Accounting for a mirror-image conformation as a subtle effect in protein folding

Khatuna Kachlishvili,* Gia G. Maisuradze,‡ Oslo A. Martin,‡ Adam Liwo,* Jorge A. Vila,‡ and Harold A. Scheraga* 1

*Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301; ‡Instituto de Matemática Aplicada San Luis, Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina, Universidad Nacional de San Luis, 5700 San Luis, Argentina; and Laboratory of Molecular Modeling, Faculty of Chemistry, University of Gdańsk, 80-308 Gdańsk, Poland

Contributed by Harold A. Scheraga, April 29, 2014 (sent for review February 25, 2014)

By using local (free-energy profiles along the amino acid sequence and 13C chemical shifts) and global (principal component) analyses to examine the molecular dynamics of protein-folding trajectories, generated with the coarse-grained united-residue force field, for the B domain of staphylococcal protein A, we are able to (i) provide the main reason for formation of the mirror-image conformation of this protein, namely, a slow formation of the second loop and part of the third helix (Asp29–Asn35), caused by the presence of multiple local conformational states in this portion of the protein; (ii) show that formation of the mirror-image topology is a subtle effect resulting from local interactions; (iii) provide a mechanism for how protein A overcomes the barrier between the metastable mirror-image state and the native state; and (iv) offer a plausible reason to explain why protein A does not remain in the metastable mirror-image state even though the mirror-image and native conformations are at least energetically compatible.

Significance

Numerous diseases, including neurodegenerative diseases, are caused by misfolding of proteins; therefore, understanding the mechanism by which a protein folds to its native conformation is one of the most important problems of biophysics. This task becomes more complicated for proteins with a high degree of symmetry that, in addition to the native structure, allows the existence of another, energetically very close to the native conformation. It would have been desirable if the mirror-image conformation and its evolution to the native state could be detected by NMR spectroscopy. Nevertheless, by using local [13C chemical shift (25) and free-energy profiles (FEPs) along the amino acid sequence (26–28)] and global [principal component (PC) (29)] analyses (SI Materials and Methods), we examined molecular dynamics (MD) trajectories of protein A, generated with the coarse-grained united-residue (UNRES) force field (27, 30–32) (Fig. S1 and SI Native and Mirror-Image Structures of Protein A). Hence, with an experimental technique such as circular dichroism, used to estimate the fraction of secondary-structure content, it is almost impossible to distinguish the mirror-image structure from the native structure. It would have been desirable if the mirror-image conformation and its evolution to the native state could be detected by NMR spectroscopy. Nevertheless, by using local [13C chemical shift (25) and free-energy profiles (FEPs) along the amino acid sequence (26–28)] and global [principal component (PC) (29)] analyses (SI Materials and Methods), we examined molecular dynamics (MD) trajectories of protein A, generated with the coarse-grained united-residue (UNRES) force field (27, 30–32) (Fig. S1 and SI Native and Mirror-Image Structures of Protein A). These analyses of the MD trajectories, in which folding from a fully unfolded conformation occurs either almost instantly or through a metastable state formed by the mirror-image topology, enabled us to elucidate the origin of the formation of a mirror-image topology and how the protein emerges from the kinetic trap and folds to the native state.

The results presented in this work are based on the analysis of four pairs of MD trajectories at 270 K (in each pair, one trajectory folds directly to the native state and the other folds...
through the metastable mirror-image state) selected from 96 MD simulations, which we carried out in a broad range of temperatures (details in Materials and Methods). The mirror-image conformation in the studied trajectories (an illustrative example of two trajectories is in Fig. S3), and these results are in agreement with those of earlier studies (12, 24).

**Results**

**Free-Energy Profile.** To understand what induces protein A to fold into its native conformation from a fully unfolded conformation almost instantly or through a kinetic trap, the local motions of each residue were examined along the sequence. In particular, the FEPs along the backbone virtual-bond angle \( \theta \) and the backbone virtual-bond dihedral angle \( \gamma \) of each residue (defined in Fig. S2) were examined.

The FEPs along the \( \theta \) and \( \gamma \) angles of the entire trajectory \( \mu(\theta) = -k_B T \ln P(\theta), \mu(\gamma) = -k_B T \ln P(\gamma) \), where \( P, T, \) and \( k_B \) are the probability distribution function (PDF), the absolute temperature, and the Boltzmann constant, respectively) are very helpful to identify the key residues in the folding process (26–28); however, the analysis of the FEPs of the entire trajectory does not provide information about the way in which each residue explores its own FEP in the course of time or to what extent the motion of each residue is coupled to the global motion of the protein as it proceeds toward its native state. To answer these questions, for selected trajectories, the FEPs along the \( \theta \) and \( \gamma \) angles for certain periods of time, during which significant structural changes occur before the protein reaches its native state, were calculated.

It should be noted that the FEPs presented here are effective FEPs because they are computed from a nonequilibrium probability density and depend on the time duration and on the initial conditions of the trajectory. The effective FEP differs from the actual FEP, which is an equilibrium thermodynamic property, and should be computed from the entire sets of trajectories (folding and nonfolding). Because of the dependence of the effective FEP on the time duration of the trajectory and on the initial conditions, the effective FEP was used (27, 28) to analyze the MD trajectories in detail and extract the reasons why a protein folds or does not in a single MD trajectory. In the present work, the effective FEPs were used to explain why protein A folds with or without a kinetic trap.

In this work, selection of time intervals, over which the FEPs were calculated, was based on significant changes in the C\(^\alpha\) root-mean-square deviation (rmsd) from the native structure of trajectories (details in Fig. S4 and SI Interval Strategies for Free-Energy Profiles).

**FEPs Along \( \theta \) and \( \gamma \) Angles of the Folding Trajectories With and Without Kinetic Traps.** By comparing the FEPs along all of the \( \theta \) and \( \gamma \) angles, computed at different time intervals, of two trajectories, one of which folds without and the other with a kinetic trap, we found that the FEPs along 17 of the \( \theta \) angles and 17 of the \( \gamma \) angles differ noticeably from each other between the two trajectories and consequently play a crucial role in a folding pathway. The discrepancies between these FEPs are caused by different behaviors of the angles in different time intervals. Based on these discrepancies, FEPs can, overall, be divided into two categories: (i) the FEPs in which differences appear at the beginning of the trajectory before the protein jumps into either the native or the mirror-image state and (ii) the FEPs in which differences appear after the collapse.

