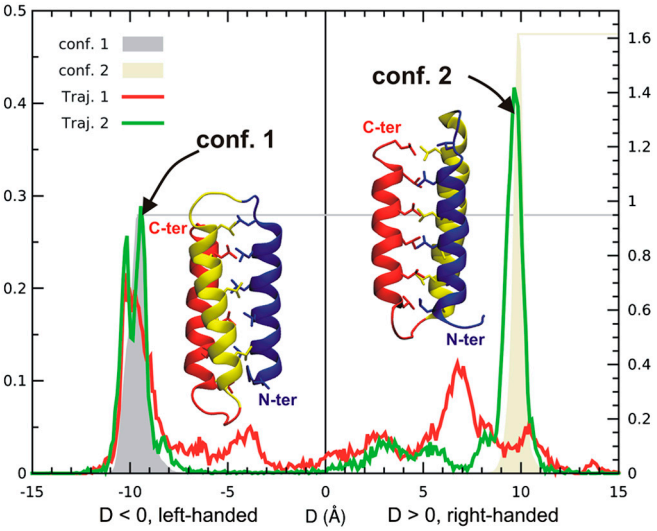

Fig. S8. The packing-chirality *D* distributions (300 K) of LALQ. Two conformations at cluster centers from cluster analyses are also shown. The colors of the three successive helices are blue, yellow, and red starting from the N to C terminus. The shaded peaks (to be read from the right vertical scale) in both panels were obtained from 300 K regular MD simulations starting from the respective cluster-center structures.

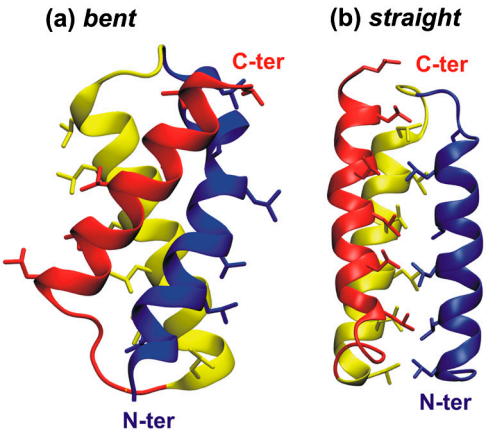

Fig. S9. Two competing conformations, bent and straight, of LALA (extracted from the two largest clusters). Leu side chains are also shown. The colors of the three successive helices A, B, and C from the N to C terminus are blue, yellow, and red, respectively. (*A*) In the bent conformation, helix A is kinked with shortened distances to helices B (7.1 Å) and C (6.8 Å). The distance between helices B and C is 9.1 Å. (*B*) In the straight conformation, the distance between B and C (6.7 Å) is shorter, while the other two distances are longer: 9.1 Å (A–B) and 9.8 Å (B–C).

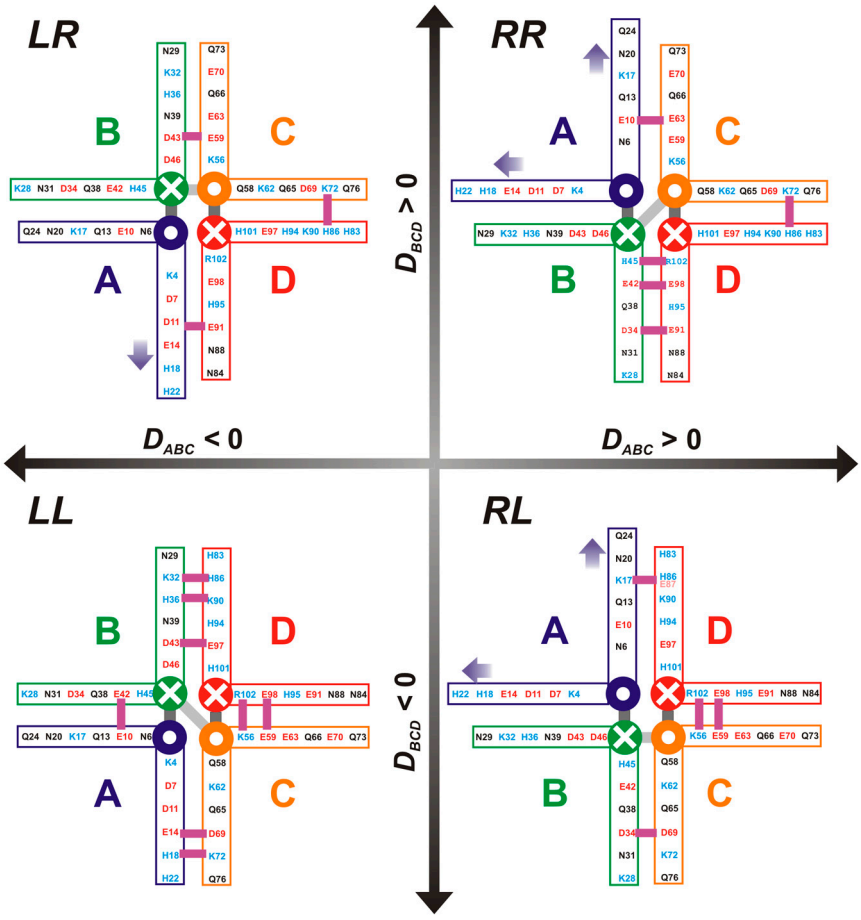

Fig. S10. The packing patterns of charged residues in the four possible chiral conformations of the four-helix bundle S-836. The chiral conformations are casted to four quadrants LR, RR, RL, and LL according to the signs of the two chiral distances D_{ABC} and D_{BCD}. Only the LR, RR, and RL conformations were observed from folding trajectories. Each helix of the protein adopts the heptad repeat *a-b-c-d-e-f-g*. Residues on *a*, *d*, and sometimes *e* positions are hydrophobic. The rest (charged/polar) on the heptad are engaged in interhelical pairing, and can be casted into two groups: *b* and *e* (or *f* if *e* is hydrophobic) residues on one side, and *c* and *g* residues the other. The two sides interact with different helices and are shown as the two "arms" in the helix wheel representation (outer residues are closer to the reader). Interactions of charged/polar residues between two different helices are shown as two aligned arms with residues on the interface. Unfavorable interactions between corresponding residues with same charges are shown by magenta bars. In the observed conformations of LR, RR, and RL, a helix often slides slightly along its axis, as indicated by blue arrows, to optimize interhelical interactions among charged residues.

