Folding helical proteins in explicit solvent using dihedral-biased tempering

Cheng Zhang[a] and Jianpeng Ma[a,b,1]

[a]Applied Physics Program and Department of Bioengineering, Rice University, Houston, TX 77005; and [b]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, BCM-125, Houston, TX 77030

Edited by Alan R Fersht, Medical Research Council Laboratory of Molecular Biology, Cambridge University, Cambridge, United Kingdom, and approved March 30, 2012 (received for review July 25, 2011)

Using a single-trajectory-based tempering method with a high-temperature dihedral bias, we repeatedly folded four helical proteins [α_3D (PDB ID: 2A3D, 73 residues), α_3W (1LQ7, 67 residues), Fap1-NR$_\alpha$ (2KUB, 81 residues) and S-836 (2JUA, 102 residues)] and some of the mutants in explicit solvent within several microseconds. The lowest root-mean-square deviations of backbone atoms from the experimentally determined structures were 1.9, 1.4, 1.0, and 2.1 Å, respectively. Cluster analyses of folding trajectories showed the native conformation usually occupied the most populated cluster. The simulation protocol can be applied to large-scale simulations of other helical proteins on commonly accessible computing platforms.

enhanced sampling | molecular dynamics | packing chirality | protein folding

Molecular dynamics (MD) allows one to study protein folding (1–5) in molecular details. It is, however, computationally demanding if we model a protein accurately using an all-atom representation and simulate it in explicitly solvent. Recent advances in simulation methods and improved computer hardware have made it possible to fold a protein up to about 30 to 80 residues (6–17).

In this study, we report folding simulations of four helical proteins of 67 to 102 amino acid residues, along with their mutants, via a dihedral-biased tempering method. The method couples accelerated dihedral dynamics (10, 18, 19) with a fast motion at a high temperature through a single-trajectory-based tempering method reported previously (12). As the dihedral bias was only added at higher temperatures, we were still able to recover unbiased room temperature properties. This method allowed us to study several interesting nonnative conformations in addition to the native one.

We simulated four proteins in Protein Data Bank (PDB): α_3D (PDB ID: 2A3D, 73 residues) (20, 21), α_3W (1LQ7, 67 residues) (22, 23), Fap1-NR$_\alpha$ (2KUB, 81 residues) (24) and S-836 (2JUA, 102 residues) (25) in a few microseconds. The first three are parallel three-helix bundles, and the last is a parallel four-helix bundle. Among the four, only Fap1-NR$_\alpha$ is a natural domain; the rest are artificially designed proteins.

Results

Overview of Folding of Helical Proteins. Protein sequences, simulation conditions, and parameters for the dihedral bias are listed in Tables S1, S2, and S3, respectively. In Fig. 1, we show the trajectory frames with the lowest backbone root-mean-square deviations (RMSDs) 1.9 (α_3D), 1.4 (α_3W), 1.0 (Fap1-NR$_\alpha$), and 2.1 Å (S-836) from the respective experimental structures, which will be referred to as "native" structures even for the designed proteins. By comparison, control simulations using regular MD starting from the native conformations yielded higher average RMSDs (3.2, 2.2, 1.5, and 3.2 Å, respectively) at 300 K, suggesting that the native conformations were stably reached.

Without the experimental structure, cluster analyses (see *Methods* and *SI Text*) showed that the native conformation occupied the largest cluster in most trajectories for α_3D, α_3W, and S-836 (see Table S3). For Fap1-NR$_\alpha$, the native conformation was found in the second largest cluster, possibly due to limited sampling and/or force field inaccuracy (see *SI Text* for a discussion).

For each protein, we show the time series of RMSD for a typical trajectory in Fig. 2. All simulations started from fully extended conformations (Fig. S1) with large RMSDs. Folding and unfolding events, as indicated by the rises and falls of RMSD, typically occurred on a time scale of 0.5 to 2 μs. We have accumulated around 10 folding events in each of α_3D, α_3W, and S-836, although longer and/or more simulations may be needed to further improve statistics.

Since the simulation method used a variable temperature, only a fraction of trajectory frames were pertinent to the room temperature. We only reported results after reweighted to 300 K using the multiple histogram method (26–28; see *SI Text*).

Geometry of a Parallel Helix Bundle. Helical proteins are geometrically interesting, as they present different ways of packing helices. Here we discuss two features in packing helices by using the three proteins α_3D, α_3W, and S-836 as examples (we omitted Fap1-NR$_\alpha$ as it shares the same overall geometry with α_3W).

The first feature is the packing "chirality." If a parallel three-helix bundle protein structure is viewed as three connected rods, after the first two helices are aligned, the third one can be packed against either side of the plane extended by the former two. We call the bundle right-handed if the first helix, when viewed along the middle helix from N to C terminus, can be rotated clockwise to overlap with the third one by an angle less than 180°; otherwise, we call the bundle left-handed. A three-helix bundle of either chirality exists; α_3D is right-handed, while α_3W and Fap1-NR$_\alpha$ are left-handed.

