

De novo Folding of Two-Helix Potassium Channel Blockers with Free-Energy Models and Molecular Dynamics

Aina Quintilla, Evgeni Starikov, and Wolfgang Wenzel*

*Forschungszentrum Karlsruhe, Institute für Nanotechnologie, P.O. Box 3640,
76021 Karlsruhe, Germany*

Abstract: We report the predictive de novo folding of three two-helix proteins using the free-energy protein forcefield PFF01. Starting from random initial conformations 40–90% of the members of the simulated ensembles converge to near-native conformations. The energetically lowest conformations approach the conserved part of the native conformations to within 1.64, 1.86, and 1.84 Å for 1WQC, 1WQD, and 1WQE, respectively. An analysis of the low-lying conformations predicts the correct topology of the disulfide bridges, which are formed in additional simulations with a constraining potential. The free energy landscapes of these proteins are very simple, suggesting them as candidates for all-atom molecular dynamics simulations. In five independent simulations we find the formation of the correct secondary structure and several folding events into the tertiary structure.

1. Introduction

The development of methods that simulate the protein folding process at the all-atom level and the elucidation of its mechanism are among the important outstanding problems of biophysical chemistry.^{1–4} Significant new insights arise from de novo folding studies of small proteins and peptides.^{5–8} We have recently demonstrated a feasible strategy for all-atom protein structure prediction^{9–11} in a minimal thermodynamic approach. We developed an all-atom free-energy forcefield for proteins (PFF01), which is primarily based on physical interactions.¹¹ We already demonstrated the reproducible and predictive folding of several proteins with 20–60 amino acids.^{9,12–14}

While many β -hairpin systems have been investigated both experimentally and theoretically,^{15–17} there is a relative scarcity of small two-helix peptides that are known to fold experimentally into well-defined tertiary structure. Since two-helix proteins constitute a minimal model, in which to investigate the interplay of hydrophobic collapse, secondary structure formation and the formation of native contacts, the identification of such systems may be helpful to elucidate

the protein folding mechanism. We have previously folded the widely studied 23 amino acid trp-cage protein,^{6,12,18,19,20–22} which has spurred many theoretical investigations because of its fast folding time.

In this investigation we fold three homologous potassium channel blockers,²³ which exhibit a parallel two-helix bundle. We find that all three peptides fold reproducibly into stable tertiary structures, with very simple free-energy funnels. We demonstrate through molecular dynamics simulation that the lack of competing metastable conformations makes these proteins ideal candidates for folding studies to elucidate the interplay of secondary and tertiary structure formation.^{15,24,25}

2. Methods

2.1. Forcefield. We have recently developed an all-atom (with the exception of apolar CH_n groups) free-energy protein forcefield (PFF01) that models the low-energy conformations of proteins with minimal computational demand.^{26,10,11} The forcefield parametrizes the internal free energy of the protein (excluding backbone entropy) and contains the following nonbonded interactions:

* Corresponding author phone: +49-7247-82-6383. E-mail: wenzel@int.fzk.de; Web: <http://www.fzk.de/biostruct>.

$$V(\{\vec{r}_i\}) = \sum_{ij} V_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^6 \right] + \sum_{ij} \frac{q_i q_j}{\epsilon_{g(i)g(j)} r_{ij}} + \sum_i \sigma_i A_i + \sum_{\text{hbonds}} V_{\text{hb}} \quad (1)$$

Here r_{ij} denotes the distance between atoms i and j and $g(i)$ the type of the amino acid i .

The Lennard-Jones parameters (V_{ij} , R_{ij} for potential depths and equilibrium distance) depend on the type of the atom pair and were adjusted to satisfy constraints derived from a set of 138 proteins of the PDB database.^{26–28} The nontrivial electrostatic interactions in proteins are represented via group-specific dielectric constants ($\epsilon_{g(i)g(j)}$ depending on the amino acids to which the atoms i and j belong). The partial charges q_i and the dielectric constants were derived in a potential-of-mean-force approach.²⁹ Interactions with the solvent were first fit in a minimal solvent accessible surface model³⁰ parametrized by free energies per unit area σ_i to reproduce the enthalpies of solvation of the Gly-X-Gly family of peptides.³¹ A_i corresponds to the area of atom i that is in contact with a fictitious solvent. Hydrogen bonds are described via dipole–dipole interactions included in the electrostatic terms and an additional short-range term for backbone–backbone hydrogen bonding (CO to NH) which depends on the OH distance, the angle between N, H, and O along the bond, and the angle between the CO and NH axis.¹¹

In the folding process under physiological conditions the degrees of freedom of a peptide are confined to rotations about single bonds. In our simulation we therefore consider only moves around the side-chain and backbone dihedral angles, which are attempted with 30% and 70% probability, respectively. The moves for the side-chain angles are drawn from an equidistributed interval with a maximal change of 5 degrees. Half of the backbone moves are generated in the same fashion; the remainder is generated from a move library that was designed to reflect the natural amino acid dependent bias toward the formation of α -helices or β -sheets. The probability distribution of the move library was fitted to experimental probabilities observed in the PDB database.³² While driving the simulation toward the formation of secondary structure, the move library introduces no bias toward helical or sheet structures beyond that encountered in nature.

2.2. Optimization Methods. The low-energy part of the free energy landscape of proteins is extremely rugged due to the comparatively close packing of the atoms in the collapsed ensemble. Rugged potential energy surfaces are characterized by the existence of many low-lying minima, which are separated by high-energy barriers. For this reason, the global optimum of such a surface is difficult to obtain computationally. Simple methods, such as steepest descent or simulated annealing, are almost always trapped in metastable conformations.

Efficient optimization methods must therefore speed up the simulation by avoiding high-energy transition states, by adapting large scale moves wherever possible, or by accepting unphysical intermediates. One of the simplest ideas to effectively eliminate high-energy transition states of the free-

energy surface is the basis of the basin hopping technique,^{33,34} also known as Monte Carlo with minimization. This method simplifies the original potential energy surface by replacing the energy of each conformation with the energy of a nearby local minimum. This replacement eliminates the high-energy barriers that are responsible for the freezing problem in simulated annealing. In many cases the additional minimization effort to find a local minimum for each starting configuration is more than compensated by the increase of efficiency of the stochastic search on the simplified potential energy surface. The basin hopping technique and derivatives²⁷ have been used previously to study the potential energy surface of model proteins³⁵ and all atom protein models.^{22,36–38}

In contrast to previous work, we use a simulated annealing process³⁹ for the minimization step. The temperature is decreased geometrically from its starting to the final value of $T_f = 2\text{K}$. Following an optimized protocol³⁸ the starting temperature T_s is drawn randomly from a distribution $p(T_s) \sim \exp(T_s/T_0)$. The performance of the method is only weakly dependent on the choice of T_0 , which was chosen as 750 K for all simulations reported here.