The first category includes the FEPs along the angles pertaining to the second loop and its vicinity including part of the third helix; the second category includes the FEPs along the angles pertaining to the first loop and its vicinity, including parts of the first and second helices. In particular, most of the angles of the first-category FEPs are flexible at the beginning of the trajectory and explore a large region of angle space before jumping into the global minimum in the trajectory with a kinetic trap, whereas the same angles of the trajectory without a kinetic trap gradually explore the region of only their own global minima during the entire trajectory (an illustrative example is shown in Fig. 1 A and B).

The behavior of the angles of the second-category FEPs is more complicated; for example, some of these angles of the trajectory without a kinetic trap completely explore the shallow local minimum before the protein jumps to the native state, whereas the same angles of the trajectory in which protein A folds through a kinetic trap jump back and forth between the local and global minima during the entire trajectory (an illustrative example is shown in Fig. 1 C and D); differences in behavior of the remaining angles of the second-category FEPs do not have any particular pattern, although protein A explores a larger portion of conformational space when it folds through a kinetic trap, manifested by a small number of regions in which the FEPs are undefined (Fig. S5 and SI FEPs Along \( \theta \) and \( \gamma \) Angles of the Folding Trajectories with and Without Kinetic Traps, in which the FEPs along all of the \( \theta \) and \( \gamma \) angles are shown).

Based on the results of the FEP analysis, it can be concluded that all residues of the first and second loops and their edges, including the parts of all three helices, play a crucial role in the folding pathway of protein A. In particular, residues of the second loop, its edges, and part of the third helix are responsible for formation of the mirror-image topology; and residues of the first loop and its edges along with the parts of the first and second helices assist the protein to overcome a barrier between the metastable mirror-image and native states.

To justify and strengthen the aforementioned statements resulting from the analysis of only one pair of trajectories, we selected three additional pairs of MD trajectories, in which protein A folds without and through a kinetic trap, respectively, and computed the FEPs along the \( \theta \) and \( \gamma \) angles for these trajectories. The obtained results are in agreement with those shown in Fig. 1 and Fig. S5; however, because of the tremendous number of FEPs (each trajectory contains 89 panels of FEPs), we do not show the results for the other three pairs. Instead, based on the results obtained from all four pairs of trajectories, we
plotted the amino acid sequence of 1BDD (Fig. 2), in which the residues in rectangles belong to helices, the residues outside of the rectangles belong to the loops, the residues in red color are parts of both $\theta$ and $\gamma$ angles along which the FEPs differ from each other, and the residues in green color are parts of only one of those angles whose FEPs differ from each other.

**$^{13}C^\alpha$ Chemical Shift Analysis.** By using the CheShift-2 Server (25), the $^{13}C^\alpha$ chemical shifts were calculated for each conformation of the same MD trajectories and during the same time intervals as in the FEP analysis (Fig. S4). The results obtained from the analysis of the $^{13}C^\alpha$ chemical shifts coincide with those from the analysis of the FEPs. In particular, the main differences between the results, depicted in Fig. 3 A and B, which corresponds to the $^{13}C^\alpha$ chemical shifts calculated before the collapse of the protein, are in the second loop region (residues 30–32) and in part of the third helix (residues 33–39). The $^{13}C^\alpha$ chemical shifts of the residues of the first loop and those from the experimental $^{13}C^\alpha$ chemical shifts of the residues of the same portions of the protein folding with a mirror image (Fig. 3A) are either acceptable (white bars) or unacceptable (red bars). A graphical representation of these differences is shown in Fig. 3C. In particular, each color bar of each residue in Fig. 3C is a difference between the corresponding color bars of the same residue in Fig. 3A and B (interpretation of Fig. 3C in SI $^{13}C^\alpha$ Chemical Shift Analysis).

Noticeable differences, between the $^{13}C^\alpha$ chemical shifts calculated during the time interval when the protein remains in the metastable mirror-image state (Fig. S6 D and E) and the full trajectory (Fig. 3 D and E), are in the region of the first loop (residues 11–15) (Fig. S6F, Fig. 3F, and SI $^{13}C^\alpha$ Chemical Shift Analysis). The reason for these differences is that the first loop does not reach the native geometry in the metastable mirror-image state and tries to emerge from the kinetic trap; during this time the structure of the first loop undergoes drastic changes. Consequently, the differences between $^{13}C^\alpha$ chemical shifts of the residues of the first loop and those from the experimental chemical shifts are larger (details in Fig. S6 and SI $^{13}C^\alpha$ Chemical Shift Analysis). As for the FEPs, the analysis of the $^{13}C^\alpha$ chemical shifts was also performed for three other pairs of trajectories. The results obtained for those other pairs are in agreement with

---

**Fig. 3.** $\Delta^\sigma_i$ represents the differences between the observed and theoretical $^{13}C^\alpha$ chemical shifts computed from the MD trajectories. The different colors of the bars indicate the magnitude of the differences ($\Delta_\sigma$) in terms of $\sigma$ ($\sigma = 1.7$ ppm (25)). In particular, blue bars correspond to $\Delta_\sigma \leq \sigma$, white bars correspond to $\sigma < \Delta_\sigma \leq 2\sigma$, and red bars correspond to $\Delta_\sigma > 2\sigma$. A full bar highlighted in yellow, as for residue Gly21, indicates that the chemical shift was not measured experimentally or (for all of the remaining partial bars highlighted in yellow) that the theoretical value could not be computed for a particular conformation. A, D and B, E illustrate the differences in $^{13}C^\alpha$ chemical shift per residue for the trajectories that fold with and without mirror image, respectively. In particular, A and B correspond to the time interval from the start of the simulation until the protein starts to collapse; and D and E correspond to the full trajectory of the simulation. C and F illustrate the second-order differences, computed as follows: $\Delta^\sigma_{i,C} = \Delta^\sigma_{i,A} - \Delta^\sigma_{i,B}$ and $\Delta^\sigma_{i,F} = \Delta^\sigma_{i,D} - \Delta^\sigma_{i,E}$.
those presented here; therefore, we present the results for only one pair.