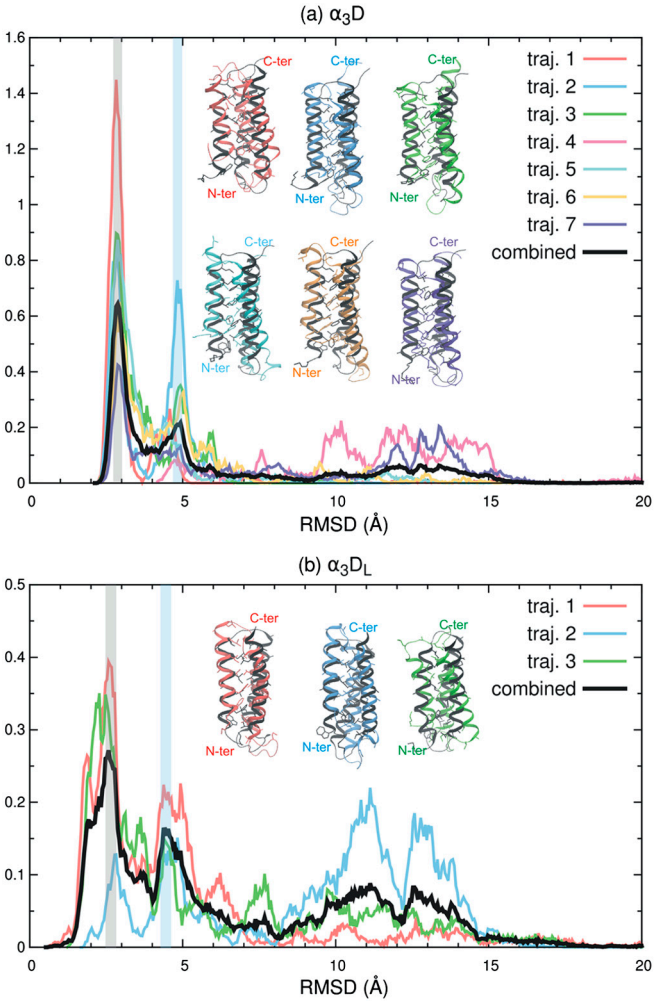

Fig. S11. The 300 K distributions of the RMSD from the (putative) native structures for (*A*) α_3D and (*B*) its mutant $\alpha_3 D_L$. The representative structure (black) at the smallest RMSD peak (gray strip) and that (colored) at the second smallest RMSD peak (light-blue strip) from each trajectory are aligned. Trajectory 4 of α_3D did not have the first peak and thus is omitted. The colored strips indicate where the representative structures were extracted, but do not represent physical clusters (same for Figs. S12 to S15).

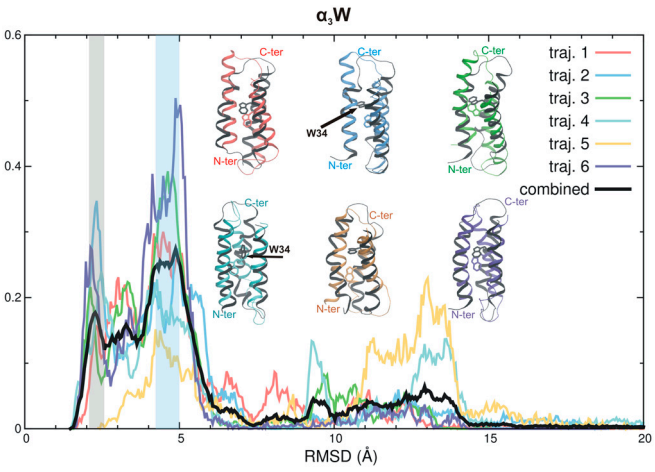

Fig. S12. The 300 K distributions of the RMSD from the native structure for α_3W. The representative structure (black) at the smallest RMSD peak (gray strip) and that (colored) at the second smallest RMSD peak (light-blue strip) from each trajectory are aligned. Trp34s in the second peaks were consistently lower than those in the first peaks.