While each simulated annealing run is typically much more expensive than a local minimization using gradient based techniques, it can nevertheless be competitive for very rugged potential energy surfaces, or when the computation of the gradient of the potential is prohibitive. The number of moves in each individual basin hopping cycle, is increased with the square root of the cycle number m as $N = 10\,000 \times \sqrt{m}$.

At the end of one annealing step the new conformation was accepted if its energy difference to the current configuration was no higher than a given threshold energy ϵ_T , an approach proven optimal for certain optimization problems.⁴⁰ Throughout this study we use a threshold acceptance criterion of 1 kcal/mol.

2.3. Molecular Dynamics. Starting from the same unfolded conformation as above, we performed all-atom implicit water molecular dynamics simulation using the AMBER8 simulation package⁴¹ with the AMBER99 force-field⁴² using the Born/SASA solvation model.^{43–45,8} The simulation was performed at the linux cluster of the KIST supercomputational materials lab with up to 16 processors in parallel. The system was first minimized by steepest descent. We generated five trajectories with 50 ns total simulation time each, three at 300 K and two at 325 K. After minimization the simulations are heated independently to their final temperature. Simulations were performed using Langevin temperature coupling and electrostatic interactions without cutoff.

3. Results

3.1. Free-Energy Folding Simulations. Structures for the peptides 1WQC, 1WQD, and 1WQE with 26, 27, and 23 amino acids, respectively, were retrieved from the PDB database⁴⁶ and unfolded by setting all backbone dihedral angles to random values until nonclashing conformations were obtained. The starting conformations had backbone root-mean-square deviations (RMSB) of 11.8 Å, 7.8 Å, and 9.7 Å to the native conformations of 1WQC, 1WQD, and

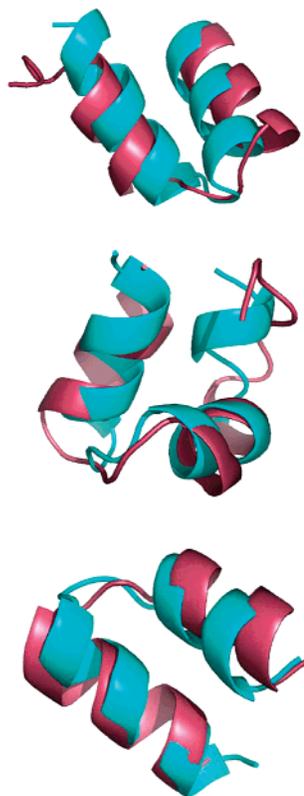


Figure 1. Overlay of the experimental (in red, first model) and the folded conformations (in blue) of 1WQC, 1WQD, and 1WQE, respectively.

Table 1. Length, Sequence, and Disulfide-Bridge Topology (Numbers of the Amino Acids) for the Peptides 1WQC, 1WQD, and 1WQE

name	#AA	sequence	DS1	DS2
1WQC	26	DPCYEVLQQHGNVKECEEACKHPVE	3/21	7/17
1WQD	27	DPCYEVLQQHGNVKECEEACKHPVEY	3/21	7/17
1WQE	23	NDPCEEVCIQHTGDVKACEEACQ	4/22	8/18

1WQE, respectively—they had no secondary structure. Table 1 shows the sequences of the peptides, which have a very high degree of homology. We note that all three peptides are stabilized by two disulfide bridges, as indicated in the table.

For each of the peptide we performed 20 independent basin hopping simulations with 200 cycles each using the protocol described above. In order to avoid any bias toward the native conformation, there was no potential representing the disulfide bridges in these simulations. Figure 2 shows the convergence of the energy and the RMSB deviation for 1WQC vs the number of function evaluations as a representative example.

Tables 2–4 summarize the energies, RMSB deviations, and secondary structure for the final population of these simulations. There are 30 NMR models for each of the peptides, which differ in the unstructured tail-fragments after amino acid 20 of the sequence. The table therefore reports the RMSB deviation to the closest model and the RMSB deviation to the structurally conserved part of each peptide.

All three proteins were folded predictively to very good resolution. Predictive folding is achieved, when near-native

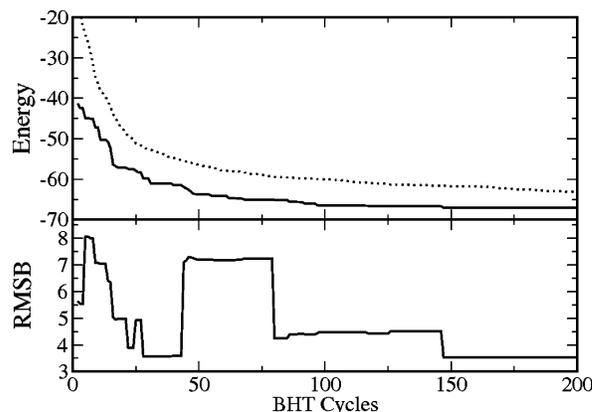


Figure 2. Convergence of best and mean energy (top panel, in kcal/mol) and RMSB (bottom panel, in Å) as a function of the number of basin hopping cycles for the simulations of 1WQC. Note that early in the simulation, one of the non-native conformations was lowest in energy, before it was overtaken by a near-native conformation. For simplicity the RMSB deviation was measured against the first, not the best NMR model.

Table 2. Final Population of Decoys of the Basin Hopping Simulations for 1WQC^a

RMSB	RMSB _{1–20}	energy	secondary structure
2.90	1.64	−67.10	CHHHHHHHHHHTCHHHHHHHHHHHC
3.21	1.67	−66.80	CHHHHHHHHHHTCHHHHHHHHTTTC
2.60	1.95	−65.40	CHHHHHHHHHHTCHHHHHHHHSCCC
6.31	4.07	−65.40	CHHHHHHHHHHTSCSHHHHHHHHSCCC
3.85	2.01	−65.00	CHHHHHHHHHHTSCHHHHHHHHTSSCC
2.55	1.92	−64.70	CHHHHHHHHHHTSHHHHHHHHSCCC
3.08	1.52	−64.40	CHHHHHHHHHHTCCHHHHHHHHHHHC
3.98	2.32	−64.20	CHHHHHHHHHHTSCHHHHHHHHTTTC
3.34	1.74	−63.90	CHHHHHHHHHHTSCHHHHHHHHHHHC
4.83	2.15	−63.40	CHHHHHHHHHHTSCHHHHHHHHTSCCC
6.45	4.57	−62.80	CHHHHHHHHHHTCCHHHHHHHHHHHC
6.38	4.45	−62.40	CHHHHHHHHHHTCCHHHHHHHHHHHC
6.34	4.41	−62.40	CHHHHHHHHHHTCCHHHHHHHHHHHC
6.30	4.42	−62.30	CHHHHHHHHHHTCCHHHHHHHHHHHC
2.45	1.67	−62.10	CHHHHHHHHHHTCCHHHHHHHHSCCC
3.74	2.81	−61.70	CHHHHHHHHHHTCCCCHHHHHHSCCC
3.26	1.75	−60.80	CHHHHHHHHHHSSHHHHHHHHCSSCC
5.01	2.85	−59.70	CHHHHHHHHHCHHHHHHHHHHTSSCC
3.14	1.92	−59.60	CHHHHHHHHSSHCHHHHHHHHHHHC
6.19	4.82	−58.30	CHHHHHHHSCTTCHHHHHHHHSCCC