It should be noted that the observed $^{13}$C$\alpha$ chemical shifts were obtained from the Biological Magnetic Resonance Bank (BMRB), deposited under accession no. 5656, which correspond to $^{13}$C$\alpha$ chemical shift values obtained from the Z domain, rather than the B domain, of protein A (33). Nevertheless, the Z domain differs from the B domain by only two substitutions, namely Ala1 → Val and Gly30 → Ala. However, because we studied the 10- to 55-residue fragment of the B domain of staphylococcal protein A, the only relevant substitution is that of Gly30 → Ala, which is equivalent to Gly21 → Ala in our renumbered sequence. For this reason, all panels in Fig. 3 and Fig. S6 show bar corresponding to residue Gly21 highlighted in yellow, i.e., indicating that the observed $^{13}$C$\alpha$ chemical shift value is missing. Overall, because the B and Z domains exhibit identical binding affinity (33), the use of the observed information from the Z rather than the B domain of protein A is reasonable.

**Principal Component Analysis of the Folding Trajectories With and Without Kinetic Traps.** Both trajectories, in which protein A folds with and without a kinetic trap, were also analyzed by principal component analysis (PCA). In particular, the contributions of the two main principal modes (with the largest eigenvalues) to the mean-square fluctuations along the $\theta$ and $\gamma$ angles were calculated. The peaks in these contributions appear exactly in the regions where the main differences between the shapes of the FEPs and those of the $^{13}$C$\alpha$ chemical shifts were found. However, differences between the contributions of the two main principal modes of these trajectories are not very noticeable, which indicates that the mirror image is a local property and cannot be detected by the global motions of PCA (details in Fig. S7 and SI Materials and Methods, Principal Component Analysis).

**Origin of Formation of the Mirror-Image Conformation.** The FEPs and $^{13}$C$\alpha$ chemical shifts analyses enabled us to identify the parts of protein A involved in formation of the mirror-image conformation. However, the question of what induces protein A either to fold almost instantly or to misfold first and then emerge from the mirror-image state still needs to be answered. Therefore, we first examined, in detail, the dynamics of the 24- to 37-residue portion of protein A, which includes the second loop and parts of the second and third helices, at the beginning of the trajectory before the protein jumps into either the native or the mirror-image state. In particular, we examined the local interactions by computing the C$\alpha$–C$\alpha$ distances between the residues as a function of time in the 24- to 37-residue portion. It appears that, in the trajectory with the kinetic trap, the residues of the second loop and part of the third helix (Asp29–Asn35) are in local conformational states that are in an extended region with larger values of virtual-bond angle $\theta$ [and, consequently, larger C$\alpha$–C$\alpha$ distances (Fig. S8)]. The residence of some residues in the extended conformational states seems to make the second loop and part of the third helix reach the native geometry much slower in the trajectory with the kinetic trap (Fig. 4B) than in the trajectory without the kinetic trap (Fig. 4A). This delay of formation in Fig. 4B is enough to direct the N-terminal portion of the chain to pack against the wrong side of the helical hairpin formed by the second and the third helix (Fig. 4 C and D).

Thus, the formation of the mirror-image conformation may be caused by the presence of multiple local conformational states in the second loop and part of the third helix (Asp29–Asn35). It should be noted that, based on the FEPs and $^{13}$C$\alpha$ chemical shift analyses, the large portion of the third helix (Ser33–Lys42) might also be involved in formation of the mirror-image conformation; however, a visual inspection of each trajectory revealed that formation of the mirror-image conformation was initiated by the residues of the second loop and the neighboring part of the third helix (Asp29–Asn35).

Another three-helix bundle domain of staphylococcal protein A, namely the E domain (PDB code: 1edk), also forms the mirror-image topology (24). Remarkably, the B and E domains of protein A have identical sequences in the Asp29–Asn35 fragment, which are present in the second loop and in part of the third helix (Fig. 2). This observation reinforces our conclusion that this portion of protein A might be responsible for formation of the mirror-image conformation.

**The Mechanism by Which Protein A Emerges from the Metastable Mirror-Image State.** The FEPs and $^{13}$C$\alpha$ chemical shifts analyses enabled us to identify the fragment of the sequence of protein A that may be actively involved in surmounting the kinetic trap. However, the mechanism of how protein A overcomes the barrier between the metastable mirror-image state and the native state and what makes the protein undergo this transition has not yet been explained. That is why we examined the behavior of the helices in both the mirror-image and the native conformations. In particular, for the trajectory that folds through the mirror image, we calculated the distances between C$\alpha$s of selected nonpolar residues, pertaining to the first and second (Fig. 5A), the first and third (Fig. 5B), and the second and third (Fig. 5C) helices over 50 ns, which form hydrophobic contacts either in both the mirror-image (during the first 28 ns) and native conformations (Fig. 5 A and C) or in one of them (Fig. 5B). *Insets* in Fig. 5 A–C represent the PDF of each distance computed for both the mirror image state and the native state.

---

**Figure 4.** (A–F) Anaglyph stereo diagrams of the 24- to 37-residue portion of protein A at 45 ps of the trajectories without (A) and with (B) a kinetic trap. At 300 ps the whole protein collapses and forms either a molten globule (C) or a mirror-image (D) conformation; at 28 ns the protein emerges from a kinetic trap by opening the conformation of the first loop (E) and then proceeds to the native state (F) by adopting a closed-loop conformation.
The average distance between C’s of nonpolar residues, pertaining to the first and second helices, and the amplitudes of the fluctuations of this distance in the metastable mirror-image state (Fig. 5 A and C) are greater than those of the other pairs (Fig. 5 B). Moreover, the bimodal broad distribution of the PDF in the mirror-image state is observed only for the pairs of residues pertaining to the first and second helices (Fig. 5 A), which is replaced by a narrow unimodal distribution in the native state. These results indicate that this portion of the protein is not stable in the metastable mirror-image state. The average distance between the nonpolar residues pertaining to the first and third helices (Fig. 5 B) is smaller in the metastable mirror-image state than in the native state. Such behavior is understandable, because the Phe5–Leu26 pair forms hydrophobic contacts only in the mirror-image conformation. The average distance between the nonpolar residues pertaining to the second and third helices (Fig. 5 C) is slightly larger in the metastable mirror-image state than in the native state. The amplitudes of fluctuations of the distances in the mirror-image state and in the native state are similar for both pairs of helices (first–third and second–third). However, the amplitudes of the fluctuations of the distance, between the nonpolar residues pertaining to the first and second (Fig. 5 A) and the first and third (Fig. 5 B) helices, increase significantly when the protein is about to overcome the barrier between the metastable mirror-image and native states. These results indicate that the first helix is the least stable helix in the metastable mirror-image state, and its instability induces the first loop to open and then close to place the third helix in the native location (Fig. 4 E and F). The behavior of the distances between the C’s of other nonpolar residues than those shown in Fig. 5 is shown in Fig. S5.

