Folding Path and Funnel Scenarios for Two Small Disulfide-Bridged Proteins[†]

Ivan Kondov and Abhinav Verma[‡]

Steinbuch Centre for Computing, Forschungszentrum Karlsruhe, P.O. Box 3640, 76021 Karlsruhe, Germany

Wolfgang Wenzel*

Institute of Nanotechnology, Forschungszentrum Karlsruhe, P.O. Box 3640, 76021 Karlsruhe, Germany, and DFG Center for Functional Nanostructures, Department of Physics, Universität Karlsruhe, Wolfgang Gaede Strasse 1, 76131 Karlsruhe, Germany
[‡]*Present address: Centro de Investigaciones Biológicas Spanish Research Council (CSIC), Madrid, Spain.*

ABSTRACT: The presence of disulfide bonds leads to an interesting interplay between noncovalent intramolecular interactions and disulfide bond formation even in small proteins. Here we have investigated the folding mechanism of the 23-residue potassium channel blocker 1WQE and the 18-residue antimicrobial peptide protegrin-1 1PG1, as two proteins containing disulfide bridges, in all-atom basin hopping simulations starting from completely extended conformations. The minimal-energy conformations deviate by only 2.1 and 1.2 Å for 1WQE and 1PG1, respectively, from their structurally conserved experimental conformations. A detailed analysis of their free energy surfaces demonstrates that the folding mechanism of disulfide-bridged proteins can vary dramatically from Levinthal's single-path scenario to a cooperative process consistent with the funnel paradigm of protein folding.

The mechanism of protein folding has been the subject of much debate and controversy over the course of several decades. Starting from the Levinthal's paradox (*1*) through Anfinsen's ground-breaking refolding experiments (*2*), a funnel-based paradigm for protein folding (*3*, *4*), in agreement with experimental observations for small globular proteins, has emerged. Disulfide bond formation has been studied and used in landmark experiments (*5*−*10*) to elucidate the mechanism of protein structure formation and folding. Because disulfide bond formation can be manipulated at various stages of the folding process, the control and probing of disulfide bonds can yield insights into the protein folding process that are difficult to obtain with alternative techniques. In addition, many small proteins, which are most amenable to computational and nuclear magnetic resonance-based investigations of protein folding, are stabilized by disulfide bonds in their native conformation.

Kinetics and thermodynamics of protein folding have been studied by manipulating the formation (*6*, *7*) or mutation (*5*) of disulfide bonds. Reconstruction of folding intermediates of bovine pancreatic trypsin inhibitor (BPTI) and RNase T1 (*6*) helped to elucidate possible scenarios for the folding pathways of these proteins. For particularly stable proteins, folding to the native state may require the formation of only a subset of native disulfide bonds. A number of studies on oxidative folding of RNase A (*7*−*9*) have indicated a high degree of cooperation between disulfide bond formation and chain folding. Moreover, it was shown for the same system that disulfide-coupled folding kinetics does not follow the "rugged funnel" models of protein folding (*7*). Furthermore, in an atomic force microscopy (AFM) study (*5*) of cysteine pair mutants of protein I27, the disulfide bond position had interesting implications for native state stability. In particular, the unfolding rate constant is decreased for cysteine pair mutants compared to that of the wild-type protein, while the distance to the transition state is not influenced by the engineered disulfide bond.

For a range of disulfide-containing proteins (*11*), enthalpic stabilization of the native state by disulfide bridges outweighs entropic destabilization by a reduction of the hydrophobic effect. Disulfide bonds may thus strongly influence protein stability. For the scorpion toxin HsTx1 (*10*), for example, the fourth disulfide bridge was shown to be mandatory for the formation of the native conformation, in contrast to its structural analogue Pi1 (*10*). While having no impact on the helical structure in HsTx1, the removal of this disulfide bond changes the two-stranded β-sheet from a twisted to a nontwisted configuration.

Using experimental techniques to control oxidative conditions during the folding process, several middle-sized disulfide-containing proteins were studied recently (*12*). Therein, two opposed models for the folding mechanisms were proposed. For proteins such as BPTI, the folding proceeds on a funnel-shaped energy landscape along a path containing a certain number of nativelike intermediates. The disulfide bonds and secondary structure are formed progressively with each intermediate step. In contrast, proteins like hirudin fold via a fast initial collapse with immediate disulfide bond formation followed by folding to the native conformation without building intermediates in terms of minima on the energy landscape.

In recent years, the folding process of several proteins without disulfide bridges could be studied in detail in simulations (*13*−*16*) using all-atom protein force fields (*17*), molecular dynamics (*18*, *19*), replica exchange (*16*), and Monte Carlo

(*14*, *20*) or nonequilibrium methods (*21*−*24*). However, all-atom simulations still face difficulty with proteins containing disulfide bridges (*25*, *26*), due to the long time scale on which these processes occur. Folding of disulfide-bridged proteins has thus been studied using conformational space annealing (*27*), lattice models (*28*), and a number of bioinformatics-based techniques (*29*−*33*). Various proteins with disulfide bridges were investigated (*27*) with the conformational space annealing method using a united-atom force field, where disulfide bonds were modeled using distance-dependent harmonic potentials. Only one of the four proteins could be folded without prior knowledge of the disulfide bond topology. Recent molecular dynamics studies (*25*) of the disulfide-bonded protein Tendamistat suggested an important role for the stabilization as well as for the folding pathway.

While disulfide bonds stabilize the protein in the thermodynamic sense, they may impede the folding process by stabilizing misfolded intermediates (*34*), especially in early folding stages (*28*). In the latter study, it was found that the stiffness of the polypeptide chain due to constrained motion results in lower folding rates. Specifically, disulfide bonds that accelerate folding seem to be mostly formed in early stages of the folding process (in the rate-determining formation of the so-called folding nucleus), while those that decelerate the folding are formed only after the rate-determining step. The competition between secondary and tertiary structure formation and the formation of disulfide bonds of varying topology offers a new degree of freedom that may be exploited in the elucidation of the folding process for small proteins accessible to both experiment and simulation.

We have recently investigated the potassium channel blocker 1WQE containing two disulfide bonds using a combination of free energy and molecular dynamics simulations (*35*, *36*) under reducing conditions. While failing to stabilize the native configuration, the protein exhibited significant native secondary structure content and even visited near-native configurations. Here we extend our approach to study folding of disulfide-bridged proteins under oxidizing conditions, including the long time scale process of disulfide bond formation.

MATERIALS AND METHODS

In recent years, we have pursued a free energy approach to protein folding and structure prediction and developed an all-atom force field based on models of the most important biophysical interactions. This method (*37*) is based on the thermodynamic hypothesis stipulating that many proteins in their native conformations are in thermodynamic equilibrium with their environment. According to this paradigm (*2*), the native conformation of a protein corresponds to the global minimum of its free energy surface. In this approach, each protein backbone conformation **r** is assigned an "internal free energy" obtained by integrating the solvent degrees of freedom [within the SASA model (*38*)], such that the relative free energy difference between two conformations i and j with energies E_i and E_j, respectively, is given by $E_i - E_j$. The advantage of this approach is that it decouples the sampling of the conformational space from the computation of relative free energies of conformations. Therefore, we can use any sampling technique, including nonequilibrium methods such as the basin hopping technique, to generate a protein conformational ensemble until the low-energy region of the free energy surface containing the native conformation is sufficiently sampled.