Another packing feature is the helix orientation. Although two successive helices are usually aligned antiparallel, a parallel alignment via a long loop is also possible.

In our trajectories, we found several compact but less stable nonnative conformations, which differed from the native one by the packing chirality or helix orientation. We wished to see what determines the natural geometry.

α_3D. We started with the first three-helix bundle, α_3D (20, 21). A typical trajectory is shown in Fig. S2*A*. The three helices are labeled as A, B, and C from N to C terminus. In Fig. 3*A*, we show the distributions of the chiral distance D (defined in *SI Text*, positive for a right-handed bundle, negative a left-handed one) from three independent folding trajectories, as well as a 300 K regular MD trajectory starting from the native structure. The dis-

Author contributions: C.Z. and J.M. designed research; C.Z. and J.M. performed research; C.Z. and J.M. analyzed data; and C.Z. and J.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: jpma@bcm.tmc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1112143109/-/DCSupplemental.

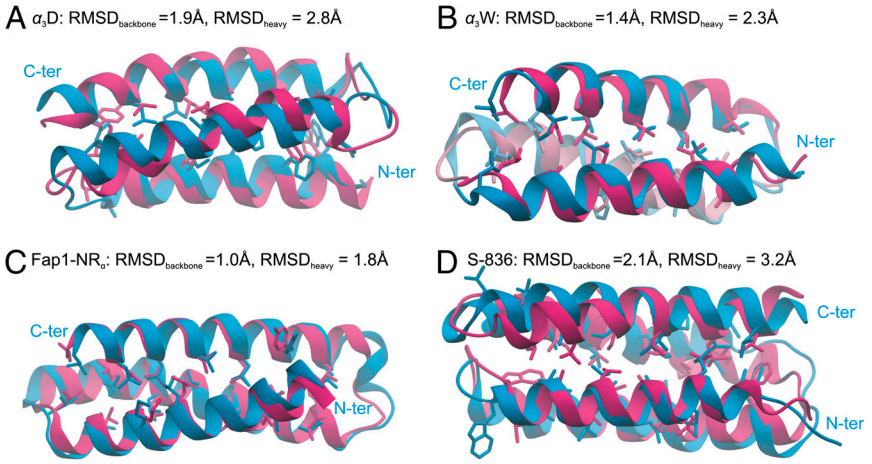

Fig. 1. Best folded structures (blue) of four proteins: (*A*) α_3D, (*B*) α_3W, (*C*) Fap1-NR$_\alpha$, and (*D*) S-836 compared with the experimentally (NMR) determined structures (red).

tributions were dominated by a right-handed (native) peak around $D = 9 \sim 10$ Å, coinciding with that at $D = 9.7$ Å from the native simulation. The widths of the peak still differed among the trajectories, however.

The dominance of the right-handed conformation can be explained by a favorable interhelical charge interaction (20). In the protein, residues in all three helices follow the heptad repeat $a-b-c-d-e-f-g$ (see Fig. 3*A, Left, Inset*), in which the *a* and *d* residues are hydrophobic and the *e* and *g* ones charged or polar. The charge interactions among the *e* and *g* residues of different helices help secure the skeleton.

For α_3D, the *e* and *g* residues on helix A are all positively charged, while those on helix B are negatively charged; on helix C, the *e* residues are positively charged but the *g* ones are negatively charged. In the center diagram of Fig. 3*A*, we show schematically the distributions of these residues in the two opposite chiral conformations (*Right*, native; *Left*, hypothetical) in both top and side views. In the right-handed native conformation (the right panel), the positively charged *e* residues (Arg57 and Arg64) on helix C contact with the negatively charged *g* residues (Glu39 and Glu32) on helix B, and the negatively charged *g* residues (Glu59 and Glu66) on helix C with the positively charged *e* residues (Lys8 and Lys15) on helix A; both are favorable. By contrast, in the left-handed conformation, the *e* and *g* residues on helix C have to contact unfavorably with their counterparts on helices A and B with the same charges.

We further show that the packing chirality can be reversed by a mutation, in which we flipped the charges on the *e* and *g* residues of helix C through two local swaps: R57E-E59R and R64E-E66R (R57E means mutating Arg57 to Glu57). We also made three local swaps (A5E-E6A, E54A-A54E, and R71N-N71R) to help stabilize the left-handed conformation. The mutant is referred to as α_3D_L. A sequence comparison with α_3D is shown in the middle of Fig. 3.

We simulated the mutant α_3D_L in three trajectories (Figs. S2*B* and S3). As shown in Fig. 3*B*, the left-handed conformation became dominant with a signature peak at $D = -9.7$ Å, while the right-handed one almost vanished. Cluster analyses confirmed that the left-handed conformation occupied the largest cluster in 300 K.