Another visual justification of our conclusions can be seen in Fig. S4, in which the structures of protein A corresponding to significant changes in both trajectories clearly show how protein A misfolds (Fig. S4 C) and then emerges from the metastable mirror-image state (Fig. S4 B). Stereo diagrams of folding pathways for both trajectories are shown in Fig. S10.

**Discussion and Conclusions**

By analyzing the MD trajectories of the 10- to 55-residue fragment of the B domain of staphylococcal protein A generated with the coarse-grained UNRES force field, by FEPs along the amino acid sequence, by $^{13}$C$^\alpha$ chemical shift, and by PCA methods, we have investigated the reasons for formation of the mirror-image topology in protein A and found the following:

i) The slow formation of the second loop and part of the third helix (Asp29–Asn35), which seems to play a crucial role in formation of the mirror-image conformation, appears to be caused by the presence of multiple local conformational states of the residues of the second-loop region (Asp29–Asn35).

ii) In the trajectory with the mirror image, the residues of the second-loop region frequently visit more extended states (larger values of $\theta$; Fig. 1 B), whereas in the trajectory leading directly to the native structure, the second-loop region is more folded (smaller values of $\theta$; Fig. 1 A). Consequently, the second-loop region of the chain seems to reach the native geometry slowly and, in this way, it can enable the N-terminal portion of the chain to pack against the wrong side of the helical hairpin formed by the second and the third helix.

iii) The “opening and closing” of the first loop may assist the first helix to jump over the plane of the second and third helices and consequently help protein A overcome the barrier, between the metastable mirror-image state and the native state, and fold to the native state. The computed difference between the total free energies of the mirror-image and the native conformations is only a few kilocalories per mole (12, 24) and, hence, it is not possible to identify only one particular type of interaction as being responsible for surmounting the kinetic trap during folding of protein A. In other words, any type of change in the intramolecular interactions could make protein A leave the energetically compatible metastable mirror-image state.

iv) PCA was unable to detect formation of the mirror-image conformation, which might indicate that this conformation is a local property of protein A.

In the end, it should be noted that, although we computationally proved the plausibility of formation of the mirror-image conformation, we did not argue whether or how the mirror-image conformation could be detected experimentally. The point is that, based on our results, the differences in the local geometry of the second loop and part of the third helix regions between the pathways leading to the native and those leading to the mirror-image structures of protein A occur only at the beginning of the trajectories, before the protein collapses, within the tenth-of-microsecond timescale (because of the extension of the timescale of UNRES [27, 32]).

Another important point, which makes this problem very interesting, is the functional consequences that protein A may have because of the possible coexistence of the native and mirror-image folds. The point is that, besides the energetic closeness of these two conformations, the frequency of the transitions between the folded and mirror-image states increases with increase
of temperature, and the mirror-image state becomes more and more probable with respect to the native state. In particular, the free-energy differences between the native and mirror-image folds that resulted from a cluster analysis of multiplexed replica exchange molecular dynamics (MREMD) simulations at different temperatures \[ \Delta F = k_B T \ln \left( \frac{\rho_{\text{native}}}{\rho_{\text{mirror}}} \right) \], where \( k_B \) is the Boltzmann constant, the absolute temperature, \( \rho_{\text{mirror}} \), and \( \rho_{\text{native}} \) are the populations of mirror-image and native clusters, respectively, are 4.15 kcal/mol (at 270 K), 3.15 kcal/mol (at 280 K), and 1.95 kcal/mol (at 300 K) and, at the folding-transition temperature, the populations of the conformations with the native and mirror-image topology are nearly equal (17). Hence, the question can be raised of whether protein A in the mirror-image conformation performs the same function as it does in the native conformation. In general, it should not be the same, because the mirror-image conformation is a misfolded conformation; however, without experimental evidence for the existence of the mirror-image conformation, we cannot argue this issue.

**Materials and Methods**

We carried out 96 canonical MD simulations in a broad range of temperatures (270–350 °C with the UNRES force field parameterized (34) on the 1GAB (35) protein. The UNRES force field takes the solvent into account implicitly, through the mean-force potential of interactions between united side chains (34). The folding was found to occur either directly to the native state or through a kinetic trap, mainly the topological mirror image of the native three-helix bundle. The latter folding scenario was observed more frequently at low temperatures (e.g., protein A folds through a kinetic trap in 7 of 16 trajectories at 270 K). Therefore, the four pairs of trajectories that were selected for detailed analysis corresponded to those at low temperature (270 K). It should be noted that a visual inspection of the remaining pairs of MD trajectories at 270 K revealed a similar mechanism for formation of the mirror-image conformation and for its emergence from the metastable mirror-image state. The Berendsen thermostat (36) was used to maintain constant temperature. The time step in molecular dynamics simulations was \( \tau = 0.1 \mu s \) (1 mu = 48.9 fs) is the "natural" time unit of molecular dynamics (37) and the coupling parameter of the Berendsen thermostat was \( s = 1 \mu s \). A total of 10^6 MD steps were run for each trajectory.