^a We computed the minimal RMSB deviation (in Å) to the 30 experimental models of the full protein and of amino acids 1–20, respectively. The secondary structure was computed with DSSP:⁶⁷ H, T, S, and C designate helix, turn, strand, and coil conformations, respectively.

structures dominate the low-energy spectrum of the simulated ensemble. In a free-energy forcefield the native conformation is selected on the basis of its estimate of internal free-energy in comparison to other conformations with well-defined secondary and tertiary structure. This is in contrast to MD or REM investigations, where occupation probability determines the thermodynamically stable conformation. Thus finding a particular conformation repeatedly with the lowest energy, as was observed for ALL proteins studied here, predicts the native conformation. Not in all simulations that reach the native conformation all stabilizing tertiary interac-

Table 3. Final Population of Decoys of the Basin Hopping Simulations for 1WQD^a

RMSB	RMSB ₁₋₂₀	energy	secondary structure
3.23	1.86	-69.30	CHHHHHHHHHHSCHHHHHHHTCHHHHC
2.82	1.91	-68.70	CHHHHHHHHHHSSHHHHHHTCHHHHC
3.87	1.93	-68.70	CHHHHHHHHHHTCHHHHHHHHHHHHC
2.92	2.08	-68.50	CHHHHHHHHHHSSHHHHHHTCHHHHC
4.89	1.72	-68.40	CHHHHHHHHHHTSHHHHHHHHCTTTC
3.43	3.32	-68.40	CCSHHHHHHHHTSCHHHHHHHHTCC
3.93	1.93	-68.00	CHHHHHHHHHHTCHHHHHHHHHHHHC
3.27	1.99	-67.70	CHHHHHHHHHHTSCHHHHHHHHTHHHC
3.30	2.51	-66.10	CHHHHHHHHHHCHHHHHHHHTHHHC
4.26	4.05	-66.10	CHHHHHHHHTTTTCHHHHHHHHHHHHC
3.63	3.67	-65.50	CHHHHHHHHHHCHHHHHHHHTTCCC
3.67	3.68	-65.50	CHHHHHHHHHHCHHHHHHHHTTCCC
2.96	2.79	-65.50	CCCHHHHHHHHTSCHHHHHHHHTTSC
3.67	3.72	-64.70	CHHHHHHHHHHCHHHHHHHHTTCCC
4.05	1.77	-63.90	CHHHHHHHHHHTCSSHHHHHHHHHHHC
4.61	4.60	-63.50	CHHHHHHHHHHTCCHHHHHHHHHHHHC
3.27	2.09	-63.40	CHHHHHHHHHHTSCHHHHHHSSHHHHC
3.67	3.70	-63.30	CHHHHHHHHHHTTCHHHHHHHHTCCC
5.16	2.97	-62.60	CHHHHHHHHHHSCHHHHHHHCTTTC
3.56	2.29	-60.20	CHHHHHHHHTTCHHHHHHHHHHHHC
4.01	3.91	-60.20	CHHHHHHHCSSSCHHHHHHHHTTCCC

^aWe computed the minimal RMSB deviation (in Å) to the 30 experimental models of the full protein and of amino acids 1-20, respectively. The secondary structure was computed with DSSP:⁶⁷ H, T, S, and C designate helix, turn, strand, and coil conformations, respectively.

Table 4. Final Population of Decoys of the Basin Hopping Simulations for 1WQE^a

RMSB	RMSB ₁₋₂₀	energy	secondary structure
1.90	1.84	-57.00	CHHHHHHHHHHTCHHHHHHHHC
1.64	1.63	-56.90	CHHHHHHHHHHTCHHHHHHHHC
1.66	1.66	-56.90	CHHHHHHHHHHTCHHHHHHHHC
1.70	1.67	-56.80	CHHHHHHHHHHTSHHHHHHHHC
1.70	1.69	-56.50	CHHHHHHHHHHTCCHHHHHHHHC
1.68	1.67	-56.50	CHHHHHHHHHHTCCHHHHHHHHC
2.13	2.10	-56.50	CHHHHHHHHHHTCCHHHHHHHHC
1.74	1.72	-56.50	CHHHHHHHHHHTCCHHHHHHHHC
1.73	1.70	-56.40	CHHHHHHHHHHTCCHHHHHHHHC
1.72	1.70	-56.40	CHHHHHHHHHHTCCHHHHHHHHC
1.68	1.67	-56.30	CHHHHHHHHHHTCCHHHHHHHHC
1.74	1.71	-56.30	CHHHHHHHHHHTCCHHHHHHHHC
1.69	1.66	-56.30	CHHHHHHHHHHTCCHHHHHHHHC
2.12	2.14	-56.20	CHHHHHHHHHHTCCHHHHHHHHC
1.69	1.66	-56.20	CHHHHHHHHHHTCCHHHHHHHHC
1.69	1.67	-56.10	CHHHHHHHHHHTCCHHHHHHHHC
1.71	1.68	-56.10	CHHHHHHHHHHTCCHHHHHHHHC
2.18	2.14	-55.00	CHHHHHHHHHHCSCHHHHHHHHHC
2.04	2.02	-54.70	CHHHHHHHHHHCSCHHHHHHHHHC
5.73	4.54	-54.40	CHHHHHHHHTCCSCHHHHHHHHC
4.71	3.84	-53.50	CHHHHHHHHTTSCHHHHHHHHHC

^aWe computed the minimal RMSB deviation (in Å) to the 30 experimental models of the full protein and of amino acids 1-20, respectively. The secondary structure was computed with DSSP:⁶⁷ H, T, S, and C designate helix, turn, strand, and coil conformations, respectively.

tions are fully formed. As a result there may be many more near-native conformations that are slightly higher in energy. We found the lowest 3, 9, and 18 of 20 simulations to converge to near-native conformations of 1WQC, 1WQD,

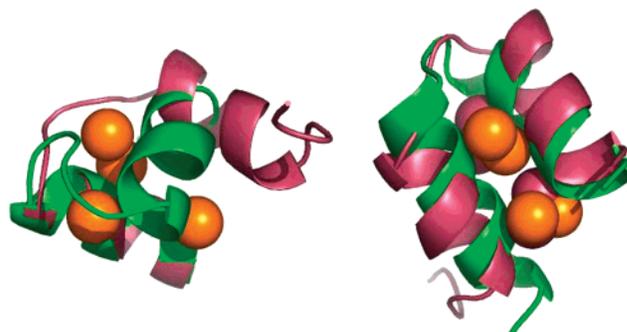


Figure 3. Left panel: Misfolded conformation (green) of 1WQC, corresponding to the fourth decoy in Table 2. Right: Folded conformation (green) of 1WQC in the presence of disulfide bridges in the simulation. The sulfur atoms are shown in orange, the experimental model in red.