We further investigated the local stability of the mutant α_3D_L in comparison with that of the original α_3D. The cluster center from trajectory 1 was extracted and energy minimized as the putative native conformation of α_3D_L, as illustrated in Fig. S4*A*. We then ran a separate 300 K regular MD simulation starting from this point for local properties (the local chiral distribution was shown as the shaded peak in Fig. 3*B*). The average RMSD among all trajectory frames from the native simulation was 2.0 Å,

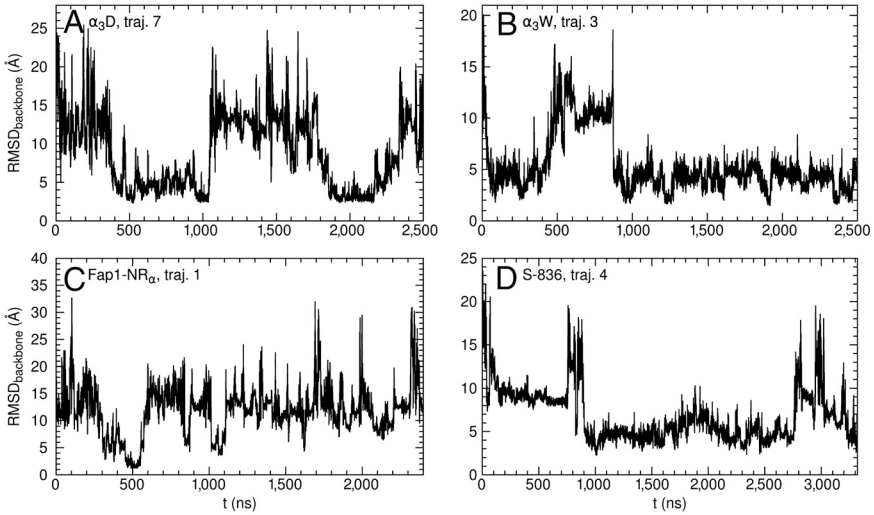

Fig. 2. The backbone root-mean-square deviation (RMSD) from the experimentally determined structure versus time in a typical simulation trajectory of (*A*) α_3D, (*B*) α_3W, (*C*) Fap1-NR$_\alpha$, (*D*) S-836.

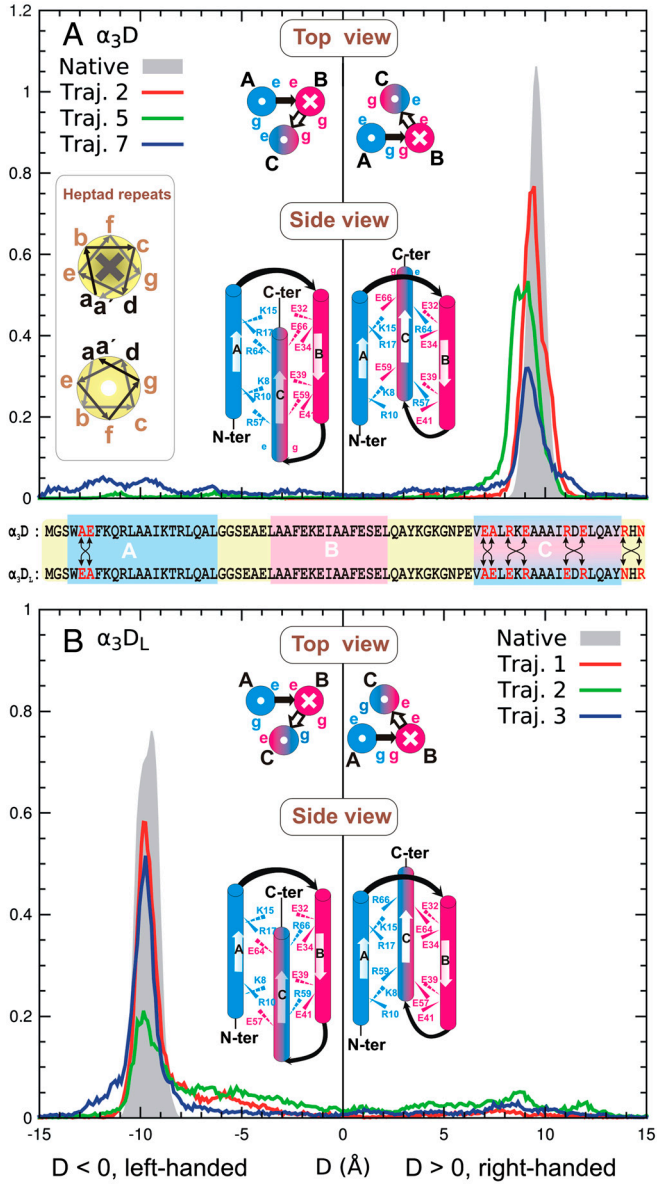

Fig. 3. The packing-chirality D distributions at 300 K of (*A*) α_3D and a mutant (*B*) α_3D_L. In each panel, the shaded peak was obtained from a 300 K regular MD simulation of the respective native conformation. For each protein, interaction patterns of charged residues are also shown schematically in top and side views, in which the right and left sides represent the right-handed and left-handed conformations, respectively. For α_3D (*A*), only the right-handed (native) conformation exists; the left-handed conformation is hypothetical. But the opposite is true for the mutant α_3D_L (*B*). *Inset* of *A*: helix heptad. In the top views and heptad illustration, crosses and dots in circles mean pointing into and out of the paper, respectively. The sequences of the two proteins are compared in the middle of the figure.