**Acknowledgments.** This work was conducted by using the resources of (i) our S88-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University; (ii) the National Science Foundation (NSF) Terascale Computing System at the Pittsburgh Supercomputer Center; (iii) the Beowulf cluster at the Department of Computer Science, Cornell University; (iv) the Informatics Center of the Metropolitan Academic Network in Gdansk; and (v) the Interdisciplinary Center of Mathematical and Computer Modeling at the University of Warsaw. This work was supported by National Institutes of Health Grant GM-14312, NSF Grant MCB10-19767, Grant PIP-121-20100030 from Instituto de Matemática Aplicada San Luis–Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina, and Project 328402 from Universidad Nacional de San Luis, Argentina.
Supporting Information

Kachlishvili et al. 10.1073/pnas.1407837111

SI Native and Mirror-Image Structures of Protein A

The mirror-image and native conformations have identical secondary structures, the only difference between them is the location of the third (C-terminal) α-helix. Fig. S1 illustrates the native structure of the 10- to 55-residue fragment of protein A from the Protein Data Bank (PDB) (Fig. S1A) and the mirror image of the native structure generated with the UNRES force field (Fig. S1B). It should be noted that the residues of the fragment of protein A studied here (main text and SI Text) are numbered starting from 1, which corresponds to residue no. 10 in the 1BDD PDB structure.

SI Materials and Methods

United Residue Model. The united residue (UNRES) model of polypeptide chains (1–10) is illustrated in Fig. S2. A polypeptide chain is represented as a sequence of α-carbon (Cα) atoms linked by virtual Cα...Cα bonds with united peptide groups halfway between the neighboring Cαs and united side chains (SC), whose sizes depend on the nature of the amino acid residues, attached to the respective Cαs by virtual Cα...SC bonds. The effective energy is expressed by Eq. S1 (9),

\[ U = w_{SC} \sum_{i<j} U_{SC,SC} + w_{SCP} \sum_{i<j} U_{SC,LP} + w_{pp} f_2(T) \sum_{i<j} U_{pp} + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i) + w_{bond} f_2(T) \sum_i U_{bond}(\gamma_i, \gamma_{i+1}) + w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\alpha_{SC}, \beta_{SC}, \theta_i) + w_{bond} \sum_i U_{bond}(d_i) + \sum_{m=3}^6 w_{m}^{(m)} f_m(T) U_{m}^{(m)} + w_{SS} \sum_i U_{SS,i} \]

[S1]

with (9)

\[ f_m(T) = \frac{\ln(e + e^{-1})}{\ln(\exp[(T/T_0)^{m-1}] + \exp[-(T/T_0)^{m-1}])}; \quad T_0 = 300 \text{ K}. \]

[S2]

where the successive terms represent side chain–side chain, side chain–peptide, peptide–peptide, torsional, double-torsional, bond-angle (3), side-chain local (dependent on the angles \( \alpha \) and \( \beta \) of Fig. S2), distortion of virtual bonds, and multibody (correlation) interactions and formation of disulfide bonds, respectively. The \( w \)'s are the relative weights of each term. The side chain–side chain potentials include interactions with the solvent (2). The correlation terms arise from a cumulant expansion (4, 11) of the restricted free-energy function of the simplified chain obtained from the all-atom energy surface by integrating out the secondary degrees of freedom. The temperature-dependent factors of Eq. S2, introduced in our later work (9) and discussed further in ref. 12, reflect the fact that the UNRES effective energy is an approximate cumulant expansion of the restricted free energy. The virtual-bond vectors are the variables used in molecular dynamics.

\(^{13}\text{C}^\alpha\) Chemical Shift. Although the details of this section have already been published (13), a brief description is provided for the convenience of the reader. For each amino acid \( \mu \), it is possible to define the differences \( \Delta_\mu \) between observed and predicted \(^{13}\text{C}^\alpha\) chemical-shift values as

\[ \Delta_\mu = \frac{13C_{\text{observed}, \mu} - 1}{\Omega} \sum_{i=1}^{13} C_{\mu,j} \]

[S3]

where \( C_{\mu,j} \) is the chemical shift of residue \( \mu \) in conformation \( i \) of \( \Omega \) conformations. The average of the predicted chemical shifts over the \( \Omega \) conformations is taken because proteins in solution exist as an ensemble of conformations.

The procedure for mapping the \( \Delta_\mu \) values onto a 3D protein model was formulated as follows. First, the \( \Delta_\mu \) value computed for each residue \( \mu \) is smoothed by averaging it over the values of the two nearest-neighbor residues. Second, the resulting averaged \( \langle \Delta_\mu \rangle \) value is discretized according to the following rule:

\[ \langle \Delta_\mu \rangle \text{discretization} = \begin{cases} 1, & \langle \Delta_\mu \rangle \leq \sigma \text{ (i.e., } 1.70 \text{ ppm)} \\ 0, & \sigma < \langle \Delta_\mu \rangle \leq 2\sigma \text{ (i.e., } 3.40 \text{ ppm)} \\ -1, & \langle \Delta_\mu \rangle > 2\sigma \end{cases} \]

[S4]

The cutoff value of 1.7 ppm was obtained as explained in ref. 13. Third, the \( \langle \Delta_\mu \rangle \text{discretization} \) values, 1, 0, and −1 are mapped onto a 3D protein model and associated with a color, blue, white, and red, respectively. Blue-colored residues are those with small differences between the observed and computed chemical shifts, and white and red indicate medium and large differences, respectively.

Principal Component Analysis. A detailed description of the principal component analysis (PCA) method is available in our previous papers (10, 14–17) and in an earlier reference (18); therefore, only a brief outline of the approach is presented here. PCA, a covariance-matrix–based mathematical technique, is an effective method for extracting important motions from molecular dynamics (MD) simulations. In PCA, the Cartesian or internal coordinate space is rotated to a new space with new coordinates, principal components (PCs), a few of which are sufficient to describe a large part of the fluctuations of a protein. Here, structural fluctuations of the UNRES \( \theta \) and \( \gamma \) angles [mean-square fluctuations (MSF)] can be decomposed into collective modes by PCA. The modes have “frequencies” and directions corresponding to the eigenvalues and eigenvectors of the covariance matrix. The modes \( k \) with the largest eigenvalues \( (\lambda^2) \) correspond to the modes that contribute the most to the structural fluctuations of the protein.

The contribution of each angle \( \theta \) and \( \gamma \) to mode \( k \) is called the influence, \( \nu_k^2 \) (17, 19, 20).