Force Field. The PFF01 (*21*, *22*) and PFF02 (*37*, *39*) free energy protein force fields have been developed to describe the free energy surface. The PFF01 force field comprises the following nonbonding interactions

$$G(\mathbf{r}) = \sum_{ij} V_{ij}\left[\left(\frac{R_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}}{r_{ij}}\right)^6\right] + \sum_{ij}\frac{q_i q_j}{\varepsilon_{g_i g_j} r_{ij}} + \sum_i \sigma_i A_i + \sum_{\text{H-bonds}} V_{\text{hb}} \qquad (1)$$

where r_{ij} values denote the interatomic distances. The Lennard-Jones potential depth and equilibrium distance, V_{ij} and R_{ij}, respectively, have been optimized imposing constraints of series of proteins from the PDB.[1] The second, electrostatic term depends on the partial atomic charges (q_i) and pairwise dielectric constants ($\varepsilon_{g_i g_j}$), with g_i being the type of amino acid residue pertaining to atom i. The last two terms provide implicit description of solvent interaction [the SASA model (*38*)], where A_i is the contact area for atom i and V_{hb} is the short-range backbone−backbone hydrogen bonding.

While the PFF01 force field was parametrized for helical proteins, its extension PFF02 (*37*, *39*, *40*) includes corrections that permit folding of β-sheets and proteins of mixed secondary structure. To date, more than 20 proteins (*40*) with 20−60 amino acids could be folded to an average backbone root-mean-square deviation (rmsd) of 2.8 Å from completely unfolded conformations. Thus, the PFF02 force field provides an unbiased and transferable description of the free energy landscape for proteins containing both β-sheet and α-helix secondary structures. The PFF02 force field was used throughout this study.

In this study, we describe disulfide bridges by a Morse potential

$$V_{\text{SS}}(r) = E_0\{[1 - e^{\beta(r - r_0)}]^2 - 1\} \qquad (2)$$

that is widely used to describe chemical bonds in classical force fields. r_0 is the equilibrium distance between the sulfur atoms forming a disulfide bridge; E_0 is the energy corresponding to r_0, and β is characteristic of the spatial extent of the potential. In the following, we used an r_0 of 2 Å and a β of 1 [1]. In this work, we choose E_0 values of 2 and 5 kcal/mol for 1WQE and 1PG1, respectively. In comparison, Czaplewski et al. (*27*) employed a harmonic potential biased by 5.5 kcal/mol to describe the disulfide bond. Investigations exploring various potential forms and parameters (*41*) show that the Morse potential with the parameters given above has the best overall performance for the proteins studied compared to other potentials. In particular, it performs better than harmonic potentials because it is short-range; i.e., it decays exponentially on long distances. The merits and drawbacks of the Morse and other constraining potentials for folding these two and further proteins are discussed elsewhere (*41*).

To interpret energy differences in this model, it is important to recall that our description assigns an internal free energy to each backbone conformation. By virtue of its parametrization, this internal free energy contains an entropic contribution of the solvent and side chain entropy (using the vacuum as a reference). Energy differences between different conformational ensembles must be corrected by the backbone entropy difference to arrive at

---

[1]Abbreviations: FNC, fraction of native contacts; PDB, Protein Data Bank; rmsd, root mean square deviation.

the physical energy scale. Such corrections are large when the differences between highly flexible unfolded ensembles and near-native collapsed conformations are considered. The late stages of the folding process, which are considered here, occur between conformations in the collapsed ensemble which have similar backbone entropy, so these corrections are small. Extensions of the basin hopping method have been suggested to account for these differences but could not yet be validated for protein simulations (*42*).

Simulation Method. The basin hopping technique (*23, 43*) employs a relatively straightforward approach to eliminating high-energy transition states of the free energy surface. The original free energy surface is simplified via replacement of the energy of each conformation with the energy of a nearby local minimum. This replacement eliminates high energy barriers in the stochastic search that are responsible for the freezing problem in simulated annealing. In many applications, the additional effort for the minimization step is compensated by the improved efficiency of the stochastic search. The basin hopping technique and derived methods (*44*) have been used previously to study the free energy surface of model proteins (*45*) and polyalanines using all-atom models (*46, 47*).

Every basin hopping cycle contains a move generation step, a minimization step, and an acceptance criterion. In the following, we use move generation methods that gradually distort the conformation by small changes of the dihedral angles of the protein, thus generating a nearly continuous protein trajectory. Side chain and backbone angles are changed in 30 and 70% of the moves, respectively. The dihedral angle changes for 50% of the moves are drawn from a uniform distribution with a maximal change of 5°. The angle changes for the remaining 50% of the moves are taken from a move library that accounts for side chain rotamer preferences, as described previously (*23*).

Here we replace the gradient-based minimization step with a simulated annealing run (*48*), because local minimization generates only very small steps on the free energy surface. Within each annealing simulation, new configurations are accepted according to the Metropolis criterion, while the temperature is decreased geometrically from its starting value to its final value. The starting temperature and cycle length determine how far the annealing step can deviate from its starting conformation. The final temperature must be small compared to typical energy differences between competing metastable conformations, to ensure convergence to a local minimum. The annealing protocol is thus parametrized by the starting temperature (T_S), the final temperature ($T_F = 2$ K), and the number of steps. We investigate various choices for the numerical parameters of the method but have always used a geometric cooling schedule (*23*).

At the end of one annealing cycle, the new conformation is accepted if its energy difference with respect to the current configuration is not higher than a given threshold energy ε_T. Throughout this study, we use a threshold acceptance criterion of 1 kcal/mol. Here we started 30 copies of the extended conformation (denoted in the following by X); i.e., all backbone dihedral angles were set to 180°, and 800 and 600 independent basin hopping cycles were performed for 1WQE and 1PG1, respectively. In every cycle, the initial temperature was chosen from an exponential distribution and the number of simulated annealing steps was increased following ref *23*. In addition, 30 simulations with a length of 300 cycles have been performed for both proteins starting from the NMR structure of each protein under the conditions outlined above. Then the sets of accepted structures of

all runs were used in the further analysis. To determine secondary structure, we used DSSP (*49*). The protein structures were visualized using PyMOL (*50*).

Native Contacts. The fraction of native contacts (FNC) was computed for each accepted conformation. The use of the contact order or FNC as folding reaction coordinates has been established (*51, 52*). We have used the MMTSB tool set (*53*) to calculate the residue−residue native contacts based on distances between heavy atoms. In this study, a threshold distance of 4.2 Å between side chain heavy atoms has been used. While the rmsd is characteristic for the quality of the secondary structure, native contacts characterize the mutual alignment of side chains in the tertiary native structure. It is noted that the NMR structures for the two proteins studied exhibit rather different side chain alignment (up to 0.4 difference in units of FNC). Thus, the best structure obtained, starting from both the extended and the natural conformations, was chosen as a reference for calculation of FNC rather than NMR structures. The accepted conformations from all independent simulations were sorted according to their amount of native contacts. Then, for each FNC, the minimal energies were calculated. The probability of the disulfide bond presence was calculated from the number of all accepted conformations with a sulfur−sulfur distance of < 3.6 Å for 1WQE and 3.0 Å for 1PG1.

Connectivity Tree. To characterize the tertiary structure of transition and intermediate states along the folding pathway of the two proteins, a decoy family tree (connectivity tree) was constructed. This tree characterizes the topology of the low-energy region of the free energy landscape. To construct the connectivity tree (*54*), we group all conformations encountered in the simulations into clusters according to energy brackets. The bottom of each tree branch corresponds to a metastable conformation of the protein, which grows with an increase in energy into a larger, structurally similar ensemble. When two such families overlap structurally, the branches of the tree are joined, illustrating the topological relationship of low-lying metastable ensembles. In this work, we applied a clustering algorithm as described and used in ref *55*. Thereby, the threshold for merging of families is 3 Å, and only families with a contribution larger than 1% in the total population were considered.

RESULTS

Two-Helix Protein. The 23-amino acid potassium channel blocker 1WQE (*56*) comprises two α-helices (Pro3−His11 and Val15−Cys22) connected by a turn region. The native conformation is stabilized by two disulfide bonds between residues Cys4 and Cys22 (bond at end region E, E-bond) and Cys8 and Cys18 (bond at turn region T, T-bond).