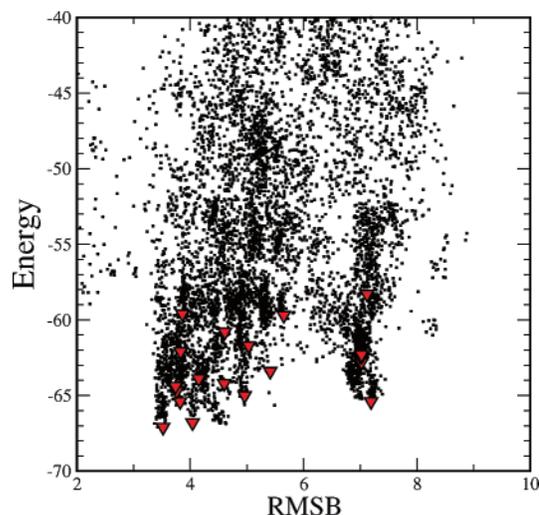


Figure 4. Plot of the energy vs the RMSB in all accepted conformations in the simulations for 1WQC, the triangles show the best conformations of the 20 simulations. There are only two structural clusters in the free-energy landscape with characteristic RMSB deviations of around 3 and 6 Å to the experimental model.

and 1WQE, respectively. The data demonstrate that the simulation method is very robust: 50% (1WQC), 40% (1WQD), and 90% (1WQE) of the simulations converge to conformations with RMSB deviations of less than 2 Å to the native conformation.

In Figure 1 we show the overlay of the lowest energy conformations with the respective experimental model. The figure demonstrates the high degree of similarity of the folded and experimental conformations. It is interesting to note that the addition of just one amino acid from 1WQC to 1WQD leads to a noticeable change in structure that is reproduced in the theoretical model.

3.2. Analysis of the Free-Energy Landscape. Next we turn to the surface of the internal free-energy (excluding backbone entropy) of 1WQC as a representative example of the three peptides. Figure 4 shows energy versus RMSB for all accepted configurations at the end of basin hopping cycles (from all simulations). The triangles indicate the terminal configurations of the individual simulations. We clearly see two broad funnels of conformations, which terminate into

low-energy structures with 3.4 Å and about 7.0 Å RMSB deviation to the native conformation, respectively. The configuration corresponding to the non-native funnel is shown in the left panel of Figure 3. This conformation is inconsistent with the formation of the correct number of native disulfide bridges of this peptide. There is only one, very broad folding funnel consistent with the native disulfide bridge topology. For this reason, the proteins studied here may be ideal examples to follow the kinetics of protein folding with molecular dynamics or replica exchange methods.^{47–49}

3.3. Simulations with Disulfide Bridges. Inspection of the final conformation of the simulations reported above suggests correct pairing for the native disulfide bridges, even if the distances between the sulfur atoms are too large in the absence of any constraining potential. We have therefore added a constraining potential, which varied with the square root of the distance between the sulfur atoms in the correct disulfide bridge topology

$$V_{ss} = V_0 \sum \sqrt{(d_i - d_0)}$$

where d_i is the distance between the sulfur atoms in the i th disulfide bridge, and $V_0 = 5$ kcal/mol, $d_0 = 2$ Å was the target distance for all disulfide bridges (typical experimental values are 2.05 Å). The functional form of the potential was chosen to obtain an appreciable force even for small bond mismatches. In the main simulation the disulfide bridge potential was only applied to the correct disulfide bridge pairing. In exploratory simulations we had applied the potential also to other pairings, but for the protein under investigation here these pairings were incommensurate with the helical starting structures formed in the unbiased simulations discussed in the previous section. Application of the incorrect pairings thus either led to conformations in which the sulfur atoms did not approach one another closely or where the helices were destroyed. In both cases the free energy increased significantly compared to the converged structures described below.

Starting from the final population of 1WQC of the preceding section, we performed 50 additional basin hopping cycles using the annealing cycle described in the methods section. The results of the simulations are summarized in Table 5. Find more near-native conformations and the energetic gap between the native and non-native conformations more than doubles from 1.7 kcal/mol to 3.9 kcal/mol. Because most of the structural differences between the native and the non-native structures arise in the unstructured tail, which is not directly affected by the formation of disulfide bridges, the overall RMSB deviation did not improve much.

This example demonstrates that unconstrained simulations can be used to predict the native topology of disulfide bridge formation, which can be later refined in constrained simulations to form ideal disulfide contacts.

3.4. Molecular Dynamics Folding Simulations. The free energy surface of 1WQE, as illustrated in Figure 4, is much more simple than that encountered for other proteins we have investigated so far.^{10,14} However the internal free-energy estimate does not contain backbone entropy; stabilization of

Table 5. Final Population of Decoys of the Basin Hopping Simulations for 1WQC with the Constraining Potential^a

RMSB	RMSB _{1–20}	energy	secondary structure
2.94	1.59	–56.50	CHHHHHHHHHHTCHHHHHHHHHHHHC
2.61	1.79	–56.20	CHHHHHHHHHHTSHHHHHHHHTSSCC
3.17	1.67	–54.80	CHHHHHHHHHHTSHHHHHHHHTTTTC
3.33	1.77	–53.50	CHHHHHHHHHHTSCHHHHHHHHTTTTC
3.12	1.47	–53.40	CHHHHHHHHHHTCCHHHHHHHHHHHHC
3.42	1.69	–53.10	CHHHHHHHHHHTSCHHHHHHHHHHHHC
2.83	1.35	–52.80	CHHHHHHHHHHTCCHHHHHHHHHHHHC
4.61	1.86	–52.60	CHHHHHHHHHHTSCHHHHHHHHTSSCC
2.76	1.88	–51.20	CHHHHHHHHHHTCCHHHHHHHHHHSCCC
2.33	1.48	–50.60	CHHHHHHHHHHTCCHHHHHHHHHHSCCC
3.33	1.79	–46.70	CCCHHHHHHHHTSCHHHHHHHHTTTTC
3.45	2.27	–46.50	CHHHHHHHHHHTSTTCHHHHHHSCCC
3.31	1.99	–46.50	CHHHHHHHHSSSCHHHHHHHHHHHHC
6.11	3.98	–45.30	CHHHHHHHHHHTSCSHHHHHHHHSCCC
5.80	4.16	–44.20	CCCHHHHHHHHTSCHHHHHHHHHHHHC
5.74	4.19	–44.00	CCCHHHHHHHHTSCHHHHHHHHHHHHC
5.96	4.27	–43.80	CCCHHHHHHHHTCCHHHHHHHHHHHHC
5.73	4.26	–43.20	CCCHHHHHHHHTSCHHHHHHHHHHHHC
4.99	2.67	–42.90	CHHHHHHHHHCHHHHHHHHHHTSSCC
4.24	3.71	–36.30	CHHHHHHHHTCCTTCTHHHHHHHSCCC