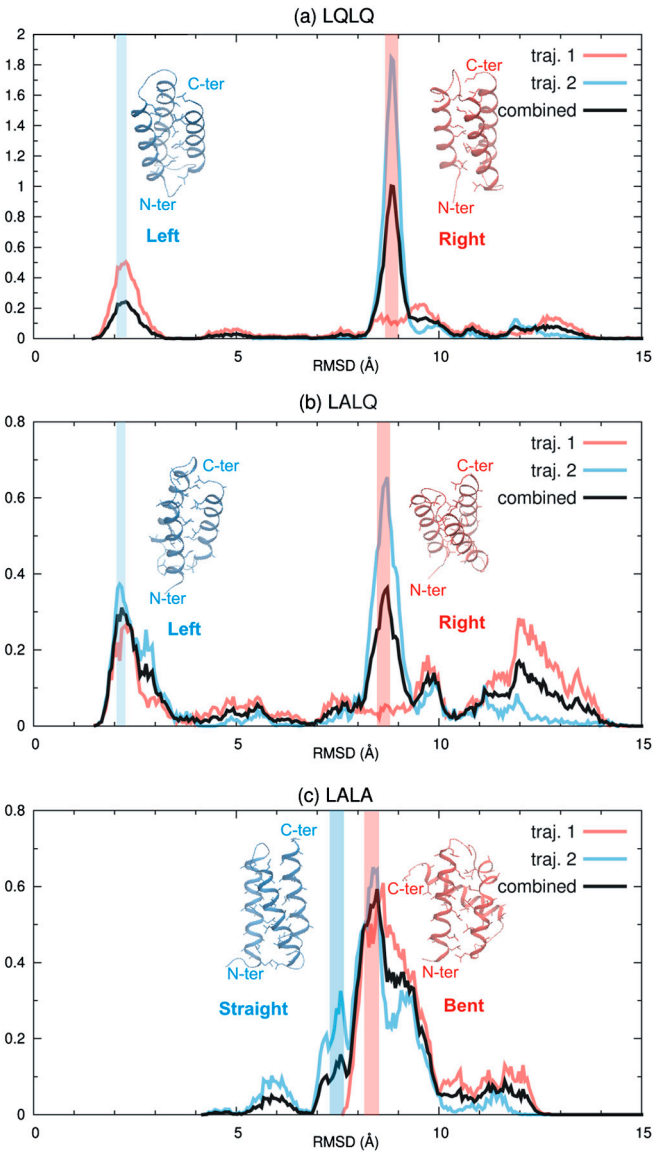

Fig. S13. The 300 K distributions of the RMSD from the putative native structures for (*A*) LQLQ, (*B*) LALQ, and (*C*) LALA. The representative structures at the distribution peaks are also shown. The RMSD ranges for selecting the representative structures are as follows: LQLQ left, (2.1, 2.4) Å; right, (8.7, 9.0) Å. LALQ left, (2.1, 2.3) Å; right: (8.5, 8.7) Å. LALA left (straight), (7.4, 7.7) Å; right (bent): (8.3, 8.6) Å (cf. Fig. S9).

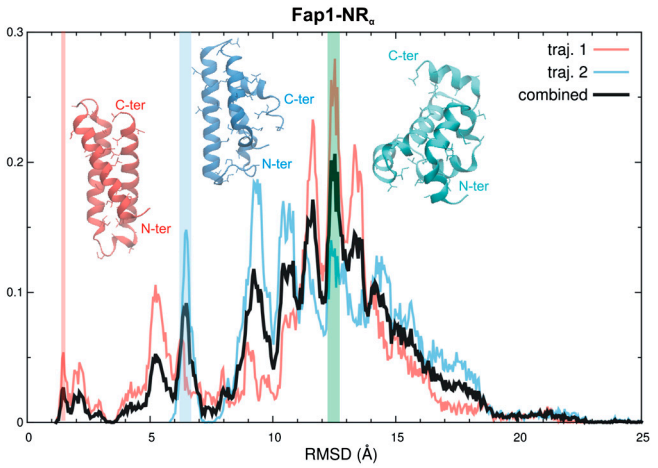

Fig. S14. The 300 K distributions of the RMSD from the native structures of Fap1-NR$_\alpha$. Three representative structures for RMSD within the ranges (1.4 Å, 1.6 Å) (red), (6.3 Å, 6.5 Å) (blue), and (12.3 Å, 12.6 Å) (cyan) are also shown.

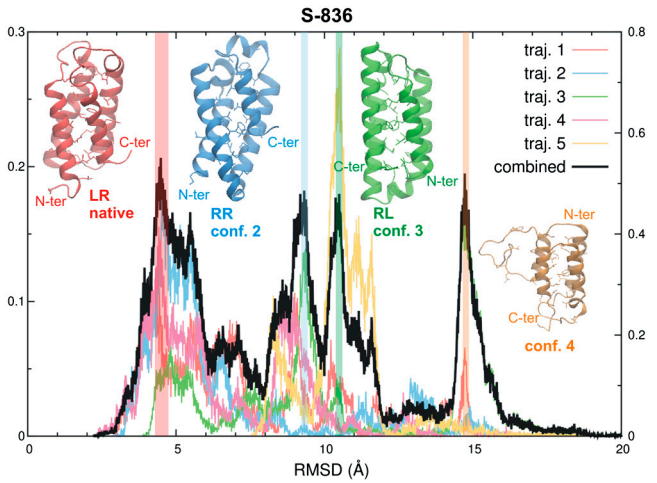

Fig. S15. The 300 K distributions of the RMSD from the native structure of S-836. The left and right vertical scales are for the combined trajectory and individual ones, respectively. Representative structures for RMSD within the ranges (4.3 Å, 4.7 Å) (native, red), (9.2 Å, 9.4 Å) (RR, blue), (10.4 Å, 10.6 Å) (RL, green), and (14.6 Å, 14.8 Å) (conf. 4 as in Fig. 5 in the text) are also shown.

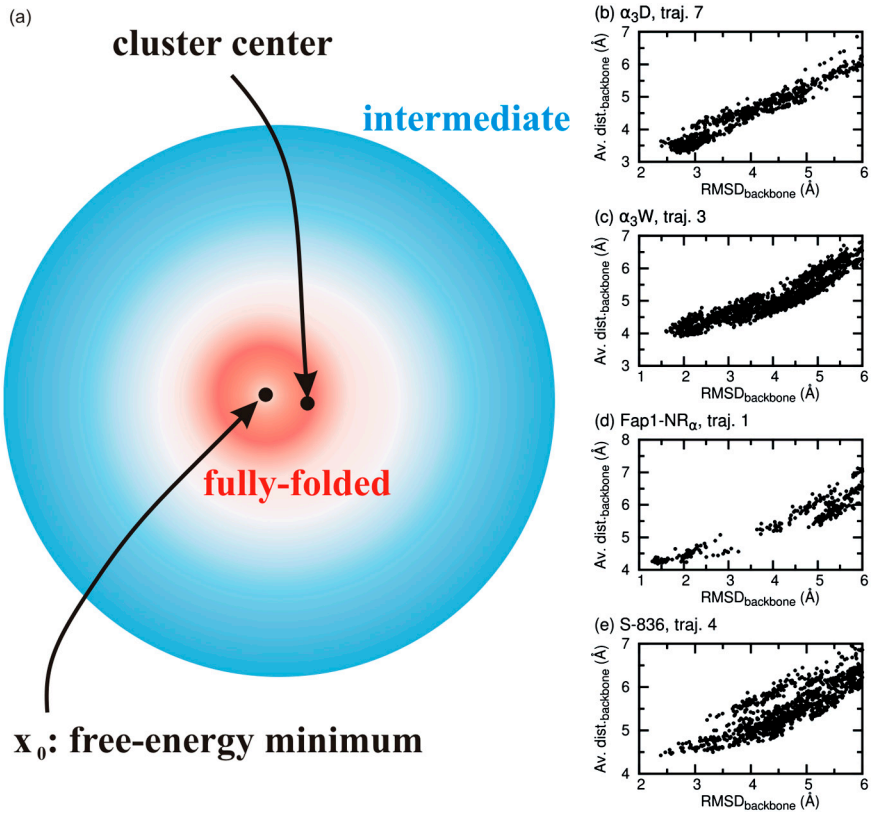

Fig. S16. (*A*) A model for the relation between the fully folded and intermediate states in the configuration space. *(B)–(E)* The average distance from other frames in the native cluster (frames with RMSD < 6.0 Å) versus the RMSD from the native conformation for α_3D, α_3W, Fap1-NR$_\alpha$, and S-836, respectively.