which was comparable to that of α_3D, 2.1 Å, suggesting a similar conformational stability. Further, as shown in Table S4, the hydrophobic contact number, interhelical distances, and angles were similar to the native conformation of α_3D. Finally, the side chain χ_1 ($C\text{-}C_\alpha\text{-}C_\beta\text{-}C_\gamma$) dihedrals of α_3D and α_3D_L had similar rotameric distributions, as shown in Fig. S5 *A* and *B*, as well as similar rotameric entropy of the hydrophobic core, as shown in Table S5. Thus, the mutations might not affect the stability of the hydrophobic core significantly, and the charge interaction should be responsible for the chirality change.

The comparison of α_3D and α_3D_L also suggested that the stability of the native right-handed conformation of α_3D was limited by the energy of forming a few favorable pairs of charged residues. Other energy contributions, such as hydrophobic interactions, were unlikely to inhibit the chirality flip.

α_3W. We now turn to the second protein α_3W (PDB ID: 1LQ7) (22, 23), which was previously simulated in implicit solvent (29–31). However, our trajectories revealed in addition to the native conformation two compact, geometrically different (albeit less populated) conformations.

Like α_3D, α_3W is a parallel bundle of three helices, again labeled as A, B, and C from N to C terminus. Its strategy to secure the chirality is similar to that of α_3D (22): the heptad pattern is also strictly followed in all three helices; the *e* and *g* residues on helix B are positively charged; those on helix C are negatively charged; and the *e* and *g* residues on helix A are negatively and positively charged, respectively. As shown in Fig. 4*A*, *Upper, Inset*, the arrangement now favors the left-handed conformation.

The chiral distribution computed from the folding trajectories indeed showed a prominent left-handed ($D < 0$) peak as well as a suppressed right-handed ($D > 0$) one, as seen in Fig. 4*A*. In fact, the right-handed peak covered two different conformations, both of which had corresponding clusters under cluster analyses (300 K, cluster centers also shown in Fig. 4*A*). We ran respective 300 K MD simulations starting from the two nonnative conformations as well as the native one. The chiral distributions of the three conformations are shown as shaded peaks in Fig. 4*A*.

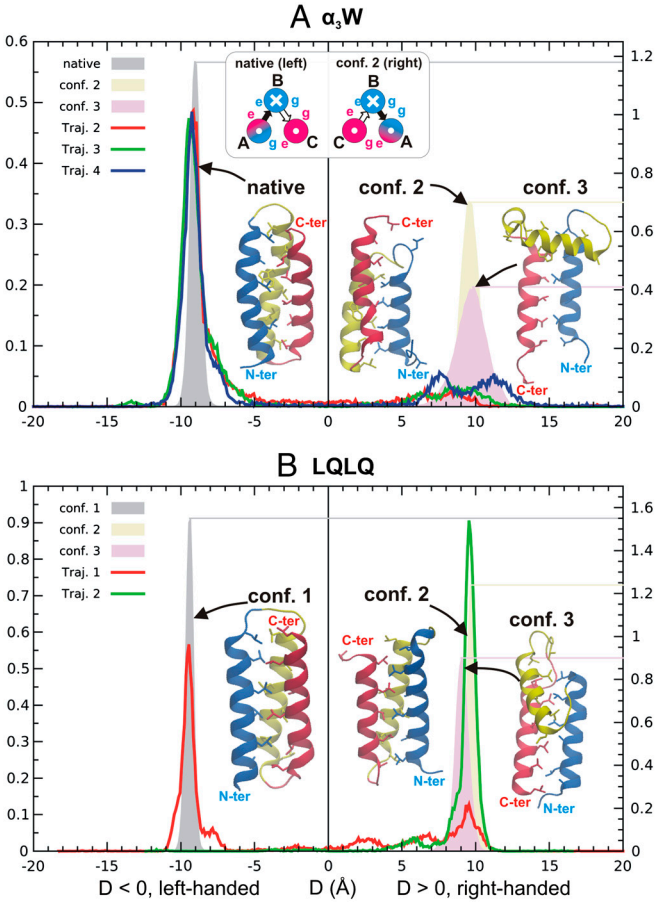

Fig. 4. The packing-chirality D distributions at 300 K of (*A*) α_3W and (*B*) a mutant LQLQ with a symmetric sequence. The native and two nonnative conformations are also shown in either case. The colors of the three helices are blue, yellow, and red, sequentially, from N to C terminus. The shade peaks (right vertical scales) in both panels were obtained from 300 K regular MD simulations starting from the representative structures.*A*, *Upper Inset* for charge interactions is similar to that in Fig. 3*A* or *B*.