^aWe computed the minimal RMSB deviation (in Å) to the 30 experimental models of the full protein and of amino acids 1–20, respectively. The secondary structure was computed with DSSP.⁶⁷ H, T, S, and C designate helix, turn, strand, and coil conformations, respectively.

one particular conformation with respect to all others does not mean that this conformation is stable with respect to the unfolded ensemble. To settle this question kinetic or thermodynamic simulations must be performed. We have therefore performed all-atom implicit water molecular dynamics simulations for this protein as described in the methods section. The results for the deviation of the actual conformation from the native structure and the two helices are shown in Figure 5. The simulations equilibrate quickly into a rapidly fluctuating ensemble with an average overall rmsd deviation between 5 and 8 Å. When we analyze the rmsd deviation of the helical segments however (Helix 1: 1–11, Helix 2: 15–21), we find that the entire simulation is dominated with conformations that are within 1–2 Å of the respective fragment of the protein. We have also analyzed the helix propensity as a function of time for each amino acid as a function of time, as measured by DSSP. Figure 6 demonstrates a very strong helical content for both segments, but the propensity of helix formation may be forcefield dependent (see below). The figure illustrates very nicely that numerous folding and unfolding events occur for each helix. Both helices disappear completely for short time windows during the simulation, only to form again on a 10 ps time scale.

Next we analyze the sulfur–sulfur distance between CYS8–CYS18 and CYS4–22 as a function of time (lower panels in Figure 5). These distances also fluctuate strongly, averaging more than 10 Å during the simulations. On occasion, however, some of the sulfur atoms approach each other to within 3–4 Å, i.e., close enough for a disulfide bridge to form. On isolated instances, which occur in three of the five simulations (in one simulation two times independently), folding events occur in which both pairs of

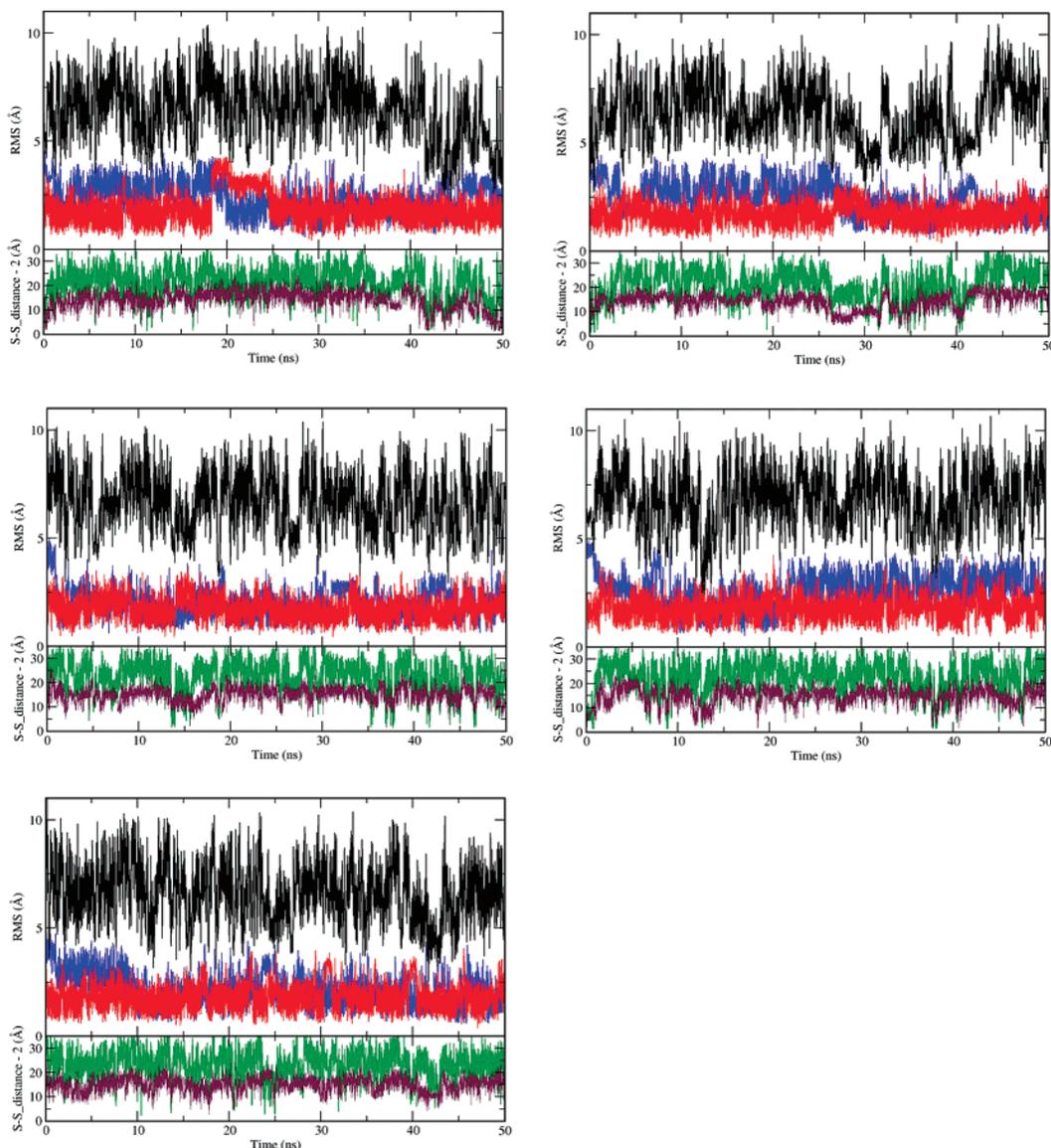


Figure 5. Analysis of the molecular dynamics trajectories as a function of simulation time. The top panel of each graph shows the rmsd of the actual conformation to the native conformation (black) and for the helical fragments only (red: helix 1–11, blue: helix 15–21). The lower panel always shows the deviation of the sulfur–sulfur distance for a potential disulfide bridge (at 2 Å distance) for the amino acids forming the first (green, CYS8–CYS18) and the second disulfide bridge (brown, CYS4–CYS22).

sulfur atoms approach one another (see Figure 7), while both helices are preformed. In those occurrences (which last several ps), the simulations attain all-atom RMSDs to native of 3.43 Å, 3.80 Å, and 3.47 Å, respectively. The intrahelix rmsd vary between 2.1 and 2.5 Å for helix 1 and between 0.8 and 1.0 Å for helix 2 in this time frame.