Previously, we have investigated this protein in a combination of free energy and molecular dynamics simulations (*35, 36*). These simulations were performed under reducing conditions. I. e., they did not employ a disulfide bond potential. Therein, the free energy protocol folded 1WQE to a low-energy conformation with correct secondary structure, while failing to fully satisfy the disulfide constraints. The sulfur atoms in the lowest-energy conformation had a d_E distance of 3.3 Å (E-bond) and a d_T distance of 5.5 Å (T-bond), compared to the distance of 2.0 Å in the native conformation for both. Subsequent molecular dynamics simulations (*35, 36*) started from this low-energy conformation suggested that the secondary structure elements are stable. The tertiary arrangement of the helices fluctuated

Table 1: Characterization of the Lowest Energy Members of Most Prominent Decoy Families for Protein 1WQE[a]

family	secondary structure	E-bond	T-bond	FNC	RMSD	E_{tot}	E_{PFF02}
N	CHHHHHHHHHHCSCHHHHHHHHC	+	+	1.00	2.11	−45.30	−42.99
IH1	CHHHHHCCSCSSSCHHHHHHHHC	−	−	0.17	6.29	−40.48	−40.47
IH2	CCSSHHHHHTCCSCHHHHHHHHC	+	−	0.42	3.90	−37.89	−36.58
IH3	CCSCHHHHHCCSSCHHHHHHHHC	−	+	0.17	6.27	−36.94	−35.60
H4	CHHHHHHHCCCSSSCCHHHHHC	+	+	0.33	4.93	−35.49	−32.73
U1	CCSSSSCCBSSSSSCSSSCSBCC	−	+	0.25	6.14	−31.99	−30.79
H5	CCSSHHHHSCCSSSCCCSHHHHC	−	−	0.17	3.85	−30.42	−30.08
H6	CCSSHHHHHHHHHHTCCHHHHC	−	−	0.00	6.06	−30.34	−30.23
U2	CCSSSSCCSEESSCCSSCSEECC	−	−	0.17	6.65	−29.53	−29.51
U3	CCSCSSSCCSTTSSSTTSHHHHC	−	−	0.17	6.89	−27.79	−27.77
U4	CCSSCSSSCCSCCSSSCCCTTTTC	−	−	0.17	6.44	−26.31	−26.26

[a] The secondary structure of 1WQE determined by NMR is CCHHHHHHHHHHTCCHHHHHHHHHC (56). The reference for FNC is conformation N. The reference for rmsd is the NMR structure (56).
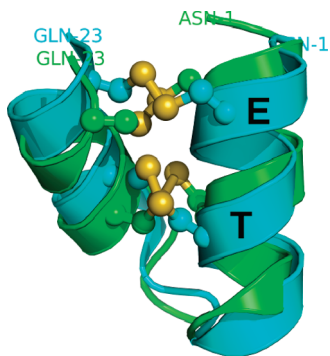

FIGURE 1: Cartoons of folded protein 1WQE (colored green) superimposed on the corresponding native structures (colored cyan). Sulfur atoms of cysteine residues forming disulfide bridges are depicted as golden spheres.

FIGURE 2: Folding minimal energy profiles (●) and probabilities $P(S\text{–}S)$ of building a disulfide bond in the end region (◆) and the turn region (▲) for protein 1WQE. Data points are interpolated using Akima splines (solid lines). Vertical straight lines separate the folding stages as discussed in the text. The regions with letters U, I, and N indicate families of unstructured, intermediate, and nativelike conformations, respectively.

significantly, but visited conformations were in agreement with the correct disulfide bonding pattern on a short (∼50 ns) time scale.

Here, we study the protein in the presence of the disulfide bridge potential described above in 30 independent basin hopping simulations starting from completely extended conformations. The results of the folding simulations are summarized in Table 1. We find a lowest-energy conformation within 2.1 Å of the native conformation. Figure 1 illustrates the excellent agreement between the folded and experimental conformation with regard to the secondary structure, the location of the turn, and the presence of the disulfide bridges. However, we note a dramatic reduction in the probability of reaching the native conformation, when comparing simulations (of equal length) under oxidizing and reducing conditions. While absolutely essential for the thermodynamic stabilization of the native conformation, the presence of the disulfide bridge potential thus seems to impede rather than enhance folding. To understand this phenomenon and elucidate the folding process, we have computed the fraction of native contacts (FNC) for all accepted
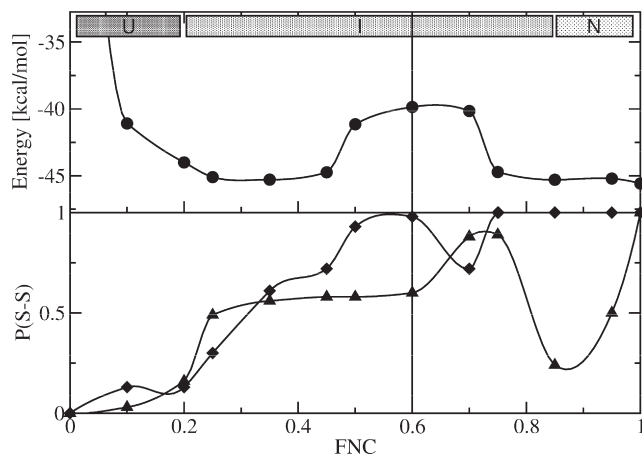
conformations in our simulations (cf. Materials and Methods) and drawn the energy profile as a function of the FNC as shown in the top panel of Figure 2. The fraction of native contacts has often been used as a one-dimensional reaction coordinate for protein folding (51, 52). To improve our understanding of the interplay between disulfide bond formation and the folding process, we have additionally calculated the probability of formation of each disulfide bond for each value of FNC. The bottom panel of Figure 2 depicts the probabilities $P(S\text{−}S)$ for disulfide bridge presence versus FNC.

For FNC values of < 0.35, we observe a monotonous decay of the energy that is accompanied by a simultaneous increase in the probability of disulfide bond formation for both bonds. For FNC values of > 0.35, further formation of the T-bond is hindered: its probability exhibits a plateau at approximately
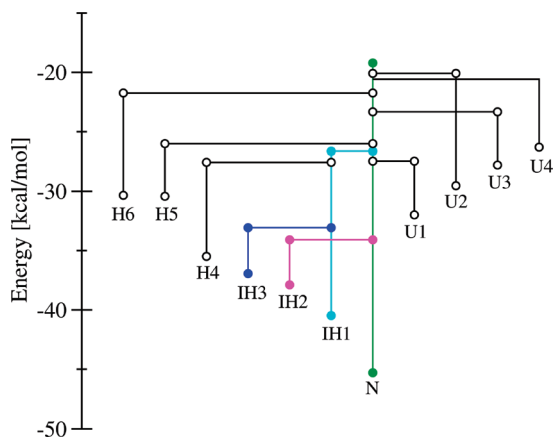
FIGURE 3: Decoy tree for protein 1WQE. The lowest energy conformations are labeled as in Table 1.
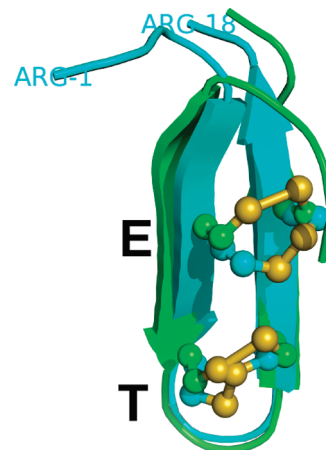

FIGURE 4: Cartoons of folded protein 1PG1 (colored green) superimposed on the corresponding native structures (colored cyan). Sulfur atoms of cysteine residues forming disulfide bridges are depicted as golden spheres.

50% for FNC values from 0.35 to 0.6. In contrast, the E-bond is formed with almost unit probability already for conformations with 50% native contacts. This region is followed by a free energy barrier (0.5 < FNC < 0.7) of ~5 kcal/mol that is associated with a significant reduction in the probability of forming the E-bond (~30%). An FNC of ~70% corresponds to the thermodynamic transition state. For larger FNC values, we observe a rapid decrease in the energy accompanied by a re-formation of the E-bond.