The first nonnative conformation (conf. 2) had the opposite chirality to the native one. Its metastability was achieved by sliding two helices along the parallel axis, which somewhat relieved unfavorable charge interactions (see Fig. S6). The conformation had similar interhelical distances and angles as the native one, but a lowered contact number [from 49 (native) to 40] among hydrophobic residues, as shown in Table S4.

The second nonnative conformation (conf. 3) differed from both the native one and conf. 2 by a reversely oriented helix C (see Fig. 4*A*). It was stabilized through a few interhelical hydrophobic and charged interactions, as shown in Fig. S7*A*.

We further demonstrate that if all charged residues on the helices were turned to polar or neutral ones, the protein could indeed switch among conformations of different geometries. We simulated a highly symmetric three-helix bundle, named LQLQ, in which all *a* and *d* residues were Leus and all *b*, *c*, *e*, *f*, and *g* residues were Gln in two independent trajectories. Cluster analyses revealed three clusters, which we refer to as conf. 1, 2, and 3, as shown in Fig. 4*B*.

Due to the symmetric sequence, conf. 1 and 2 (left- and right-handed, respectively) were similar in all geometric measures except the opposite chirality, as shown in Table S4. They represented the two most populated clusters with similar populations. Conf. 3 (a smaller cluster) was similar to, but more compact than, the conf. 3 in α_3W, with the middle helix B being almost parallel, instead of orthogonal, to the other two helices. Note helix B was closer to helix C, not A, although the two choices appeared to be symmetric; we propose a possible explanation in Fig. S7*B*.

Additionally, we simulated another mutant LALQ in which the *b* and *c* residues were Alas instead Gln with similar results (Fig. S8), suggesting some generality of LQLQ in modeling a symmetric three-helix bundle. However if the *e*, *f*, and *g* residues were also replaced Alas, helices in the resulting mutant LALA tended to be bent (see Fig. S9), suggesting that the polar and/or charged residues with long side chains might help maintaining a straight skeleton.

S-836. The last protein is the four-helix bundle S-836 (PDB ID: 2JUA) (25). The four helices are labeled as A, B, C, and D from N to C terminus. In five trajectories (the total simulation time was 17 μs), the first four reached the native conformation, while the last sampled two different nonnative ones.

To describe the geometry of the protein, we first extended the packing chirality for a three-helix bundle to a four-helix one. We view the four helices as two successive three-helix groups, A-B-C and B-C-D, each of which is thus equivalent to a three-helix bundle and has its own chirality. We partition the conformation space to four quadrants according to the signs of the two chiral distances D_{ABC} and D_{BCD}, denoted as LR ($D_{ABC} < 0$, $D_{BCD} > 0$), RR ($D_{ABC} > 0$, $D_{BCD} > 0$), RL ($D_{ABC} > 0$, $D_{BCD} < 0$) and LL ($D_{ABC} < 0$, $D_{BCD} < 0$). A quadrant can also be represented by either the native (LR) conformation or some hypothetic one that differs from the native only by chirality; e.g., the representative RR conformation differs from the native only by the chirality of the A-B-C helix group.

We computed the joint distribution of D_{ABC} and D_{BCD} after combining data from all trajectories. As shown in Fig. 5, the distribution revealed four major clusters. The four corresponding conformations were the native one (LR), conf. 2 (RR) and 3 (RL), and conf. 4, which differed from conf. 3 by a reversely oriented helix A. Although the competing conformations exist, the native one still occupied the most populated cluster in the first four folding trajectories, and the population ratios to the first runner-up clusters were 2.1, 3.8, 1.1, and 1.9, respectively.

Except conf. 4 (which is not a simple chiral image of the native), the chiral conformations could be sorted as LR (native) >RR and RL (conf. 2 and 3) >LL (missing) by their populations. We attempt to explain the order by the pairing pattern of charged residues, as shown in Fig. S10. The missing LL conformation was possibly due to inescapable repulsion between His83, His86, Lys90, His94 on helix D, and Lys32, His36 on helix B (all His are positively charged). Helices A and D in the RL conformation encountered a similar but less severe problem, causing a kink in helix D to avoid direct repulsion between His86 and Lys17. Unfavorable interactions between helices B and D in the RR conformation could bend or dislocate helix D. The LR (native) conformation appeared to have the best pairing pattern, although unfavorable contacts exist; e.g., Glu91 and Asp11.