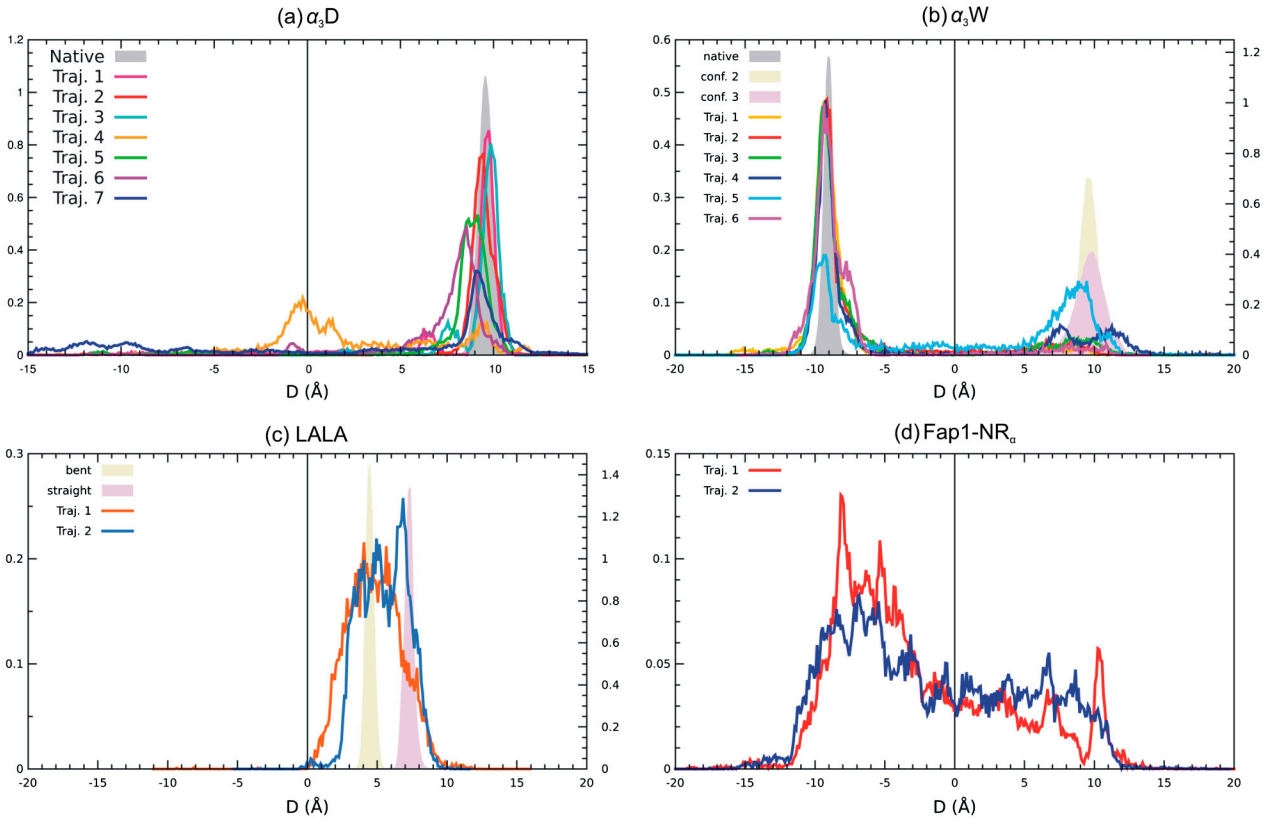

Fig. S17. The packing-chirality D distributions of all trajectories of (*A*) α_3D, (*B*) α_3W, (*C*) LALA, and (*D*) Fap1-NR$_\alpha$. $D > 0$: right-handed, $D < 0$: left-handed. Since trajectory 4 of α_3D and trajectory 5 of α_3W reached the respective folded states with lower accuracy (cf. Table S3), their distributions differed considerably from the rest distributions.

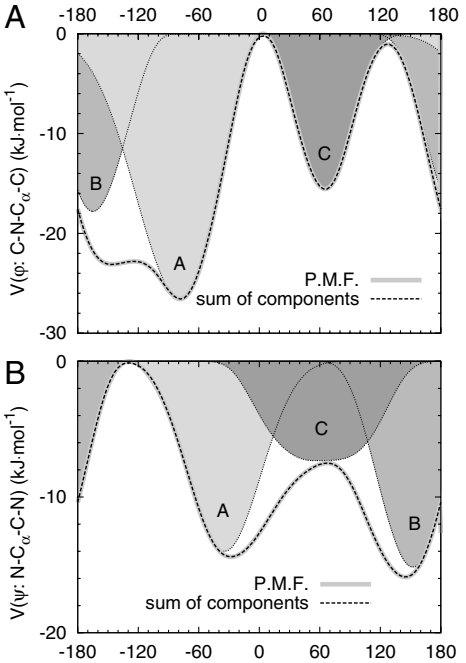

Fig. S18. The potential of mean force (computed from a special simulation of α_3D) and three decomposed dihedral modes for φ and ψ in panels (*A*) and (*B*), respectively. Dashed line: sum of modes; light solid line: the original potential of mean force. The maximum of the potential is shift to zero.