Thus, folding seems to proceed in two major phases. In the first phase, both disulfide bridges are partially formed in a compact ensemble which is inconsistent with the final tertiary structure of the protein. In the second phase, the T-bond must, on average, be broken and re-formed once the correct tertiary arrangement is approached. Both the formation of the disulfide bridges and the correct tertiary structure are required for thermodynamic stabilization of the protein. Observation of a low disulfide bond probability at early stages is consistent with the finding that "early" disulfide bonds accelerate while "late" ones decelerate the total folding process (28). Since formation of the T-bond occurs after the transition state, folding into the native ensemble is rapid. As a result, there are few data in the region of FNC values between 0.85 and 0.95, but there is clear evidence for disulfide bond rearrangement.

To go beyond the description based on a one-dimensional reaction coordinate and elucidate the thermodynamics of the folding process, we constructed the connectivity tree of the protein, as shown in Figure 3. Connectivity trees (see Materials and Methods) completely characterize the complex low-energy conformational ensemble of the protein and the topology of the free energy surface. Each branch terminates in a metastable conformation representing a local minimum in the internal free energy of the protein. The structural parameters characterizing the 11 families with more than 1% probability of occurrence (of 67 families in total) are summarized in Table 1. The lowest-energy conformation of the first family, shown at the bottom of the innermost branch of the tree and labeled N, is the native fold exhibiting both α-helices connected by a turn passage in the middle and two disulfide bonds. The next three families, sorted by energy, still have significant resemblance with the native conformations but have incompletely formed helical regions. Families IH1, IH3, and H4 merge together at a rather low energy, while all other families merge as single branches to the trunk. Conformation H5 exhibits less pronounced native structure

elements, while conformations from families U1, U2, U3, and U4 have large unstructured regions. The low degree of local structure in the tree and the absence of clearly structured branches are indicative of a funnel-shaped free energy landscape. Overall, we observe a good correlation between the presence of disulfide bonds and an increasing level of native content (measured by FNC). Structure H6 has no native contacts and disulfide bonds but a low energy which results from significant helical content. Conformation H4 has two disulfide bonds but a smaller FNC than conformation IH2 due to less well formed secondary structure. Furthermore, conformation U1 has one disulfide bond and therefore a larger FNC, although the secondary structure does not resemble that in the native conformation at all.

β-Hairpin Folding. In the second part of this study, we investigate a β-sheet protein, 1PG1 (57), which has a native conformation with two β-sheet strands connected by a turn backbone region and two disulfide bridges near the backbone termini (labeled E for end) and the turn (labeled T for turn) between Cys6 and Cys15 and between Cys8 and Cys13, respectively. We first performed simulations starting from the native conformation, as retrieved from the PDB, with and without disulfide bond potential. In the simulation with the disulfide bond potential, the protein is stable in the native conformation, while the simulation without the disulfide bonding potential diverges from the native structure by as much as 7.26 Å. This shows that the disulfide bridges are essential for stabilizing this protein, which is structurally well reproduced in their presence.

We have then performed 30 independent folding simulations starting from a completely extended conformation X, which differed by ~18 Å from the native conformation and has no secondary structure and also no disulfide bonds. The simulation with the lowest energy converges to a conformation which, in the structurally defined region, agrees well with the experimental structure (first NMR conformation in the PDB set), as illustrated in Figure 4. For the whole peptide, the large overall backbone rmsd of 3.78 Å results from the outlying four-residue fragment (Arg-Gly-Gly-Arg) at the N-terminus that is most likely unstructured under physiological conditions. When this region is excluded, the backbone rmsd between the folded and simulated structure is just 1.26 Å, indicating that the simulation reproduces the native structure within the experimental resolution. Table 2

Table 2: Lowest Energy Members of the Most Prominent Decoy Families for Protein 1PG1[a]

family	secondary structure	E-bond	T-bond	FNC	RMSD	RMSD−	E_{tot}	E_{PFF02}
N	CCSSEEEEETTEEEEECC	+	+	1.00	3.78	1.26	−48.19	−40.63
B1	CTTCSSEEETTEEECCCC	+	+	0.41	7.60	3.53	−44.20	−36.62
U1	CCSSSBCCTTTSCCCBCC	+	−	0.36	5.98	4.63	−43.68	−39.91
U2	CCSBTTBCSSSCCBSCCC	+	+	0.36	6.39	5.31	−43.22	−35.72
B2	CBTTTEEETTTSEEEECC	+	−	0.27	7.23	6.69	−43.03	−39.24
B3	CTTTSCEEETTEEESSCC	+	+	0.50	6.77	6.10	−42.14	−34.50
B4	CCCCEESCTTTSEECSCC	−	−	0.28	8.22	4.02	−38.75	−38.59
U3	CTTSCBSSSSSBSCCCCC	+	−	0.27	7.27	4.63	−37.68	−33.86
H1	CTTSHHHHHHHSCCCCC	+	−	0.23	5.58	4.22	−36.52	−32.71
H2	CHHHHHHHHHHSCCTTC	+	−	0.09	5.75	5.52	−36.23	−32.50
H3	CHHHHHHHHHSSCCSCC	−	+	0.23	4.96	4.52	−35.60	−32.03
H4	CCSSCTTSHHHHHHHHTC	+	−	0.14	6.86	6.07	−33.67	−29.99

[a] The secondary structure of 1PG1 determined by NMR is CCSEEEEEETTEEEEEC (57). The reference for FNC is conformation N. The reference for rmsd is the NMR structure (57). rmsd− is calculated without the first four residues (see the text).
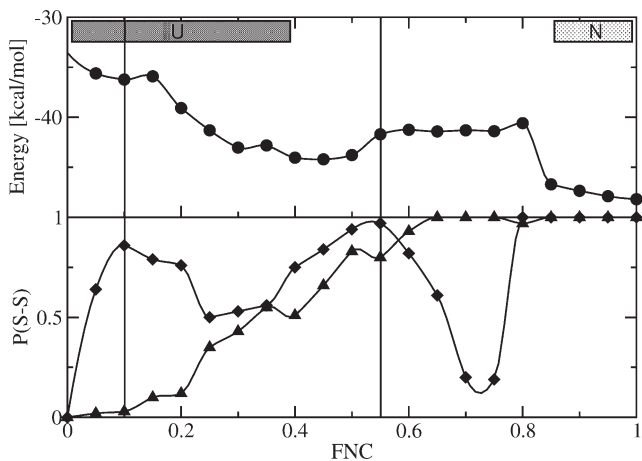

FIGURE 5: Folding minimal energy profiles (●) and probabilities $P(S\text{ }S)$ of building a disulfide bond in the end region (◆) and the turn region (▲) for protein 1PG1. Data points are interpolated using Akima splines (solid lines). Vertical straight lines separate the folding stages as discussed in the text. The regions containing letters U and N indicate families of unstructured and nativelike conformations, respectively.

summarizes the structural features of the lowest-energy conformations of 12 selected families (see the decoy tree in the text below). Several of these conformations exhibit partially β-sheet segments on both sides of the turn. However, not all of the conformations contain closed disulfide bridges, as indicated in the table. Finally, there are some high-energy conformations (H1−H4) that do not even possess the correct secondary structure.

It is interesting to note that several designed β-hairpin peptides without disulfide bonds, such as the Trp-zipper peptides, fold very rapidly with the same protocol, requiring less than 2% of the number of steps with a folding probability of >90% (39). According to the "new paradigm" of protein folding, it is generally assumed that the protein folding process is guided by

a funnel in the free energy landscape, which accelerates the process and permits the molecule to find its unique native conformation with high probability (58). Here, the folding simulations in the absence of the disulfide bridge potential show that 1PG1 must fold against the gradient of its internal free energy profile.