4. Discussion

From this analysis emerges a picture of the folding process for 1WQE: the low-energy part of the folding funnel is characterized by fluctuating conformation in which both helices are preformed. Both helices fold and unfold repeatedly during the simulation. As the protein explores this landscape it occasionally visits conformations that can lead to the formation of the correct disulfide bridges that would stabilize the native conformation. We note that neither the MD simulations nor the free-folding simulations in PFF01 produced conformations that are consistent with a non-native

disulfide bridge pairing. These events can occur in sequence on a time scale below 1 ns but happen even concurrently on a time scale of the order of 100 ns.

These observations agree with the predictions of the free-energy folding investigation reported above. The free-energy model also predicts the existence of an exclusively helical low-energy ensemble, which collapses into the native conformation at the bottom of the free-energy funnel. Because the free-energy model contains no backbone entropy, the native conformation is found with high probability in the free-energy approach, even though it is not stable (without disulfide bridges) under physiological conditions.

These results are best put into perspective in the context of the framework^{50–52} or diffusion-collision^{53,54} approach of protein folding, where secondary structure fragments of the protein assemble first, which then assume their final tertiary structure by docking into one another. The folding process may thus conceptually be divided into two steps: the

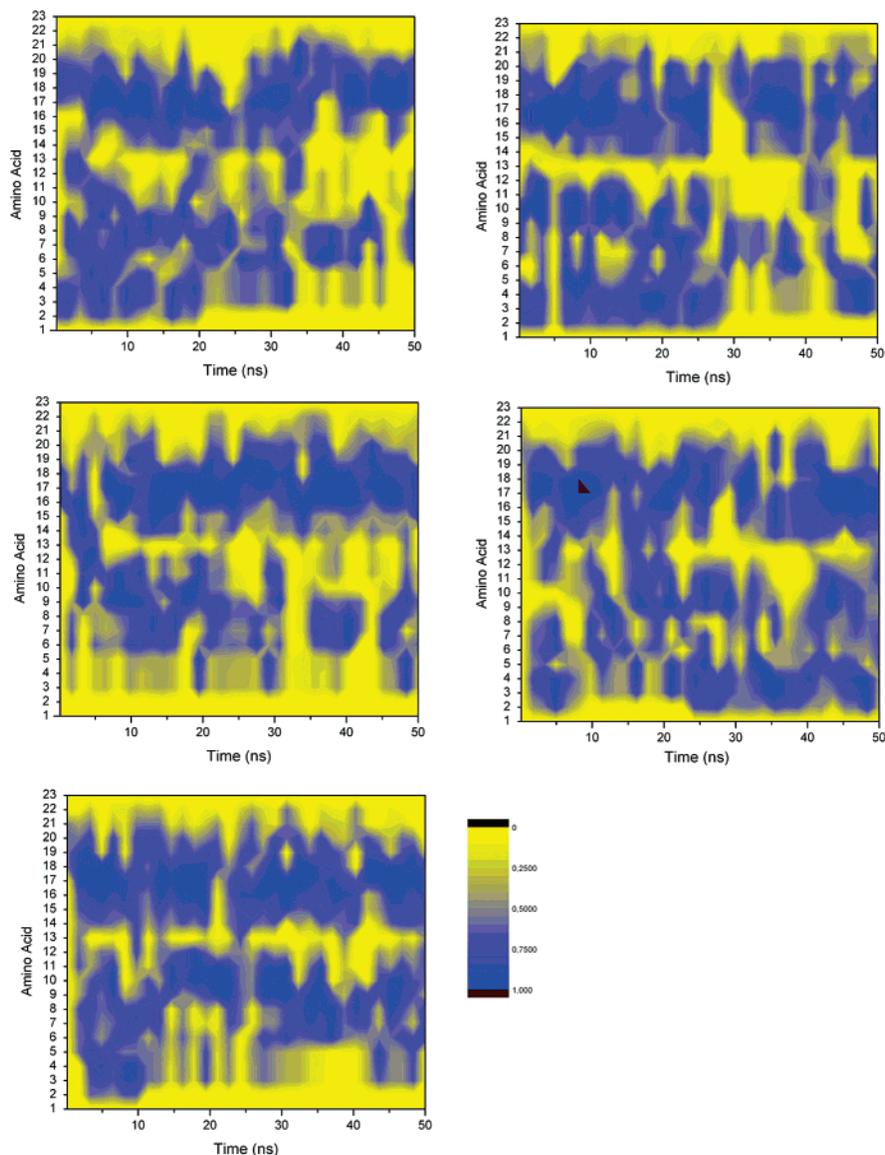


Figure 6. Time average over a 100-ps moving window of the helix propensity of each amino acid in the molecular dynamics simulations. In the native conformation the first helix spans amino acids 1–11, and the second helix spans amino acids 14–21, respectively. The scale for helix probability is the same in all panels.

assembly of the secondary structure precursors and the final collapse into the tertiary structure. For the proteins described here the latter requires the formation of disulfide bridges for enthalpic stabilization. We note that both simulations suggest the existence of preformed helical sections; for the specific systems in question the failure to connect the disulfide bridges in any other than the native topology from the low-energy conformations further supports the idea that the helices must be largely formed for the assembly of tertiary structure. However, in accordance with the funnel paradigm of protein folding,^{3,4} there is no single unique intermediate conformation that must be passed in the folding process to the native state. Instead our simulations suggest the existence of a wide ensemble of two-helix structures that precede the final collapse to the native conformation. In this final collapse there is a huge loss of configurational entropy that is apparently not compensated by the weak hydrophobic free-energy gain (increase in solvent entropy) afforded by this

small system. Therefore an enthalpic contribution from the disulfide bridges is required to stabilize the native conformations.

In this picture the folding time is determined by the rate of helix formation and the rate of disulfide bridge formation. Many present molecular dynamics forcefields, in particular those with implicit solvent models,⁸ may contain a bias toward on particular secondary structure. The combination of Amber99/GBSA that was used in the present study was reported to overemphasize helical secondary structure elements.^{55–58} This would influence the frequency with which the helical precursors for disulfide bridge formation are visited, and further studies are required for quantitative results.