SI Energetic Competitiveness of Native and Mirror-Image Structures

To determine whether the mirror-image structure is energetically competitive with the native conformation, we have selected two MD trajectories at 270 K, as an example. Both of them start from the fully extended unfolded conformation; in one of them, protein A folds to its native conformation almost instantly from a fully unfolded conformation, and in the second trajectory, before jumping to the native state, protein A becomes trapped in a metastable state formed by a mirror-image conformation. Fig. S3, in which the \( \Delta\gamma \) root-mean-square deviation (rmsd) from the native structure vs. energy of each snapshot for these two trajectories is plotted, shows that a difference between the lowest energies of the mirror-image and native conformations (indicated by two blue arrows in Fig. S3B) is only ~2.5 kcal/mol; this
indicates that the mirror-image structure is energetically competitive with the native structure. This result is in almost quantitative agreement with the energy difference (1.5 kcal/mol), between the lowest energies of the native and mirror-image conformations, obtained from an atomically detailed folding simulation of protein A (21). Fig. S3A corresponds to the MD trajectory, in which protein A folds to its native conformation almost instantly from a fully unfolded conformation, and hence only one cluster is seen. There are two clusters in Fig. S3B, representing the MD trajectory in which protein A folds through the kinetic trap; the solid circles with 8.0 Å ≤ rmsd ≤ 10.0 Å correspond to the mirror-image topology, and the solid circles with rmsd ≤ 4.0 Å correspond to the native state.

### S1 Time Intervals for Free-Energy Profiles

Fig. S4 shows how the periods of time, over which the free-energy profiles (FEPs) were calculated, were selected for one of the pairs of MD trajectories (the same principle was used for the selection of time intervals for the other pairs of trajectories). Fig. S4A illustrates the rmsd as a function of time of the folding trajectory without a kinetic trap (black) and the folding trajectory with a kinetic trap (red). The first ∼29 ns (Fig. S4B) and then the first ∼700 ps (Fig. S4C) of both trajectories were selected and expanded. Moreover, the structures of protein A corresponding to significant changes in both trajectories are shown in Fig. S4 B and C. During the first ∼250 ps protein A remains unfolded in both trajectories (brown rectangle in Fig. S4C), although one of them (the trajectory without a kinetic trap) makes an attempt to fold to either its native or mirror-image topology within the first ∼100 ps (pink rectangle in Fig. S4C). Between ∼250 ps and ∼300 ps (brown and blue rectangles, respectively, in Fig. S4C), the system in both trajectories collapses, in one of them (the trajectory with a kinetic trap) to the mirror-image conformation, and remains in a metastable state for the next ∼27 ns (green rectangle in Fig. S4B), and, as shown in the representative structures, the mirror-image conformation converts to the native conformation at ∼28 ns; and in the second one (the trajectory without a kinetic trap) it first forms a molten-globule structure, as shown in the representative structure (red rectangle in Fig. S4C), and then (∼650 ps) jumps into the native state (illustrated by the representative structure in Fig. S4C).

### S1 FEPs Along θi and γi Angles of the Folding Trajectories with and Without Kinetic Traps

Fig. S5 shows the FEPs along all of the θi and γi angles computed at different times from the whole folding trajectory without a kinetic trap (Fig. S5 A and B, respectively) and the folding trajectory with a kinetic trap (Fig. S5 C and D, respectively). As in Fig. S4, the blue, red, green, and black curves correspond to the FEPs calculated over 300 ps, 650 ps, and 27 ns and over the entire duration of the trajectories, respectively. Because protein A jumps almost instantly (Fig. S4) into either its metastable state formed by the mirror-image conformation or its native state, the FEPs calculated over 100 ps and 250 ps (the pink and brown rectangles, respectively) are very shallow and overlap each other, and for most of the angles they do not contain extra information, which was not revealed by FEPs over 300 ps; therefore, the FEPs over these time intervals are omitted in Fig. S5. However, taking into account the importance of the initial time intervals in formation of the mirror-image topology, the important differences observed between the FEPs over 100 ps and 250 ps along a few angles are discussed in the main text.

By comparing the FEPs along the θ angles of the folding trajectory without a kinetic trap (Fig. S5 A and B) to the FEPs along the θ and γ angles of the folding trajectory with a kinetic trap (Fig. S5 C and D), we found that 17 θ angles and 17 γ angles differ noticeably from each other and consequently play a crucial role in a folding pathway. In particular, of these 17 θ angles, 4 (θi = 13, 14, 15, 30, 31) are from loops (blue numbers), 7 (θi = 8, 17, 34, 35, 41–43) represent helices (red numbers), and 6 (θi = 10, 11, 15, 16, 32, 33) represent edges, which include residues from both loop and α-helix (green numbers). Of these 17 γ angles, 3 (γi = 12, 13, 30) are from loops, 7 (γi = 7, 17, 20, 21, 26, 37, 40) are from helices, and 7 (γi = 9, 10, 14–16, 28, 31) are from edges. The plots in Fig. S5 show that the discrepancies between the FEPs with and without a mirror image are caused by different behaviors of the angles in the different time intervals. In particular, the θ angles (Fig. S5A) of the first loop (θi = 13, 14) and the edge (θi = 15), for the trajectory in which protein A folds without a kinetic trap, completely explore the shallow local minimum or “shoulder” in the region of ∼110°–140° before the protein jumps to the native state, whereas the same θ angles of the trajectory in which protein A folds through a kinetic trap jump back and forth between the local and global minima during the entire trajectory (Fig. S5C). The θ angles pertaining to the edge (θi = 10) of both trajectories, plotted in Fig. S5 A and C, have similar FEPs for the entire trajectory. However, the θ10 angle in the trajectory with a kinetic trap explores mainly the local minimum or a shoulder at ∼120° while the whole system is in a kinetic trap, and after the whole system jumps to the native state, the θ10 angle starts exploring the global minimum. The same θ10 angle of the trajectory without a kinetic trap explores the region of 100°–130° at the beginning of the trajectory, while the whole system is in the non-native state, and after the whole system jumps to the native state, the θ10 angle starts jumping back and forth between the local (shoulder) and global minimum.