If we decompose the energies into contributions associated with the disulfide potential and the rest of the energy function, we find that the near-native conformations found in the simulation are separated by a wide energy gap from all competing conformations when the disulfide bridge term is considered. With the disulfide bridge potential, the native conformation (N) is separated by 5 kcal/mol from unstructured conformation U1, for example. If the disulfide bridge potential is switched off, this difference shrinks to <1 kcal/mol, indicating that the native conformation is losing its thermodynamic stability.

To analyze the folding scenario, we have gathered all accepted conformations encountered in the folding simulations and reconstructed the free energy profile. Figure 5 shows the free energy versus the FNC (cf. Materials and Methods). Conformations with low FNC values are characterized by high energies. With increasing contact order, the free energy falls slightly and after a rather low barrier at an FNC of 0.15 undergoes a subsequent much stronger decay (of ~8 kcal/mol). In the FNC region from 0.3 to 0.5, the free energy profile is almost constant. A very well pronounced plateau-like transition state along the reaction coordinate can be identified between FNC values of 0.55 and 0.8. It is characterized by an activation free energy of 3.6 kcal/mol.

The energetic contribution of the disulfide binding potential changes as a function of FNC. As the repulsive part of the Morse potential is essentially not sampled in our simulations, the total energy (including the disulfide bond potential) should be always less than the PFF02 energy. We find that the energy contribution from the disulfide bond potential increased from approximately 4 kcal/mol (corresponding roughly to one disulfide bond) for an
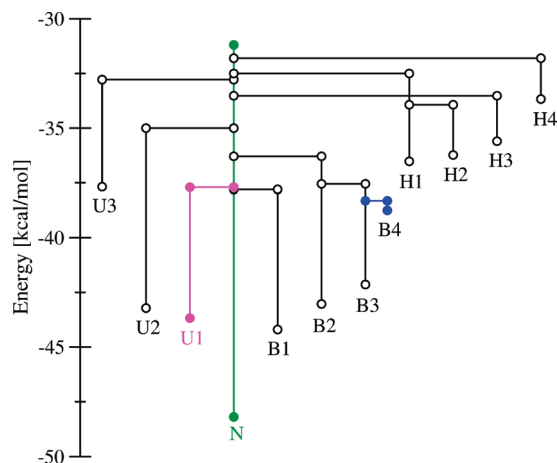
FIGURE 6: Decoy family tree for protein 1PG1. The lowest energy conformations are labeled as in Table 2.

FNC of 0.1 to 8 kcal/mol for the near-native conformations (cf. Table 2), on average. This agrees with the probability of disulfide bond formation $P(S-S)$, as indicated in the bottom panel of Figure 5. For the extended conformation with an FNC of 0, the probabilities for occurrence of both disulfide bonds are vanishing. Subsequently, we observe three stages in the formation of the disulfide bonds. In the first stage, where the FNC increases from 0 to 0.1, the E-bond is rapidly formed with probabilities of up to 0.86. In contrast, the bond at the turn region (T-bond) stays mostly open (closure probability of 3%). This stage is characterized by a barrierless decay of the energy (cf. Figure 5, top panel). The moderate energy decrease is indicative of a simple bond formation that is not accompanied by major reorganization of the secondary structure. The presence of this bond may hinder further formation of secondary and tertiary structure. In the second stage, we observe a decay and subsequent plateau of reduced $P(S-S) \approx 0.5$ for the E-bond in the FNC range from 0.15 to 0.35, which follows the small free energy barrier at an FNC of 0.15. In this region, we find a rapid increase in $P(S-S)$ for the T-bond from near zero to 0.50, indicating an increasing level of formation of the second disulfide bond. This increase continues to 75% probability at an FNC of 0.55. In this region, the nucleation at the turn competes with the formation of the E-bond. Similar competition between disulfide bond formation and nucleation has been discussed in ref *28*. In the third stage, for FNC values of >0.55, we observe a significant decrease in the probability for the E-bond, while the T-bond is completely formed. This reorganization correlates well with the broad barrier in the energy profile. For FNC values of >0.8, both disulfide bridges are closed.

To characterize the relevant conformations, we computed the connectivity tree (see Figure 6). The structural information for the relevant conformations is summarized in Table 2. We find that 99% of the sampled population can be represented by 12 families (of a total of 52 families). The lowest-energy conformation of the first family (N) exhibits the native fold (cf. Figure 4).

The second family (B1) exhibits a similar overall structure and two completely formed disulfide bonds. The significantly larger rmsd of 7.6 Å is due to the incompletely formed secondary structure with a reduced β-sheet content. Family U1 represents a conformation lacking the β-sheet strands but featuring a turn in the backbone middle and a disulfide bond at the end. The reason for the stabilization of a coil-like structure is the large energy gain (~3.8 kcal/mol) resulting from formation of the single disulfide

bond at the backbone end. The family associated with this conformation merges with the main family at an energy only slightly above that of family B1. The lowest-energy conformation of the next family (U2) is unstructured but includes two disulfide bonds. Again, the stabilization is primarily due to the formation of two disulfide bonds. This family comprises structures of the crossover region [0.5 < FNC < 0.6 (cf. Figure 5)] and merges to the trunk at a very high energy. The next two families, B2 and B3, are dominated by nativelike but incompletely folded structures. The former has a better formed secondary structure (and thus lower total energy) and one disulfide bond at the end. The latter family has less pronounced β-sheet regions but two well-formed disulfide bonds and a high fraction of native contacts (50%). These two structures are separated by a small local barrier and merge together with the trunk family after families B1 and U1. Family B4, exhibiting a very short β-sheet region and no disulfide bonds, merges with B3 after a very small barrier. The top conformation of family U3 is very similar to U2 except for the absence of the disulfide bond at the turn. It merges independently at a very high energy to the trunk. The last four families shown in Figure 6 comprise an α-helix and a coil connected by a small turn region. Their lowest-energy conformations are well separated by more than 10 kcal/mol from the first seven families and contain one disulfide bond, predominantly at the backbone ends (H1, H2, and H4).

Energy Landscapes. While the connectivity tree elucidates the topology of the free energy landscape, the folding process can be interpreted more easily on the basis of projected low-dimensional free energy surfaces, which exhibit the low-lying metastable regions. Disulfide bond formation occurs on time scales that are difficult to sample with present day kinetic simulation methods. In the absence of explicit kinetics, we have taken the approach to characterize the internal free energy landscape of the protein using efficient sampling techniques. Folding of the protein can then be understood in terms of transitions among the low-energy conformations of the protein and visualized by suitable projections to low-dimensional reaction coordinates. Such projections always run the risk of omitting important details of the folding process. For this reason, we have computed the connectivity trees, which are free of any assumption regarding the reaction coordinates and fully characterize all relevant low-energy conformations. In Figure 7, we plot the free energy of the locally most stable conformation as a function of disulfide bond distances $r_{SS,E}$ and $r_{SS,T}$ for the E-bond and T-bond, respectively. In such coordinates, the extended conformation X of 1WQE is located at $(r_{SS,E}, r_{SS,T})$ = (65 Å, 38 Å). In agreement with the topological information for the decoy tree, we observe a funnel-like structure of the free energy landscape with an overall downhill slope toward the native conformation.