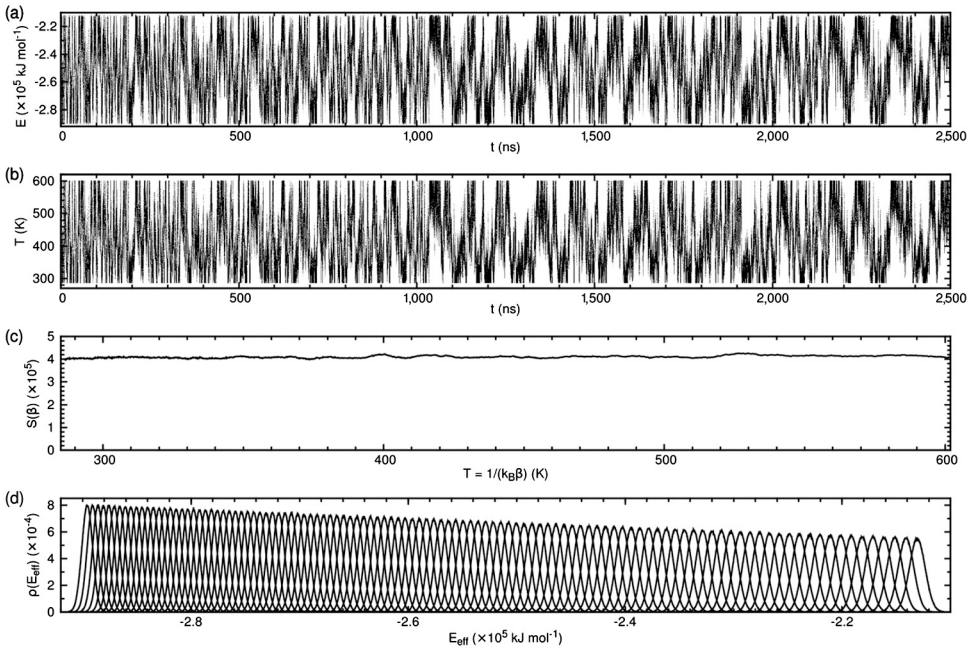

Fig. S19. Folding trajectory 7 of α_3D. (*A*) and (*B*) The instantaneous potential energy and temperature versus simulation time *t*. (*C*) The temperature histogram $H(\beta)$ divided by the intended temperature density $w(\beta)$, a quantity expected to be constant after convergence. (*D*) 111 energy histograms at different temperatures with an interval $\Delta\beta = 0.002$ from $\beta = 0.20$ to 0.42 (cf. the temperature bin size is $\delta\beta = 0.0001$).

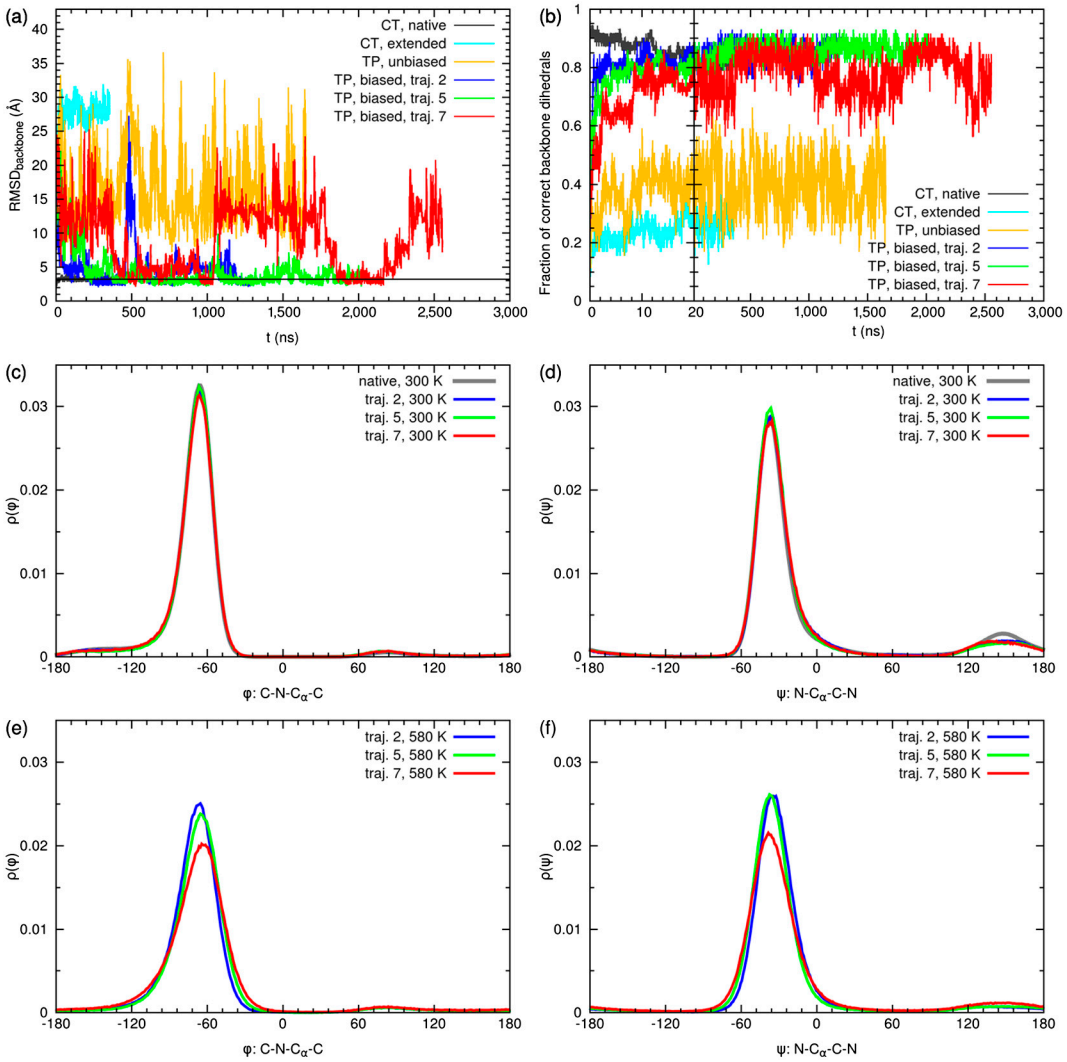

Fig. S20. For α_3D: (*A*) backbone RMSD from the native structure versus time for a regular constant-temperature molecular dynamics (CTMD) simulation starting from a fully extended chain, CTMD from the experimentally determined structure, unbiased (TP) and dihedral-biased (this method, trajectories 2, 5, and 7) tempering from an extended chain. (*B*) The fraction of correct backbone dihedrals versus time. (*C*) and (*D*) Residue-averaged backbone dihedrals distributions of φ (C-N-C_α-C) and ψ (N-C_α-C-N), respectively, at 300 K via the multiple histogram method computed from native-like conformations. (*E*) and (*F*) The corresponding distributions at 580 K. Note in producing panels *C–F*, we only included the folded state (frames with RMSD < 6 Å) to avoid the disturbance from the conformational diversity; thus, they are only used to study local dihedral equilibration and not for representing the free-energy landscape.