5. Conclusions

In this study we have demonstrated the reproducible folding of three small two-helix proteins, which were recently

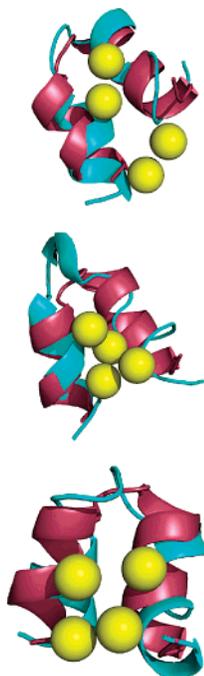


Figure 7. Figure: Overlay of conformations of three folding events from different trajectories of the molecular dynamics simulations (in blue) with the experimental conformation (red). Both helices perfectly formed in all simulations and agree well with the native conformation in their tertiary alignment. Both pairs of sulfur atoms approach each other concurrently for the formation of the disulfide bridges in the correct topology. The sulfur atoms of the MD simulations (not of the native conformation) are shown.

discovered to function as potassium channel blockers. Our simulations yielded final populations which reproducibly and predictively approach the native conformations to within 2 Å. Near native conformations are selected alone on the basis of their free-energy. The three peptides have a high degree of homology but nevertheless differ slightly in the tertiary and secondary structures.

We found that many independent simulations converge to the native conformations, which were predictively selected on the basis of the energy criterion of our force field PFF01. Our characterization of the low-energy part of their free energy surface suggests the existence of very broad and simple folding funnels.^{3,4} Folding of the protein thus proceeds by a diffusion-collision mechanism⁵³ where preformed helices approach one another occasionally to form the disulfide bridges which ultimately stabilize the native conformation.

This prediction was validated in five independent implicit solvent all-atom molecular dynamics simulations, which demonstrated that the individual helix segments preform in equilibrium under physiological conditions. On a time scale of 50 ns we observed several folding events to near native conformation, which validated the folding scenario discussed above.

Our results indicate that for these peptides secondary structure formation precedes hydrophobic collapse,^{59,60} in contrast to most standard folding scenarios. This raises the intriguing question whether it is possible to substitute the

cysteine residues to hydrophobic residues leading to hydrophobic collapse of the preformed helical ensemble into a well-defined tertiary structure that requires no stabilization by disulfide bridges. Such design exercises may help to guide the design of stable hydrophobic cores for such small proteins, which would have implications for important challenges in protein design, e.g., for zinc-finger design.^{61–65} In a recent folding study we could demonstrate that the ATF zinc finger (PDB code 1BHI) folds in the absence of the ion into a preformed ensemble with 3 Å of the native conformation, in a scenario similar to that encountered here.⁶⁶

Acknowledgment. We thank the Deutsche Forschungsgemeinschaft (grants WE 1863/10-2, WE 1863/14-1) and the Kurt Eberhard Bode Stiftung for financial support. Part of the simulations were performed at the KIST teraflop cluster.

References

- (1) Baker, D.; Sali, A. Protein structure prediction and structural genomics. *Science* **2001**, *294*, 93–96.
- (2) Schonbrunn, J.; Wedemeyer, W. J.; Baker, D. Protein structure prediction in 2002. *Curr. Opin. Struct. Biol.* **2002**, *12*, 348–352.
- (3) Dill, K. A.; Chan, H. S. From Levinthal to pathways to funnels: The “new view” of protein folding kinetics. *Nat. Struct. Biol.* **1997**, *4*, 10–19.
- (4) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
- (5) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. Absolute comparison of simulated and experimental protein folding dynamics. *Nature* **2002**, *420*, 102–106.
- (6) Simmerling, C.; Strockbine, B.; Roitberg, A. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **2002**, *124*, 11258–11259.
- (7) Snow, C. D.; Qiu, L.; Du, D.; Gai, F.; Hagen, S. J.; Pande, V. S. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4077–4082.
- (8) Chen, J.; Im, W.; Brooks, C. L., III. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.
- (9) Schug, A.; Herges, T.; Wenzel, W. Reproducible protein folding with the stochastic tunnelling method. *Phys. Rev. Lett.* **2003**, *91*, 158102.
- (10) Herges, T.; Wenzel, W. Reproducible in-silico folding of a three-helix protein and characterization of its free energy landscape in a transferable all-atom forcefield. *Phys. Rev. Lett.* **2005**, *94*, 018101.
- (11) Herges, T.; Wenzel, W. An All-Atom Force Field for Tertiary Structure Prediction of Helical Proteins. *Biophys. J.* **2004**, *87*, 3100–3109.
- (12) Schug, A.; Wenzel, W. All-atom folding of the trp-cage protein in an all-atom forcefield. *Europhys. Lett.* **2004**, *67*, 307–313.
- (13) Schug, A.; Wenzel, W. Predictive in-silico all-atom folding of a four helix protein with a free-energy model. *J. Am. Chem. Soc.* **2004**, *126*, 16736–16737.