The differences between the FEPs of the θ1 angles of these trajectories are that the θ1 angles in the trajectory with a kinetic trap explores mainly the region around 120° while the whole system is in the kinetic trap; after the whole system jumps to the native state, the θ1 angle starts exploring global (∼110°) and local (∼90°) minima. The θ1 angle of the trajectory without a kinetic trap gradually explores its own global minimum (∼110°) during the entire trajectory.

The differences between the FEPs along the θ angles pertaining to the second loop (θi = 30, 31), the edge between the second loop and third α-helix (θi = 32, 33), and the third α-helix (θi = 34, 35, 41–43) of the trajectories, in which the protein folds with and without the kinetic trap, are that the θ32, θ33, θ34, θ35, θ36, θ37, and θ38 angles in the trajectory with a kinetic trap are more flexible at the beginning of the trajectory (Fig. S5C) either exploring only the region of 105°–140° (θ30, θ31) or jumping back and forth between the global minimum (90°) and the region of 100°–130° (θ32, θ33, θ34, θ35, θ36, θ37, and θ38), whereas the corresponding angles of the trajectory without a kinetic trap (Fig. S5A) gradually explore only their own global minima (90°) during the entire trajectory [except θ30, which starts jumping back and forth between the global minimum (90°) and the region of 100°–140° after the protein reaches the native state]. The differences between the FEPs along the θ angles pertaining to the helices and edge (θi = 8, 10, 17) of these trajectories are quite similar. In particular, the θ8, θ10, and θ17 angles of the trajectory with a kinetic trap are flexible only at the beginning of the trajectory (650 ps), exploring a large range of FEPs between 90° and 140°, whereas the same angles of the trajectory without a kinetic trap are flexible for ∼27 ns (θ8) and even longer (∼45 ns) with (θ10, θ17) exploring the same regions of the FEPs. After these time intervals, all these θ angles of both trajectories jump to the global minimum (90°).

Based on the FEPs along the γ angles of the selected time periods and the entire trajectory plotted in Fig. S5 B and D, it can be observed that the γ angles of the first and the second loops and edges (γi = 10, 13–15, 28, 30, 31) for the trajectory with the kinetic trap (Fig. S5D) are more flexible and explore the regions of the FEPs (∼70° to 30° for γ10, −160° to −50° for γ13, −180° to −50° for γ14, −150° to −30° for γ15, 60°–160° for
The differences between the rest of the FEPs along the γ angles, pertaining to the first loop and edge (γ, i = 9, 12), are less prominent and appear before the protein jumps to the native state. The differences between the FEPs along the γ angles, pertaining to the helices of these trajectories, are noticeable and appear either before the collapse (γ, i = 26, 37, 40) or after the collapse (γ, i = 7, 17, 20, 21). It is important to mention that, as for the θ angles, the main discrepancies between the FEPs along the γ angles (γ30, γ31) pertaining to the second loop and edge of the second loop and the third α-helix appear at the beginning of the trajectory before the protein jumps into either the native or the mirror-image state (for γ30, the difference can also be observed after the protein remains in the mirror-image metastable state), whereas the differences between the FEPs pertaining to the first loop and edges are related to the time intervals after the collapse.

SI 13C\textsuperscript{a} Chemical Shift Analysis

Fig. S6 illustrates the differences between experimental 13C\textsuperscript{a} chemical shifts and those calculated from the MD trajectories. Fig. S6 A, C, and E and B, D, and F illustrate the 13C\textsuperscript{a} chemical shifts of the trajectories that fold with and without a mirror image, respectively. The time intervals used in the calculations of the 13C\textsuperscript{a} chemical shifts correspond to those used in the analysis of the FEPs. In particular, the results illustrated in Fig. S6 A and B are calculated before the collapse of the protein [brown rectangle (∼250 ps) in Fig. S4C]. The results illustrated in Fig. S6 D and E are calculated when the protein remains in the metastable mirror-image state [green rectangle (∼28 ns) in Fig. S4B], and the results illustrated in Fig. S6 G and H are calculated for the full trajectory (Fig. S4A).

Noticeable differences between the 13C\textsuperscript{a} chemical shifts calculated during the time interval when the protein remains in the metastable mirror-image state (Fig. S6 D and E) and the full trajectory (Fig. S6 G and H) are in the region of the first loop (residues 11–15). For example, in the full trajectory, the percentage of blue bars for the trajectory that folds without mirror image is >45% (Fig. S6H), whereas the percentage of blue bars for the trajectory that folds with a mirror image is ∼35% (Fig. S6G). The reason for these differences is that the first loop does not reach the native geometry in the metastable mirror-image state and, to emerge from the mirror-image to the native conformation, the structure of the first loop undergoes drastic changes during this time; consequently, the differences between the 13C\textsuperscript{a} chemical shifts of the residues of the first loop and those from the experimental chemical shifts are larger.

The graphical representations of these differences are shown in Fig. S6 C, F, and I for 250-ps, 28-ns time intervals, and the full trajectory, respectively. In particular, each color bar of each residue in Fig. S6 C, F, and I is a difference between the corresponding color bars of the same residue in Fig. S6 A and B, D and E, and G and H, respectively. If the heights of the bars in Fig. S6 A, D, and G (trajectory with mirror image) are higher than the heights of the corresponding bars in Fig. S6 B, E, and H (trajectory without mirror image), the differences between them, illustrated in Fig. S6 C, F, and I, are represented by the bars with positive value and vice versa for the negative values. As was expected, the negative tail blue bars in Fig. S6C appear in the regions of the second loop and part of the third helix and in Fig. S6 F and I in the region of the first loop.

We also calculated the 13C\textsuperscript{a} chemical shifts within the exact time interval when these changes occur and, as expected, the heights of the blue bars in the region of the first loop were lower (∼20%) for the mirror-image trajectory.


The native (A) and mirror-image (B) conformations of protein A.