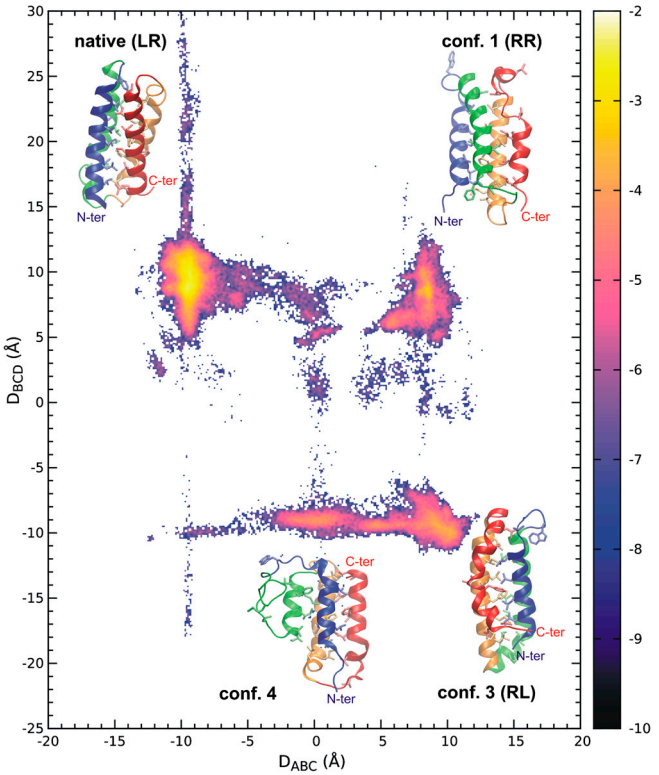

Fig. 5. The logarithm of the packing-chirality distribution $\log \rho(D_{ABC}, D_{BCD})$ of S-386 at 300 K with representative conformations at the four most populated clusters. The colors of the four helices are blue, green, orange, and red, sequentially, from N to C terminus.

In nature, four-helix bundles adopting the RR, RL, and LL conformations do exist. In fact, the majority [e.g., cytochrome ć (PDB ID: 3ZTM) (32) and cytochrome b_{562} (PDB ID: 1QPU) (33)] adopt the RL conformation, the opposite of the native LR conformation of S-836. Two examples of the RR and LL conformations are the C-terminal domain of KaiA (PDB ID: 1Q6A) (34) and domain I of the transcription factor TFIIS (PDB ID: 1EO0) (35), respectively.

Additional results of the simulated proteins are discussed in the *SI Text* and Figs. S11–S17.

Conclusions

We have shown the usefulness of the single-trajectory-based tempering with a dihedral bias in studying helical proteins in explicit solvent. We were able to fold, within a few microseconds, four helical proteins (with the native conformation usually found in the largest cluster in simulation trajectories) as well as their mutants.

Our trajectories showed that a parallel helix bundle, especially one with a symmetric sequence, often has competing conformations that differ from the native one only by packing chirality and/or helix orientation. Similarly, the existence of competing conformations was previously reported for the Rop-dimer and its mutants (36–38). While the ability to identify the native geometry is essential in studying helical proteins, the competing conforma-

tions can help to design new proteins, either as targets or as pitfalls to avoid (39).

Despite continuous improvement, the current force field might have yet reached perfection. As suggested by Best and Hummer (40), the current AMBER force field still underestimates both the enthalpy of hydrogen bonds (rendering helical conformations less energetically favorable) and the entropy of unfolded states. Our sampling strategy, which used a helical bias and a high-temperature motion, could hopefully alleviate both problems.

Methods

We describe the principles of computational methods below; details are supplied in the *SI Text*. The source code can be downloaded from http://sigler.bioch.bcm.tmc.edu/MaLab/.

Dihedral-biased Tempering. Tempering is a technique (12, 41–48) that accelerates sampling by occasionally raising and lowering the temperature. In this study, we adopted a single-trajectory-based tempering method reported previously (12) for computing efficiency. We then added a dihedral bias on top of it to facilitate simulations of helical proteins.

In the tempering method, temperature T is a continuous variable in (T_L, T_H) covering the room temperature T_0. Here, T_L is slightly below T_0, while $T_H (T_H > T_0)$ accelerates the system motion. Tempering preserves the Boltzmann distribution at any temperature cross-section such that the corresponding generalized ensemble is simply a superposition of canonical ensembles of the temperature range.

We now introduce a bias potential V to the generalized ensemble through a temperature-dependent Hamiltonian; i.e., the energy function, $H(T) = H_0 + (T/T_0 - 1)V$, where H_0 is the unbiased Hamiltonian. Since V has the maximal magnitude at T_H, but is zero at the room temperature T_0, the new Hamiltonian is reduced to the original one at $T = T_0$, and unbiased room temperature T_0 properties can thus be recovered through a modification of the multiple histogram method (26–28).