Late events in the folding process are characterized by transitions between the low-energy regions of the free energy landscape. The lowest-energy conformation of family IH1 (cf. Figure 3, depicted in cyan) is characterized by partially formed helical regions and the absence of disulfide bridges. It is connected to two major metastable basins at (10.8 Å, 2.9 Å) and (2.9 Å, 8.9 Å), and to the global minimum (colored green). Conformations in the region around IH3 (colored blue in Figure 7) feature fully formed helices and a disulfide bond near the turn. The energy barrier for transitions from structure IH1 to the transient structure IH3 at (10.8 Å, 2.9 Å) is ~10 kcal/mol. The structure associated with the other local minimum, again with almost fully formed secondary structure, is connected to
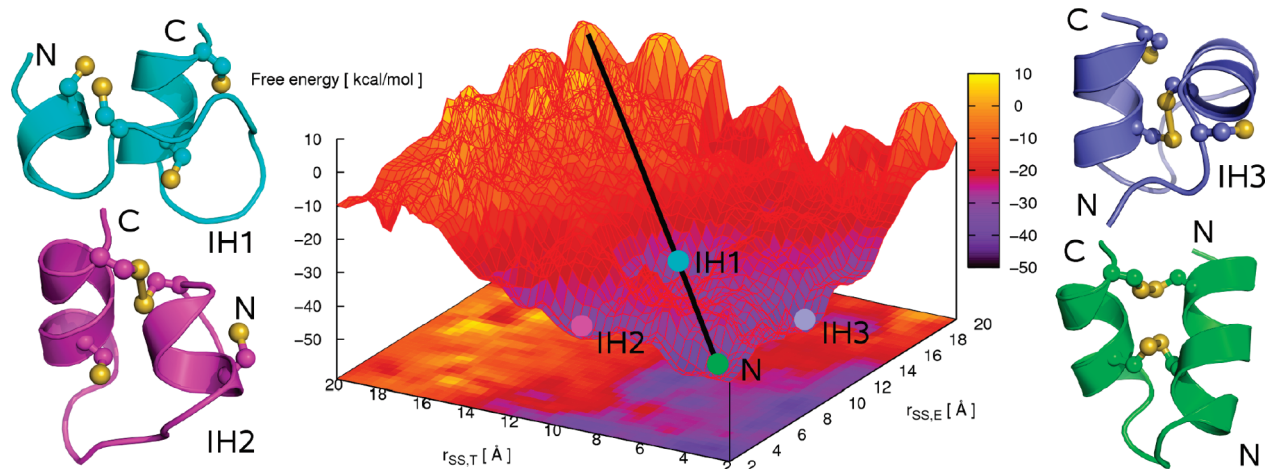
FIGURE 7: Minimum free energy landscape along the distances between sulfur atoms of bonds in the turn (T) and the end (E) of protein 1WQE. Cartoons of four selected conformations are colored cyan (IH1), magenta (IH2), blue (IH3), and green (N), and their locations on the surface are denoted by bullets of the same color. The possible reaction pathway is drawn schematically in black.

the global minimum (conformation N) via a free energy barrier of 9 kcal/mol. The barrier between basins IH1 and IH2 (colored cyan and magenta, respectively) at (2.9,8.9) is approximately 14 kcal/mol, while the barrier between IH2 and N is as small as 3 kcal/mol. A direct path connects structure IH1 with the global minimum crossing two barriers of approximately 10 and 6 kcal/mol.

Overall, the surface is thus characterized by an overall downhill gradient starting from the extended (X) and ending in the native conformation (N), but several pathways to reach this low-energy conformation seem possible. The time spent in each basin will depend on the barrier separating it from the adjacent basins. Because of the lower free energy of activation, the direct channel and the channel via structures with a disulfide bridge at the turn are presumably the more preferred folding pathways. In addition to a folding scenario, where both disulfide bridges form essentially simultaneously (schematically indicated by the black line on the surface), there are other pathways where one disulfide bond is formed preferentially first. This is in agreement with the observed correlation between the FNC and formation of a disulfide bridge. The figure thus clearly illustrates that 1WQE folds through several low-lying metastable conformations with at least one disulfide bond, but tertiary arrangements are inconsistent with the native fold. Thus, partially formed bonds must be broken, to reach the native conformation in a slow process on a rather rugged surface. The folding routes discussed above are summarized in Scheme 1.

For 1PG1 (see Figure 8) the extended conformation X is located at $(r_{SS,E}, r_{SS,T}) = (32.9, 19.9)$ and the native conformation at $(r_{SS,E}, r_{SS,T}) = (2.6, 2.6)$. The energy landscape reveals multiple wells and saddle points (transition states) situated on a path starting from the extended conformation (not shown) and leading to the native conformation. We shall start our discussion with the metastable ensemble at (6.5,7.2), a weakly populated state. Its corresponding lowest-energy conformation B4, colored blue, exhibits a well-formed turn and a partially formed β-sheet region (see Table 2), while disulfide interactions are almost absent ($\approx$0.15 kcal/mol). The basin at (6.5,7.2) is connected to another one at (2.5,9.8) via a transition state located near (5.0,7.4) with a barrier of approximately 15 kcal/mol. This step includes formation of the disulfide bond at the end and corresponds to the FNC range from 0.3 to 0.6 in Figure 5. This region is populated by

Scheme 1: Folding Pathways for 1WQE
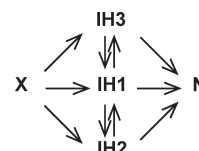
conformations U1, B2, H1, and H2. The lowest-energy conformation, U1, is colored magenta in Figure 8. The minimum at (2.5,9.8) is bridged with the global minimum by two transition states at (2.8,6.8) and (3.9,3.8) with barriers ~5 and 6 kcal/mol, respectively. These barriers correspond to the cleavage of the E-bond for FNC values between 0.6 and 0.8 shown in Figure 5. Nativelike conformations B1, B2, and B3 (Figure 6 and Table 2) are located in the funnel leading directly to the native conformation. The transitions discussed above can be summarized symbolically as shown in Scheme 2.

In contrast to 1WQE, 1PG1 exhibits energy differences as large as 10 kcal/mol between the metastable conformations. Although there is a steep decrease in the energy as a function of FNC, 1PG1 emerges as a prototype for a protein with a golf-course folding landscape. Its native conformation is stabilized by contact interactions [similar to Go models (59)], which are fully present only in the native ensemble. The folding process of 1PG1 is accompanied by two sequential stages of disulfide bond formation. In the first stage, only the E-bond is formed in an ensemble, which is incapable of accommodating the T-bond. In contrast to 1WQE, the formation of disulfide bonds for 1PG1 does not follow but rather precedes the formation of the secondary and tertiary structure.

DISCUSSION

One of the most intriguing questions regarding disulfide bridges is their interplay with other intramolecular interactions in the course of protein folding. In the classical Anfinsen scenario, the formation of the disulfide bonds follows after the complete secondary and tertiary structures are formed. Early reduction of the disulfide bonds can thus be used to halt the folding process (6, 7). In this view, disulfide bond formation improves the thermodynamic stability of the protein after the process of
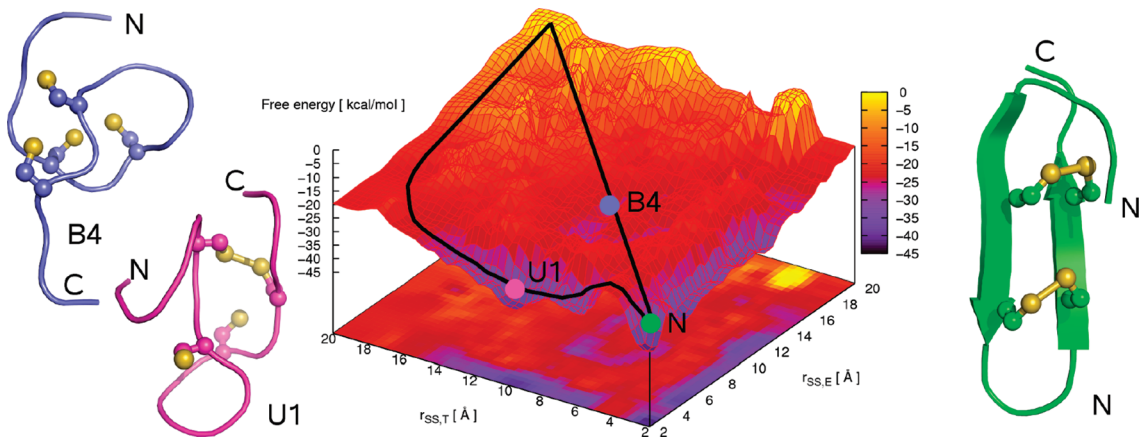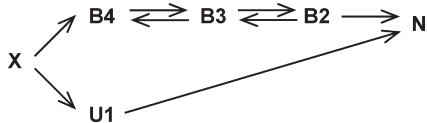
FIGURE 8: Minimum free energy landscape along the interatomic distances between the turn (T) and the end (E) sulfur atoms for protein 1PG1. Three selected conformations are colored blue (B4), magenta (U1), and green (N), and their locations on the surface are designated by bullets of the same colors.