Table S1. The amino acid sequences of α_3D, α_3W, Fap1-NR$_\alpha$, and S-836 and the mutants

Prot.	Sequence
α_3D	MGSWAEFKQRLAAIKTRLQALGGSEAELAAFEKEIAAFESELQAYKGKGNPEVEALRKEAAAIRDELQAYRHN
α_3D_L	MGSWEAFKQRLAAIKTRLQALGGSEAELAAFEKEIAAFESELQAYKGKGNPEVAELEKRAAAIEDRLQAYNHR
α_3W	GSRVKALEEKVKALEEKVKALGGGGRIEELKKKWEELKKKIEELGGGGEVKKVEEEVKKLEEEIKKL
LQLQ	GSGLQQLQQQLQQLQQQLQQLGGGGGLQQLQQQLQQLQQQLQQLGGGGGLQQLQQQLQQLQQQLQQL
LALQ	GSGLAALQQQLAALQQQLAALGGGGGLAALQQQLAALQQQLAALGGGGGLAALQQQLAALQQQLAAL
LALA	GSGLAALAAALAALAAALAALGGGGGLAALAAALAALAAALAALGGGGGLAALAAALAALAAALAAL
Fap1-NR$_\alpha$	ENLDKMISEAEVLNDMAARKLITLDAEQQLELMKSLVATQSQLEATKNLIGDPNATVADLQIAYTTLGNNTQALGNELIKL
S-836	MYGKLNDLLEDLQEVLKHVNQHWQGGQKNMNKVDHHLQNVIEDIHDFMQGGGSGGKLQEMMKEFQQVLDEIK QQLQGGDNSLHNVHENIKEIFHHLEELVHR

Table S2. Simulation conditions for α_3D, α_3W, Fap1-NR$_\alpha$, and S-836 and the mutants

Prot.	PDB ID	$N_{res.}$	charge*	box (Å) †	N_{water}	$RMSD_{bb}^{min}$ (Å)‡	$RMSD_{bb}^{300\ K}$ (Å)§	$t^{300\ K}$ (ns)¶
α_3D	2A3D	73	−1	68	6989	1.9	3.2 ± 0.2	410
α_3D_L		73	−1	68	6971	*n/a*	*n/a*	423
α_3W	1LQ7	67	+2	65	6055	1.4	2.2 ± 0.3	436
LQLQ		67	0	65	6096	*n/a*	*n/a*	*n/a*
LALQ		67	0	68	7075	*n/a*	*n/a*	*n/a*
LALA		67	0	68	7120	*n/a*	*n/a*	*n/a*
Fap1-NR$_\alpha$	2KUB	81	−6	70	7515	1.0	1.5 ± 0.2	500
S-836	2JUA	102	+2	80	11440	2.1	3.1 ± 0.4	400

*MARKER1 The total charge of the protein; sodium (Na^+) and chloride (Cl^-) ions were added to neutralize the respective system, e.g., two chloride ions were added to α_3W.

†MARKER2 The box size represented by the side length of the bottom square of the dodecahedron.

‡The minimal root-mean-square deviation (RMSDs) of backbone atoms from the experimentally determined structure.

§The average backbone RMSD in the control simulation performed at 300K using regular MD.

¶The corresponding simulation time. n/a: RMSDs are not applicable to designed proteins.

Table S3. Trajectory parameters of α_3D, α_3W, Fap1-NR$_\alpha$, and S-836 and the mutants