We let the bias potential V target the two backbone dihedral angles [i.e., φ (C-N-C_α-C) and ψ(N-C_α-C-N)], as they critically influence protein dynamics (10, 18, 19). It is parameterized in terms of three intrinsic dihedral modes (see Fig. S18 and Table S6), one of which corresponds to the helical conformation. The magnitudes of the modes are adjustable in each trajectory. The bias potential V applies uniformly to all except glycine and proline residues.

The method was coded to GROMACS 4.0.7 (49). We used a modified version (40) of AMBER03 (50, 51) as the force field (the modification slightly reduced helical content), TIP3P (52) as the water model, the particle-meshed Ewald (PME) method (53) for long-range electrostatic interaction, and the velocity-rescaling method for thermostat (54); SETTLE (55) and the parallel LINCS (56) were used as constraint algorithms for waters and proteins, respectively.

The verifications of the methods are shown in *SI Text*; see also Figs. S19 and S20.

Although the method is designed for helical proteins, extensions to other types of proteins might be achieved via a random trial of various bias potentials, as discussed in the *SI Text*. To confirm a folding, we can compare several independent trajectories to see if a common structure is reached.

Clustering Method. We used a deterministic clustering method. The method finds a balance between the intracluster distance (the distance of two protein structures is defined as their backbone RMSD after alignment; it is then averaged over all pairs of frames within the cluster) and the total number of clusters. It also allows each frame to carry the weight from the multiple histogram method (26–28) to filter out frames irrelevant to 300 K.

ACKNOWLEDGMENTS. We thank Fengyun Ni and Dr. Qinghua Wang for helpful discussions. J.M. acknowledges support from National Institutes of Health (Grant R01-GM067801), National Science Foundation (NSF) (Grant MCB-0818353), The Welch Foundation (Grant Q-1512), the Welch Chemistry and Biology Collaborative Grant and the Faculty Initiatives Fund from Rice University. This work was supported in part by the Shared University Grid at Rice funded by NSF under Grant EIA-0216467, a 2010 IBM Shared University Research (SUR) Award on IBM's Power7 high performance cluster (BlueBioU) to Rice University, the Cyberinfrastructure for Computational Research funded by NSF under Grant CNS-0821727, and the NSF through TeraGrid resources provided by Texas Advanced Computing Center under Grant TG-MCB100013. Use of Visual Molecular Dynamics (57) is gratefully acknowledged.

1. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 48:545–600.
2. Daggett V, Fersht A (2003) The present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol* 4:497–502.
3. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding 'speed limit'. *Curr Opin Struct Biol* 14:76–88.
4. Scheraga HA, Khalili M, Liwo A (2007) Protein-folding dynamics: Overview of molecular simulation techniques. *Annu Rev Phys Chem* 58:57–83.
5. Dill KA, Ozkan SB, Weikl TR, Chodera JD, Voelz VA (2007) The protein folding problem: When will it be solved? *Curr Opin Struct Biol* 17:342–346.
6. Duan Y, Kollman PA (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282:740–744.
7. Seibert MM, Patriksson A, Hess B, van der Spoel D (2005) Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *J Mol Biol* 354:173–183.
8. Ensign DL, Kasson PM, Pande VS (2007) Heterogeneity even at the speed limit of folding: Large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. *J Mol Biol* 374:806–816.
9. Juraszek J, Bolhuis PG (2008) Rate constant and reaction coordinate of Trp-cage folding in explicit water. *Biophys J* 95:4246–4257.
10. 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.
11. Freddolino PL, Schulten K (2009) Common structural transitions in explicit-solvent simulations of villin headpiece folding. *Biophys J* 97:2338–2347.
12. Zhang C, Ma J (2010) Enhanced sampling and applications in protein folding in explicit solvent. *J Chem Phys* 132:244101.
13. Shaw DE, et al. (2010) Atomic-level characterization of the structural dynamics of proteins. *Science* 330:341–346.
14. Best RB, Mittal J (2010) Balance between alpha and beta structures in ab initio protein folding. *J Phys Chem B* 114:8790–8798.
15. Day R, Paschek D, Garcia AE (2010) Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins* 78:1889–1899.
16. Ikebe J, Standley DM, Nakamura H, Higo J (2011) Ab initio simulation of a 57-residue protein in explicit solvent reproduces the native conformation in the lowest free-energy cluster. *Protein Sci* 20:187–196.
17. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334:517–520.
18. Zhou Y, Karplus M (1999) Folding of a model three-helix bundle protein: A thermodynamic and kinetic analysis. *J Mol Biol* 293:917–951.
19. Jang S, Shin S, Pak Y (2003) Replica-exchange method using the generalized effective potential. *Phys Rev Lett* 91:058305.
20. Bryson JW, Desjarlais JR, Handel TM, DeGrado WF (1998) From coiled coils to small globular proteins: Design of a native-like three-helix bundle. *Protein Sci* 7:1404–1414.
21. 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.
22. Johansson JS, Gibney BR, Skalicky JJ, Wand AJ, Dutton PL (1998) A native-like three-alpha-helix bundle protein from structure-based redesign: A novel maquette scaffold. *J Am Chem Soc* 120:3881–3886.
23. Dai QH, et al. (2002) Structure of a de novo designed protein model of radical enzymes. *J Am Chem Soc* 124:10952–10953.
24. Ramboarina S, et al. (2010) Structural insights into serine-rich fimbriae from Gram-positive bacteria. *J Biol Chem* 285:32446–32457.
25. Go A, Kim S, Baum J, Hecht MH (2008) Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. *Protein Sci* 17:821–832.
26. Ferrenberg AM, Swendsen RH (1988) New Monte-Carlo technique for studying phase-transitions. *Phys Rev Lett* 61:2635–2638.
27. Ferrenberg AM, Swendsen RH (1989) Optimized Monte-Carlo data-analysis. *Phys Rev Lett* 63:1195–1198.
28. 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.
29. Yang JS, Chen WW, Skolnick J, Shakhnovich EI (2007) All-atom ab initio folding of a diverse set of proteins. *Structure* 15:53–63.
30. Meinke JH, Hansmann UH (2009) Free-energy-driven folding and thermodynamics of the 67-residue protein GS-alpha3W—a large-scale Monte Carlo study. *J Comput Chem* 30:1642–1648.
31. Czaplewski C, Kalinowski S, Liwo A, Scheraga HA (2009) Application of multiplexed replica exchange molecular dynamics to the UNRES force field: Tests with alpha and alpha+beta proteins. *J Chem Theory Comput* 5:627–640.
32. Antonyuk SV, et al. (2011) Carbon monoxide poisoning is prevented by the energy costs of conformational changes in gas-binding haemproteins. *Proc Natl Acad Sci USA* 108:15780–15785.
33. Arnesano F, et al. (1999) The solution structure of oxidized Escherichia coli cytochrome b562. *Biochemistry* 38:8657–8670.
34. Vakonakis I, Langenhan T, Promel S, Russ A, Campbell ID (2008) Solution structure and sugar-binding mechanism of mouse latrophilin-1 RBL: A 7TM receptor-attached lectin-like domain. *Structure* 16:944–953.
35. Booth V, Koth CM, Edwards AM, Arrowsmith CH (2000) Structure of a conserved domain common to the transcription factors TFIIS, elongin A, and CRSP70. *J Biol Chem* 275:31266–31268.