Scheme 2: Folding Pathways for 1PG1

structure formation is essentially complete. The disulfide bond formation process can thus be viewed as a selection mechanism that significantly increases the probability of finding the protein in its native conformation.

In this investigation, we have sought to characterize the folding scenario for two peptides using stochastic all-atom simulations in a transferable force field. We have performed folding simulations of the antimicrobial peptide protegrin-1 1PG1 and the potassium channel blocker 1WQE starting from completely extended conformations. For both proteins, independent simulations identify the native conformation as the global minimum of the free energy landscape in the presence of a Morse potential modeling the disulfide bridges. These minimal energy conformations deviate by only 2.1 and 1.2 Å from the structurally conserved regions of the NMR models for 1WQE and 1PG1, respectively. On the basis of these de novo folding simulations, we have completely characterized the topology of the free energy surfaces of the two mini-proteins using connectivity graphs. While the free energy surface of both proteins is characterized by landscapes comprising just a few (order of 10) distinct structural families, the energetic arrangement and topological connection differ significantly. Projecting the free energy onto low-dimensional folding reaction coordinates, such as the fraction of native contacts or the disulfide bond distance, we can discuss likely folding scenarios.

For 1PG1, we found that disulfide bonds have a strong influence on the preferred secondary structure as was also indicated in ref 10. Thermodynamic stabilization due to disulfide bridges in 1PG1 is crucial for the preferred native conformation in 1PG1, and this has also been observed in other proteins (11). In good agreement with earlier work, we find that the rugged funnel folding model is not generally valid (7). In particular, disulfide bond formation either may be a prerequisite for conformational folding (as in the case of 1PG1) or can proceed in steps only after certain folding intermediates have been formed (as in the case of 1WQE). Similar implications of disulfide bonding for folding kinetics have been discussed recently in a review (8).

Our simulations suggest completely different folding scenarios for the two proteins: while the two-helix bundle 1WQE folds cooperatively via a long-lived intermediate in which the disulfide bridge near the sequence termini forms first, the β-hairpin follows a folding scenario in which a preformed disulfide bridge must be broken in the vicinity of the thermodynamic transition state to reach the native conformation.

Similar to proteins such as BPTI (12), the two-helix 1WQE folds through several intermediates, progressively realizing its native conformation. Folding of this protein thus appears to follow Anfinsen's scenario: the secondary structure forms early in the folding process, and various disulfide bridge arrangements exist in the low-energy part of the free energy surface. Some of these arrangements are in disagreement with the native tertiary structure, which is selected on the basis of intramolecular interactions and disulfide bond formation. In the majority of folding paths, disulfide content must be partially rearranged in the course of the folding process among low-energy conformations for the full formation of the native structure.

In contrast, the near-native ensemble of 1PG1 contains only conformations with the presence of stabilizing disulfide bridges. This protein rather follows a folding scenario very similar to that proposed recently for hirudin (12). First, 1PG1 collapses into an unstructured ensemble, where the loss of backbone entropy is compensated by the gain of a terminal disulfide bond. To form the correct secondary structure, this bond must be partially broken to generate the correct three-dimensional structure near the turn, including the local disulfide bond. Folding into the native ensemble thus follows a path, in the sense of Levinthal, via a well-characterized intermediate ensemble that is separated from the native ensemble by high-energy transition states.

Our investigation thus demonstrates how disulfide bridge interactions can modulate the folding scenario for small proteins to explore completely different folding mechanisms, ranging from Levinthal's scenario featuring a single dominating folding path to Anfinsen's scenario, which is conceptually realized in the "rugged funnel paradigm" for the folding of most globular proteins today.