- (14) Herges, T.; Wenzel, W. Free energy landscape of the villin headpiece in an all-atom forcefield. *Structure (London)* **2005**, *13*, 661.
- (15) Dinner, A. R.; Sali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M. Understanding protein folding via free energy surfaces from theory and experiment. *Trends Struct. Biol.* **2001**, *25*, 331.
- (16) Pande, V. S.; Rokhsar, D. S. Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062–9067.
- (17) Bonvin, A. M. J. J.; van Gunsteren, W. F. beta-hairpin stability and folding: Molecular dynamics studies of the first beta-hairpin of tendamistat. *J. Mol. Biol.* **2000**, *296*, 255–268.
- (18) Neidigh, J. W.; Fesinmeyer, R. M.; Anderson, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **2002**, *9*, 425–430.
- (19) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. Smaller and faster: The 20-residue trp-cage protein folds in 4 microseconds. *J. Am. Chem. Soc.* **2002**, *124*, 12952.
- (20) Pitera, J.; Swope, W. Understanding folding and design: Replica-exchange simulations of “trp-cage” miniproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.
- (21) Zhou, R. Trp-cage: Folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.
- (22) Schug, A.; Verma, A.; Herges, T.; Lee, K. H.; Wenzel, W. Comparison of stochastic optimization methods for all-atom folding of the trp-cage protein. *ChemPhysChem* **2005**, *6*, 2640–2646.
- (23) Chagot, B.; Pimentel, C.; Da, L.; Pil, J.; Tytgat, J.; Nakajima, T.; Corzo, G.; Darbon, H.; Ferrat, G. An unusual fold for potassium channel blockers: Nmr structure of three toxins from the scorpion *opisthacanthus madagascariensis*. *Biochem. J.* **2005**, *388*, 263–271.
- (24) Dinner, A. R.; Lazaridis, T.; Karplus, M. Understanding beta-hairpin formation. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068–9073.
- (25) Themis, Lazaridis, Martin, Karplus. “new view” of protein folding reconciled with the old through multiple unfolding simulations. *Science* **1997**, *278*, 1928–1931.
- (26) Herges, T.; Merlitz, H.; Wenzel, W. Stochastic optimization methods for biomolecular structure prediction. *J. Assoc. Lab. Autom.* **2002**, *7*, 98–104.
- (27) Abagyan, R. A.; Totrov, M. Biased probability Monte Carlo conformation searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **1994**, *235*, 983–1002.
- (28) Herges, T.; Schug, A.; Wenzel, W. Exploration of the free energy surface of a three helix peptide with stochastic optimization methods. *Int. J. Quantum Chem.* **2004**, *99*, 854–893.
- (29) Avbelj, F.; Moult, J. Role of electrostatic screening in determining protein main chain conformational preferences. *Biochemistry* **1995**, *34*, 755–764.
- (30) Eisenberg, D.; McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* **1986**, *319*, 199–203.
- (31) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry* **1991**, *30*, 9686–9697.
- (32) Pedersen, J. T.; Moult, J. Protein folding simulations with genetic algorithms and a detailed molecular description. *J. Mol. Biol.* **1997**, *269*, 240–259.
- (33) Nayeem, A.; Vila, J.; Scheraga, H. A. A comparative study of the simulated-annealing and monte carlo-with-minimization approaches to the minimum-energy structures of polypeptides: [met]-enkephalin. *J. Comput. Chem.* **1991**, *12*, 594–605.
- (34) Leitner, D. M.; Chakravarty, C.; Hinde, R. J.; Wales, D. J. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *Phys. Rev. E* **1997**, *56*, 363.
- (35) Wales, D. J.; Dewbury, P. E. J. Effect of salt bridges on the energy landscape of a model protein. *J. Chem. Phys.* **2004**, *121*, 10284–10290.
- (36) Mortenson, P. N.; Wales, D. J. Energy landscapes, global optimization and dynamics of poly-alanine Ac(ala) 8 nhme. *J. Chem. Phys.* **2004**, *114*, 6443–6454.
- (37) Mortenson, P. N.; Evans, D. A.; Wales, D. J. Energy landscapes of model polyalanines. *J. Chem. Phys.* **2002**, *117*, 1363–1376.
- (38) Verma, A.; Schug, A.; Lee, K. H.; Wenzel, W. Basin hopping simulations for all-atom protein folding. *J. Chem. Phys.* **2006**, *124*, 044515.
- (39) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680.
- (40) Schneider, J.; Morgenstern, I.; Singer, J. M. Bouncing towards the optimum: Improving the results of monte carlo optimisation algorithms. *Phys. Rev. E* **1998**, *58*, 5085–5095.
- (41) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *Amber 8*; University of California, San Francisco, 2004.
- (42) Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049.
- (43) Born, M. Volumes and hydration warmth of ions. *Z. Phys.* **1920**, *1*, 45–48.
- (44) Hawkins, G. D.; Kramer, C. J.; Truhlar, D. B. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
- (45) Hawkins, G. D.; Kramer, C. J.; Truhlar, D. B. Parametrized models of aqueous free energies of solvation based on pairwise solute descreening of solute charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.
- (46) Berman, H. M.; Westbrook, J.; Feng, Z.; Gillil, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucl. Acids Res.* **2000**, *28*, 235–242.
- (47) Hukushima, K.; Nemoto, K. Exchange monte carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.

- (48) Marinari, E.; Parisi, G. Simulated tempering: a new Monte Carlo scheme. *Eur. Phys. Lett.* **1992**, *19*, 451–458.
- (49) Garcia, A. E.; Onuchic, N. Folding a protein in a computer: An atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13898–13903.
- (50) Ptitsyn, O. B. Sequential mechanism of protein folding. *Dokl. Akad. Nauk SSSR* **1973**, *210*, 1213–1215.
- (51) Kim, P. S.; Baldwin, R. L. Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **1990**, *59*, 631–660.
- (52) Dyson, H. J.; Wright, P. E. Peptide conformation and protein folding. *Curr. Opin. Struct. Biol.* **1993**, *3*, 60–65.
- (53) Karplus, M.; Weaver, D. L. Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Sci.* **1994**, *3*, 650–668.
- (54) Islam, S. A.; Karplus, M.; Weaver, D. L. Application of the diffusion-collision model to the folding of three-helix bundle proteins. *J. Mol. Biol.* **2002**, *318*, 199–215.
- (55) Yoda, T.; Sugita, Y.; Okamoto, Y. Comparisons of force fields for proteins by generalized-ensemble simulations. *Chem. Phys. Lett.* **2004**, *386*, 460–467.
- (56) Wang, T.; Wade, R. Force field effects on a β -sheet protein domain structure in thermal unfolding simulations. *J. Chem. Theor. Comput.* **2005**, *2*, 140–148.
- (57) Lwin, T. Z.; Luo, R. Force field influences in beta-hairpin folding simulations. *Protein Sci.* **2006**, *15*, 2642–2655.
- (58) Hornak, V.; Abel, R.; Stockbrine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinformatics* **2006**, *65*, 712–725.
- (59) Šali, A.; Shakhnovich, E.; Karplus, M. How does a protein fold? *Nature* **1994**, *369*, 248–251.
- (60) Lazaridis, T.; Karplus, M. ‘‘new view’’ of protein folding reconciled with the old through multiple unfolding simulations. *Science* **1997**, *278*, 1928–1931.
- (61) Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. Threedimensional solution structure of a single zinc finger dna-binding domain. *Science* **1989**, *245*, 635–637.
- (62) Kochoyan, M.; Keutmann, H. T.; Weiss, M. A. Architectural rules of the zinc-finger motif: Comparative two-dimensional nmr studies of native and ‘‘aromatic-swap’’ domains define a ‘‘weakly polar switch’’. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8455–8459.
- (63) Choo, Y.; Isalan, M. Advances in zinc finger engineering. *Curr. Opin. Struct. Biol.* **2000**, *10*, 411–416.
- (64) Laity, J. H.; Lee, B. M.; Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **2001**, *11*, 39–49.
- (65) (65) Urnov, F. D.; Miller, J. C.; Lee, Y. L.; Beausejour, C. M.; Rock, J. M.; Augustus, S.; Jamieson, A. C.; Porteus, M. H.; Gregory, P. D.; Holmes, Michael C. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **2005**, *435*, 646–651.
- (66) Gopal, S. M.; Wenzel, W. De-novo folding of the DNA-binding ATF-2 zinc finger motif in an all-atom free energy forcefield. *Angew. Chem., Int. Ed.* **2006**, *118*, 7890–7892.
- (67) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.

CT600274A