The UNRES model of polypeptide chains. The interaction sites are peptide-bond centers \((p)\) and side-chain ellipsoids of different sizes \((SC)\) are attached to the corresponding \(\alpha\)-carbons with different "bond lengths," \(b_{SC}\). The \(\alpha\)-carbon atoms are represented by small open circles. The equilibrium distance of the \(C^\alpha-C^\alpha\) virtual bonds is taken as 3.8 Å, which corresponds to planar trans-peptide groups. The geometry of the chain can be described either by the virtual-bond vectors \(dC_i(C^\alpha_1...C^\alpha_i)\), \(i = 1, 2, ..., N - 1\) and \(dX_i(C^\alpha_i...SC_i)\), \(i = 2, 3, ..., N - 1\) represented by thick dashed arrows, where \(N\) is the number of residues, or in terms of virtual-bond lengths, backbone virtual-bond angles \(\theta_i\), \(i = 2, 3, ..., N - 1\), backbone virtual-bond-dihedral angles \(\gamma_i\), \(i = 2, 3, ..., N - 2\), and the angles \(\alpha_i\) and \(\beta_i\), \(i = 2, 3, ..., N - 1\) that describe the location of a side chain with respect to the coordinate frame defined by \(C^\alpha_1\), \(C^\alpha_i\), and \(C^\alpha_{i+1}\).

\[\text{Fig. S3.} \quad \text{C}^\alpha_{\text{rmsd}} \text{ vs. energy for the two MD trajectories of protein A at 270 K. The black (A) and red (B) solid circles correspond to the trajectories in which protein A folds without and with a kinetic trap, respectively. Blue arrows indicate the conformations with the lowest energy in each cluster. In B, the top cluster corresponds to mirror-image conformations, and the bottom cluster corresponds to native conformations.}\]
Fig. S4. (A) $C^\alpha$ rmsds vs. time for the MD folding trajectories without a kinetic trap (black) and with a kinetic trap (red). (B and C) $C^\alpha$ rmsds of the first 29-ns period (B) and the first 700-ps period (C) of both trajectories. The pink, brown, blue, and red rectangles in C correspond to $\sim$100-ps, $\sim$250-ps, $\sim$300-ps, and $\sim$650-ps time periods, respectively. The green rectangle in B corresponds to a 28-ns time interval. The structures of protein A in B and C correspond to significant changes in both trajectories.
Fig. S5. (Continued)
Fig. S5. (A–D) FEPs, $\mu(\theta)$ and $\mu(\gamma)$, along the $\theta$ and $\gamma$ angles, respectively, for the folding trajectory without a kinetic trap (A and B, respectively) and the folding trajectory with a kinetic trap (C and D, respectively) of protein A. Blue, red, green, and black curves correspond to FEPs computed over 300 ps, 650 ps, and 28 ns and over the entire MD trajectories, respectively. The blue numbers pertain to FEPs along the $\theta$ and $\gamma$ angles that include only residues of loops, the red numbers pertain to FEPs along the $\theta$ and $\gamma$ angles that include only residues of $\alpha$-helices, and the green numbers pertain to FEPs along $\theta$ and $\gamma$ angles that include residues from edges containing both a loop and an $\alpha$-helix. The NMR-derived structural data (small red solid circles at the bottom of each panel) are computed from the first model of the PDB ID code 1BDD (1).

Fig. S6. (Continued)
Fig. S6. (Continued)
Fig. S6. $\Delta \mu$ represents the differences between the observed and theoretical $^{13}$C$\alpha$ chemical shifts computed from the MD trajectories, and $\sigma = 1.7$ ppm. The different colors of the bars indicate the magnitude of the differences ($\Delta \mu$) in terms of $\sigma$. A full bar highlighted in yellow (Nan, not a number), as for residue Gly21, indicates that the experimental chemical shift was not measured experimentally or that the theoretical value could not be computed for a particular conformation (for all of the remaining partial bars highlighted in yellow). A, D, and G and B, E, and H illustrate the differences in $^{13}$C$\alpha$ chemical shift per residue for the trajectories that fold with and without mirror image, respectively. In particular, A and B correspond to the time interval between the start of the simulation and the collapse of the protein, D and E correspond to the time interval in which the protein remains at the metastable mirror-image state, and G and H correspond to the full trajectory of the simulation. C, F, and I illustrate the second-order differences, computed as follows: $\Delta \Delta \mu_{A(C)} = \Delta \mu_{(A)} - \Delta \mu_{(C)}$, $\Delta \Delta \mu_{D(F)} = \Delta \mu_{(D)} - \Delta \mu_{(F)}$, and $\Delta \Delta \mu_{G(I)} = \Delta \mu_{(G)} - \Delta \mu_{(I)}$. 
Fig. S7. (A–H) Contributions ($v_{ik}^2$) of principal mode 1 (A, C, E, and G) and mode 2 (B, D, F, and H) to the mean-square fluctuations along the $\theta$ (A, B, E, and F) and $\gamma$ (C, D, G, and H) angles for the folding trajectory without a kinetic trap (A–D) and with a kinetic trap (E–H) of protein A.
Fig. S8. (A–D) The distances between C\(^\alpha\)s of Asp29–Ser31 (A), Pro30–Gln32 (B), Ser31–Ser33 (C), and Gln32–Ala34 (D) as function of time for the trajectory without (black line) and with (green line) a kinetic trap. Red line corresponds to experimental distance (1) between the C\(^\alpha\)s of these selected residues.

Fig. S9. (A–C) The distances between Cαs of Phe5–Leu26 (A), Leu9–Ala40 (B), and Ile23–Leu36 (C) as functions of time for the 50-ns time interval (the 0- to 28-ns time interval corresponds to the mirror-image state, and the 28- to 50-ns time interval corresponds to the native state). Insets show probability distribution functions of distances between these residues computed for both mirror image and native states. Red lines correspond to experimental distances (NMR) (1) between the Cαs of these selected residues. Green lines indicate the time when the protein jumps from the metastable mirror-image state into the native state.

Fig. S10. (A and B) Stereo diagrams of the 24- to 37-residue portion of protein A at 45 ps of the trajectories without (A) and with (B) a kinetic trap. (C and D) At 300 ps the whole protein collapses and forms either a molten globule (C) or a mirror-image (D) conformation. (E and F) At 28 ns the protein emerges from a kinetic trap by opening the conformation of the first loop (E) and then proceeds to the native state (F) by adopting a closed-loop conformation.