Prot.	ID	t (μs) *	φ : p_A, p_B, p_C, p_{min}	ψ : p_A, p_B, p_C, p_{min}	r_s, r_v, r_c, Δx_{PMF} (Å) †	$RMSD_{min}/RMSD_{cls.}$ (Å) ‡
α_3D	0	1.7	1.0, 1.0, 1.0, 1.0	1.0, 1.0, 1.0, 1.0	7, 8, 10, 1.31	6.27/na
	1	1.0	1.0, 0.1, 0.3, 3.0	9.0, 0.4, 1.0, 3.0	7, 8, 10, 1.31	$1.94/2.5_1$
	2	1.3	1.0, 0.2, 0.3, 3.0	9.0, 0.4, 1.0, 3.0	7, 8, 10, 1.31	$1.87/2.6_1$
	3	1.2	1.0, 0.2, 0.5, 3.0	9.0, 0.5, 1.0, 3.0	7, 8, 10, 1.31	$1.90/2.5_1$
	4	1.3	0.15, 0.05, 0.1, 3.0	1.5, 0.5, 0.5, 3.0	7, 8, 10, 1.31	$2.85/4.6_{2,99\%}$
	5	2.0	0.25, 0.05, 0.1, 1.0	3.0, 0.2, 0.2, 1.0	8, 9, 11, 1.42	$1.97/2.7_1$
	6	1.0	1.0, 0.3, 0.3, 1.0	1.0, 0.3, 0.3, 1.0	8, 9, 11, 1.42	$2.12/2.7_1$
	7	2.5	0.1, 0.05,0.1, 1.0	1.0, 0.3, 0.3, 1.0	8, 9, 11, 1.31	$2.03/2.6_1$
α_3D_L	1	2.5	1.0, 0.3, 0.3, 1.0	1.0, 0.3, 0.3, 1.0	8, 9, 11, 1.31	na
	2	2.5	1.0, 0.5, 1.0, 1.0	1.0, 0.3, 0.5, 1.0	8, 9, 11, 1.31	na
	3	2.5	0.5, 0.1, 0.3, 1.0	2.0, 0.3, 0.5, 1.0	8, 9, 11, 1.31	na
α_3W	0	1.0	1.0, 1.0, 1.0, 1.0	1.0, 1.0, 1.0, 1.0	7, 8, 10, 1.33	7.18/na
	1	1.6	0.15, 0.05, 0.1, 1.0	1.5, 0.3, 0.3, 1.0	7, 8, 10, 1.33	$1.47/2.2_1$
	2	1.8	0.25, 0.05, 0.1, 1.0	2.0, 0.2, 0.2, 1.0	7, 8, 10, 1.33	$1.39/1.7_1$
	3	2.5	0.5, 0.1, 0.3, 1.0	5.0, 0.3, 0.5, 1.0	7, 8, 10, 1.33	$1.38/2.2_1$
	4	1.8	0.2, 0.05, 0.1, 1.0	2.0, 0.3, 0.3, 1.0	7, 8, 10, 1.33	$1.40/2.4_1$
	5	1.0	1.0, 0.1, 0.3, 3.0	9.0, 0.4, 1.0, 3.0	7, 8, 10, 1.33	$1.83/3.6_{2,93\%}$
	6	1.0	1.0, 0.1, 0.3, 3.0	10, 0.5, 0.3, 3.0	7, 8, 10, 1.33	$1.36/3.2_1$
LQLQ	1	1.7	1.0, 0.3, 0.3, 1.0	1.0, 0.3, 0.3, 1.0	8, 9, 11, 1.33	na
	2	1.5	0.5, 0.1, 0.3, 1.0	5.0, 0.3, 0.5, 1.0	8, 9, 11, 1.33	na
LALQ	1	1.0	1.0, 0.3, 0.3, 1.0	1.0, 0.3, 0.3, 1.0	8, 9, 11, 1.31	na
	2	1.0	0.5, 0.1, 0.3, 1.0	5.0, 0.3, 0.5, 1.0	8, 9, 11, 1.31	na
LALA	1	1.2	1.0, 0.3, 0.3, 1.0	1.0, 0.3, 0.3, 1.0	8, 9, 11, 1.31	na
	2	1.2	0.5, 0.1, 0.3, 1.0	2.0, 0.3, 0.5, 1.0	8, 9, 11, 1.31	na
Fap1-NR$_\alpha$	1	2.4	0.25, 0.05, 0.1, 1.0	3.0, 0.3, 0.3, 0.0	8, 9, 11, 1.46	$1.00/1.4_{2,45\%}$
	2	2.5	0.2, 0.05, 0.1, 1.0	2.0, 0.25, 0.5, 1.0	8, 9, 11, 1.46	3.11/na
S-836	1	3.1	0.2, 0.05, 0.1, 1.0	2.0, 0.3, 0.3, 1.0	8, 10, 12, 1.45	$2.09/3.8_1$
	2	3.3	0.2, 0.05, 0.1, 1.0	2.0, 0.4, 0.0, 1.0	8, 10, 12, 1.45	$2.21/3.8_1$
	3	4.2	0.25, 0.05, 0.1, 1.0	2.5, 0.3, 0.1, 1.0	7, 8, 10, 1.25	$2.30/4.3_1$
	4	3.3	0.5, 0.1, 0.3, 1.0	5.0, 0.3, 0.5, 1.0	8, 9, 11, 1.43	$2.13/2.5_1$
	5	3.1	0.25, 0.05, 0.1, 1.0	3.0, 0.15, 0.3, 1.0	8, 9, 11, 1.43	7.31/na

The temperature range $\beta \in (0.2, 0.42)$, with the bin size $\delta\beta = 0.0001$ for the first three, or 0.00005 for S-836. The thermostat temperature $T_0 = 480$ K. For the integral identity, the temperature window for integral identity was $(\beta - \Delta\beta, \beta + \Delta\beta)$ where $\Delta\beta/\beta = 0.03$ for the first three, or 0.02 for S-836. The time step for integrating Langevin equation $\Delta t_{kT} = 2 \times 10^{-5}$ for the first three, or 10^{-5} for S-836. The amplitude for adaptive averaging $C_\gamma = 1.0$

*Simulation time.

†The switching radius r_s for van der Waals interactions, cutoff radius r_v for van der Waals interactions, real-space cutoff r_c and Fourier grid spacing Δx_{PMF} for PME.

‡The minimal and cluster-center backbone root-mean-square deviations (RMSD) from the experimentally determined structure; the latter was measured from the cluster containing the native conformation (RMSD < 5 Å, na if no such cluster exists); the subscript was its rank of in size, and if not 1, then followed by the population to the largest cluster, e.g., 2.5_1 means the native cluster was the largest and its RMSD is 2.5 Å, $1.4_{2,45\%}$ means the native cluster was in the second largest, and its population ratio to the largest is 45%.

Table S4. Geometry measures of various conformations of three-helical bundles

	t (ns)	NC	D_{ABC} (Å)	d_{AB} (Å)	d_{AC} (Å)	d_{BC} (Å)	θ_{AB}(°)	θ_{AC}(°)	θ_{BC}(°)
α_3D, native	410	49 ± 2	9.6 ± 0.4	9.8 ± 0.3	9.7 ± 0.4	9.8 ± 0.6	168 ± 5	11 ± 4	167 ± 4
α_3D_L, native	423	50 ± 2	−9.6 ± 0.5	9.6 ± 0.4	9.8 ± 0.3	9.6 ± 0.5	174 ± 3	8 ± 4	168 ± 4
α_3W, native	436	49 ± 2	−9.0 ± 0.4	9.9 ± 0.3	9.6 ± 0.3	9.6 ± 0.4	163 ± 4	20 ± 5	165 ± 5
α_3W, conf. 2	240	40 ± 3	9.6 ± 0.6	9.6 ± 0.5	9.3 ± 0.5	10.0 ± 0.4	152 ± 10	20 ± 8	168 ± 5
α_3W, conf. 3	236	32 ± 2	9.7 ± 1.1	12.0 ± 0.4	8.5 ± 0.3	15.0 ± 0.4	72 ± 7	168 ± 6	108 ± 9
LQLQ, conf. 1	205	51 ± 2	−9.4 ± 0.3	9.6 ± 0.2	9.8 ± 0.3	9.4 ± 0.3	165 ± 4	14 ± 3	172 ± 4
LQLQ, conf. 2	216	51 ± 2	9.8 ± 0.4	9.2 ± 0.3	10.2 ± 0.9	9.6 ± 0.8	171 ± 4	13 ± 5	165 ± 5
LQLQ, conf. 3	197	47 ± 2	9.1 ± 0.5	10.6 ± 0.6	9.2 ± 0.4	10.2 ± 0.5	29 ± 8	160 ± 4	165 ± 6
LALQ, conf. 1	201	51 ± 2	−9.6 ± 0.5	9.5 ± 0.6	10.2 ± 0.7	9.4 ± 0.3	165 ± 4	15 ± 4	172 ± 4
LALQ, conf. 2	197	51 ± 2	9.8 ± 0.3	9.3 ± 0.3	11.1 ± 0.8	9.1 ± 0.7	171 ± 4	11 ± 3	168 ± 4
LALA, bent	205	40 ± 2	4.5 ± 0.3	7.1 ± 0.2	6.9 ± 0.2	9.4 ± 0.2	141 ± 3	44 ± 2	119 ± 3
LALA, straight	197	49 ± 2	7.3 ± 0.3	9.1 ± 0.2	9.8 ± 0.3	6.9 ± 0.2	162 ± 3	21 ± 5	168 ± 5