36. Levy Y, Cho SS, Shen T, Onuchic JN, Wolynes PG (2005) Symmetry and frustration in protein energy landscapes: A near degeneracy resolves the Rop dimer-folding mystery. *Proc Natl Acad Sci USA* 102:2373–2378.
37. Schug A, Whitford PC, Levy Y, Onuchic JN (2007) Mutations as trapdoors to two competing native conformations of the Rop-dimer. *Proc Natl Acad Sci USA* 104:17674–17679.
38. Gambin Y, et al. (2009) Direct single-molecule observation of a protein living in two opposed native structures. *Proc Natl Acad Sci USA* 106:10153–10158.
39. Baltzer L, Nilsson H, Nilsson J (2001) De novo design of proteins—what are the rules? *Chem Rev* 101:3153–3163.
40. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* 113:9004–9015.
41. Lyubartsev AP, Martsinovski AA, Shevkunov SV, Vorontsovvelyaminov PN (1992) New approach to Monte-Carlo calculation of the free-energy—method of expanded ensembles. *J Chem Phys* 96:1776–1783.
42. Marinari E, Parisi G (1992) Simulated tempering—a new Monte-Carlo scheme. *Europhys Lett* 19:451–458.
43. Zhang C, Ma J (2007) Simulation via direct computation of partition functions. *Phys Rev E* 76:036708.
44. Li H, Fajer M, Yang W (2007) Simulated scaling method for localized enhanced sampling and simultaneous "alchemical" free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations. *J Chem Phys* 126:024106.
45. Geyer CJ (1991) *Proceedings of the 23rd Symposium on the Interface* (Am Statistical Association, New York).
46. Hansmann UHE (1997) Parallel tempering algorithm for conformational studies of biological molecules. *Chem Phys Lett* 281:140–150.
47. Hukushima K, Nemoto K (1996) Exchange Monte Carlo method and application to spin glass simulations. *J Physical Soc Japan* 65:1604–1608.
48. Swendsen RH, Wang JS (1986) Replica Monte-Carlo simulation of spin-glasses. *Phys Rev Lett* 57:2607–2609.
49. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447.
50. Duan Y, et al. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24:1999–2012.
51. Sorin EJ, Pande VS (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys J* 88:2472–2493.
52. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
53. Essmann U, et al. (1995) A smooth particle messh Ewald method. *J Chem Phys* 103:8577–8593.
54. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101.
55. Miyamoto S, Kollman PA (1992) SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J Comput Chem* 13:952–962.
56. Hess B (2007) P-LINCS: A parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* 4:116–142.
57. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14:33–38.