ACKNOWLEDGMENT

REFERENCES

1. Levinthal, C. (1968) Are there pathways for protein folding? *J. Chim. Phys. 65*, 44 45.
2. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science 181*, 223 230.
3. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol. 4*, 10 19.
4. Onuchic, J. N., Luthey Schulten, Z., and Wolynes, P. G. (1997) Theory of protein folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem. 48*, 545 600.
5. Ainavarapu, S. R. K., Brujic, J., Huang, H. H., Wiita, A. P., Lu, H., Li, L., Walther, K. A., Carrion Vazquez, M., Li, H., and Fernandez, J. M. (2007) Contour Length and Refolding Rate of a Small Protein Controlled by Engineered Disulfide Bonds. *Biophys. J. 92*, 225 233.
6. Darby, N., and Creighton, T. E. (1997) Probing Protein Folding and Stability Using Disulfide Bonds. *Mol. Biotechnol. 7*, 57 77.
7. Wedemeyer, W. J., Welker, E., Narayan, M., and Scheraga, H. A. (2000) Disulfide Bonds and Protein Folding. *Biochemistry 39*, 4207 4216.
8. Wedemeyer, W. J., and Scheraga, H. A. (2000) Protein Folding: Overview of Pathways. *Encycl. Life Sci.,* 0003016.
9. Narayan, M., Welker, E., Wedemeyer, W. J., and Scheraga, H. A. (2000) Oxidative Folding of Proteins. *Acc. Chem. Res. 33*, 805 812.
10. Carrega, L., Mosbah, A., Ferrat, G., Beeton, C., Andreotti, N., Mansuelle, P., Darbon, H., de Waard, M., and Sabatier, J. M. (2005) The Impact of the Fourth Disulfide Bridge in Scorpion Toxins of the α KTx6 Subfamily. *Proteins: Struct., Funct., Bioinf. 61*, 1010 1023.
11. Doig, A. J., and Williams, D. H. (1991) Is the hydrophobic effect stabilizing or destabilizing in proteins? The contribution of disulphide bonds to protein stability. *J. Mol. Biol. 217*, 389 398.
12. Arolas, J. L., Aviles, F. X., Chang, J. Y., and Ventura, S. (2006) Folding of small disulfide rich proteins: Clarifying the puzzle. *Trends Biochem. Sci. 31*, 292 301.
13. Simmerling, C., Strockbine, B., and Roitberg, A. E. (2002) All Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc. 124*, 11258 11259.
14. Hubner, I. A., Deeds, E. J., and Shakhnovich, E. I. (2005) High resolution protein folding with a transferable potential. *Proc. Natl. Acad. Sci. U.S.A. 102*, 18914 18919.
15. Snow, C. D., Zagrovic, B., and Pande, V. S. (2002) The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations. *J. Am. Chem. Soc. 124*, 14548 14549.
16. Garcia, A. E., and Onuchic, J. N. (2003) Folding a protein in a computer: An atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. U.S.A. 100*, 13898 13903.
17. Duan, Y., and Kollman, P. A. (1998) Pathways to a Protein Folding Intermediate Observed in a 1 Microsecond Simulation in Aqueous Solution. *Science 282*, 740 744.
18. Simmerling, C., Strockbine, B., and Roitberg, A. (2002) All Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc. 124*, 11258 11259.
19. Snow, C. D., Nguyen, H., Pande, V. S., and Gruebele, M. (2002) Absolute Comparison of simulated and experimental protein folding dynamics. *Nature 420*, 102 106.
20. Yang, J. S., Chen, W. W., Skolnick, J., and Shakhnovich, E. I. (2007) All Atom Ab Initio Folding of a Diverse Set of Proteins. *Structure 15*, 53 63.
21. Herges, T., and Wenzel, W. (2004) An All Atom Force Field for Tertiary Structure Prediction of Helical Proteins. *Biophys. J. 87*, 3100 3109.
22. Herges, T., and Wenzel, W. (2005) In Silico Folding of a Three Helix Protein and Characterization of Its Free Energy Landscape in an All Atom Force Field. *Phys. Rev. Lett. 94*, 018101.
23. Verma, A., Schug, A., Lee, K. H., and Wenzel, W. (2006) Basin hopping simulations for all atom protein folding. *J. Chem. Phys. 124*, 044515.
24. Verma, A., Gopal, S. M., Oh, J. S., Lee, K. H., and Wenzel, W. (2007) All atom de novo protein folding with a scalable evolutionary algorithm. *J. Comput. Chem. 28*, 2552 2558.
25. Qin, M., Zhang, J., and Wang, W. (2006) Effects of Disulfide Bonds on Folding Behavior and Mechanism of the β Sheet Protein Tenda mistat. *Biophys. J. 90*, 272 286.
26. Martí Renom, M. A., Stote, R. H., Querol, E., Avilés, F. X., and Karplus, M. (1998) Refolding of Potato Carboxypeptidase Inhibitor by Molecular Dynamics Simulations with Disulfide Bond Con straints. *J. Mol. Biol. 284*, 145 172.
27. Czaplewski, C., Ołdziej, S., Liwo, A., and Scheraga, H. A. (2004) Prediction of the structures of proteins with the UNRES force field, including dynamic formation and breaking of disulfide bonds. *Protein Eng., Des. Sel. 17*, 29 36.
28. Abkevich, V. I., and Shakhnovich, E. I. (2000) What can Disulfide Bonds Tell Us about Protein Energetics, Function and Folding: Simulations and Bioninformatics Analysis. *J. Mol. Biol. 300*, 975 985.
29. Micheletti, C., de Filippis, V., Maritan, A., and Seno, F. (2003) Elucidation of the Disulfide Folding Pathway of Hirudin by a Topology Based Approach. *Proteins 53*, 720 730.
30. Huang, E. S., Samudrala, R., and Ponder, J. W. (1999) Ab initio fold prediction of small helical proteins using distance geometry and knowledge based scoring functions. *J. Mol. Biol. 290*, 267 281.
31. Muskal, S. M., Holbrook, S. R., and Kim, S. H. (1990) Prediction of the disulfide bonding state of cysteine in proteins. *Protein Eng. 3*, 667 672.
32. Kobayashi, Y., Sasabe, I., Akutsu, T., and Saito, N. (1992) Mechan ism of protein folding. IV. Forming and breaking of disulfide bonds in bovine pancreatic tripsin inhibitor. *Biophys. Chem. 44*, 113 127.
33. Watanabe, K., Nakamura, A., Fukuda, Y., and Saitô, N. (1991) Mechanism of protein folding: III. Disulfide bonding. *Biophys. Chem. 40*, 293 301.
34. Rey, A., and Skolnick, J. (1994) Computer simulation of the folding of coiled coils. *J. Chem. Phys. 100*, 2267 2276.
35. Quintilla, A., Starikov, E., and Wenzel, W. (2007) De novo Folding of Two Helix Potassium Channel Blockers with Free Energy Models and Molecular Dynamics. *J. Chem. Theory Comput. 3*, 1183 1192.
36. Quintilla, A., and Wenzel, W. (2007) Folding of Two Helical Peptide with Free Energy Methods and Molecular Dynamics. In NIC Series: From Computational Biophysics to Systems Biology (CBSB07) (Hansmann, U. H. E., Meinke, J., Mohanty, S., and Zimmermann, O., Eds.) pp 267−269, John von Neumann Institute for Computing, Julich, Germany.
37. Verma, A., and Wenzel, W. (2009) A free energy approach for all atom protein simulation. *Biophys. J. 96*, 3483 3494.
38. Eisenberg, D., and McLachlan, A. D. (1986) Solvation energy in protein folding and binding. *Nature 319*, 199 203.
39. Wenzel, W. (2006) Predictive folding of a β hairpin protein in an all atom free energy model. *Europhys. Lett. 76*, 156 162.
40. Verma, A., and Wenzel, W. (2007) Predictive and reproducible de novo all atom folding of a β hairpin loop in an improved free energy forcefield. *J. Phys.: Condens. Matter 19*, 285213.
41. Kondov, I., Verma, A., and Wenzel, W., manuscript in preparation.
42. Bogdan, T. V., Wales, D. J., and Calvo, F. (2006) Equilibrium thermodynamics from basin sampling. *J. Chem. Phys. 124*, 044102.
43. Nayeem, A., Vila, J., and Scheraga, H. A. (1991) A comparative study of the simulated annealing and Monte Carlo with minimization ap proaches to the minimum energy structures of polypeptides: [Met] enkephalin. *J. Comput. Chem. 12*, 594 605.
44. Abagyan, R. A., and Totrov, M. (1994) Biased Probability Monte Carlo Conformation Searches and Electrostatic Calculations for Peptides and Proteins. *J. Mol. Biol. 235*, 983 1002.
45. Wales, D. J., and Dewbury, P. E. J. (2004) Effect of salt bridges on the energy landscape of a model protein. *J. Chem. Phys. 121*, 10284 10290.
46. Mortenson, P. N., and Wales, D. J. (2004) Energy landscapes, global optimization and dynamics of poly alanine Ac(ala)$_8$NHMe. *J. Chem. Phys. 114*, 6443 6454.
47. Mortenson, P. N., Evans, D. A., and Wales, D. J. (2002) Energy landscapes of model polyalanies. *J. Chem. Phys. 117*, 1363 1376.
48. Kirkpatrick, S., Gelatt, C. D.Jr., and Vecchi, M. P. (1983) Optimiza tion by Simulated Annealing. *Science 220*, 671 680.
49. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers 22*, 2577 2637.
50. DeLano, W. L. (2002) The PyMOL Molecular Graphics System , DeLano Scientific, San Carlos, CA.
51. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact Order, Transition State Placement and the Refolding Rates of Single Domain Proteins. *J. Mol. Biol. 277*, 985 994.
52. Paci, E., Lindorff Larsen, K., Dobson, C. M., Karplus, M., and Vendruscolo, M. (2005) Transition State Contact Orders Correlate with Protein Folding Rates. *J. Mol. Biol. 352*, 495 500.
53. Feig, M., Karanicolas, J., and Brooks, C. L.III (2004) MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for appli cations in structural biology. *J. Mol. Graphics Modell. 22*, 377 395.
54. Brooks, C. L.III, Onuchic, J. N., and Wales, D. J. (2001) Taking a Walk on a Landscape. *Science 293*, 612 613.
55. Herges, T., and Wenzel, W. (2005) Free Energy Landscape of the Villin Headpiece in an All Atom Force Field. *Structure 13*, 661 668.

56. Chagot, B., Pimentel, C., Dai, L., Pil, J., Tytgat, J., Nakajima, T., Corzo, G., Darbon, H., and Ferrat, G. (2005) An unusual fold for potassium channel blockers: NMR structure of three toxins from the scorpion *Opisthacanthus madagascariensis*. *Biochem. J. 388*, 263 271.

57. Fahrner, R. L., Dieckmann, T., Harwig, S. S. L, Lehrer, R. I., Eisenberg, D., and Feigon, J. (1996) Solution structure of protegrin 1, a broad spectrum antimicrobial peptide from porcine leukocytes. *Chem. Biol. 3*, 543 550.

58. Onuchic, J. N., Luthey Schulten, Z., and Wolynes, P. G. (1997) Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem. 48*, 545 600.

59. Das, P., Matysiak, S., and Clementi, C. (2005) Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes. *Proc. Natl. Acad. Sci. U.S.A. 102*, 10141 10146.