NC: the total number of contacting hydrophobic residues between different helices with a 12 Å cutoff; D_{ABC}: the chiral distance. d_{AB}, d_{AC} and d_{BC}: the interhelical distances between helices A-B, A-C, and B-C, respectively; θ_{AB}, θ_{AC} and θ_{BC}: the interhelical angles between helices A-B, A-C, and B-C, respectively; Residues that define helices A, B and C: for α_3D and α_3L, A: 4–21, B: 28–45, C: 53–70; α_3W, LQLQ, LALQ and LALA, A: 4–21, B: 27–44, C: 50–67. Numbers followed ± are standard deviations.

Table S5. The rotameric entropy *S* of collective χ_1-dihedral conformations for several proteins

Protein	N_{res}	S/N_{res}, S, e^{S/k_B} (native) residues	S/N_{res}, S, e^{S/k_B} (folding)
α_3D, core	17	0.35, 5.9, 3.5×10^2 W4 F7 L11 I14 L18 L21 L28 F31 I35 F38 L42 Y45 V53 L56 I63 L67 Y70	0.45, 7.6, 2.1×10^3
α_3D, exposed	12	0.56, 6.7, 7.9×10^2 K8 R10 K15 R17 E32 E34 E39 E41 R57 E59 R64 E66	0.67, 8.0, 3.0×10^3
α_3D_L, core	17	0.31, 5.3, 2.0×10^2 W4 F7 L11 I14 L18 L21 L28 F31 I35 F38 L42 Y45 V53 L56 I63 L67 Y70	0.44, 7.5, 1.8×10^3
α_3D_L, exposed	12	0.55, 6.6, 7.6×10^2 K8 R10 K15 R17 E32 E34 E39 E41 E57 R59 E64 R66	0.65, 7.8, 2.5×10^3
α_3W, core	18	0.38, 6.8, 9.1×10^2 V4 L7 V11 L14 V18 L21 I27 L30 W34 L37 I41 L44 V40 V53 V57 L60 I64 L67	0.47, 8.5, 5.0×10^3
α_3W, exposed	12	0.49, 5.9, 3.5×10^2 E8 K10 E15 K17 K31 K33 K38 K40 E54 E56 E61 E63	0.65, 7.8, 2.5×10^3
Fap1-NR$_\alpha$, core	18	0.28, 5.1, 1.6×10^2 L3 M6 I7 L13 L21 L32 L36 T39 L43 T46 L49 I50 L60 Y64 L67 T71 L74 L78	0.31, 5.6, $2.7 \times 10^{2*}$
Fap1-NR$_\alpha$, exposed	12	0.61, 7.4, 1.6×10^3 K5 E9 E11 R19 E31 K34 Q40 E44 Q61 Q72 E77 K80	0.51, 6.2, $4.8 \times 10^{2*}$
S-836, core	32	0.25, 7.9, 2.6×10^3 L5 L8 L9 L12 V15 L16 V19 W23 M30 V33 L37 V40 I41 I44 F47 M48 L57 M60 M61 F64 V67 L68 I71 L75 L82 V85 I89 I92 F93 L96 L99 L100	0.27, 8.80, 6.5×10^3
S-836, exposed	20	0.43, 8.6, 5.7×10^3 K4 E10 Q13 K17 Q21 K28 K32 Q38 E42 H45 K56 E59 K62 Q66 K72 E87 E91 H94 E98 H102	0.46, 9.2, 9.5×10^3

For each protein, we define two groups of residues: *core* for residues engaged in hydrophobic core, and exposed for exposed hydrophilic residues (which are the *e*, *g* residues on the heptad for α_3D, α_3D_L and α_3W). The former group reveals the flexibility of the hydrophobic core, while the latter group serves as a control for comparison. N_{res} is the number of residues involved in a rotameric pattern. Unit of entropy is the Boltzmann constant k_B. A larger entropy suggests more rotamer combinations. For each protein, the entropy is calculated from both the 300 K native MD trajectory, and a combination of all folding trajectories. The entropy from the latter is often larger due to a more thorough sampling. The results for Fap1-NR$_\alpha$ from folding trajectories (marked by "*") are likely underestimated due to limited folding trajectories

Table S6. The parameters of the dihedral modes from decomposing the potentials of mean force along the two backbone dihedrals φ and Ψ

	M(kJ mol^{-1})	φ_c or ψ_c(rad)	a_2^-	a_4^-	a_6^-	a_2^+	a_4^+	a_6^+
φ, **A**	53.1498	−1.35644	1.13993	0	0.35771	0.82287	0	0
φ, **B**	35.6254	−2.88178	1.62469	0	0.43748	1.90438	0.28168	0.23115
φ, **C**	31.1770	1.14492	2.38089	0.38803	0	2.07291	0	1.78939
ψ, **A**	28.0089	−0.60321	1.29464	0	0.02106	1.3024	0	0.08051
ψ, **B**	30.3457	2.65584	1.55338	0.13681	0.37975	1.47698	0	0.20224
ψ, **C**	14.6367	0.93443	0	0.25603	0.01605	0.45829	0.22928	0.06936

For each dihedral, the three modes are successively labeled as A, B, and C