

# Performance assessment of different constraining potentials in computational structure prediction for disulfide-bridged proteins

Ivan Kondov<sup>a,\*</sup>, Abhinav Verma<sup>a,1</sup>, Wolfgang Wenzel<sup>b</sup>

<sup>a</sup> Steinbuch Centre for Computing, Karlsruhe Institute of Technology, P.O. Box 3640, 76021 Karlsruhe, Germany

<sup>b</sup> Institute of Nanotechnology, Karlsruhe Institute of Technology, P.O. Box 3640, 76021 Karlsruhe, Germany

## ARTICLE INFO

### Keywords:

Protein simulation  
Disulfide bond  
Structural restraints  
Constraining potential

## ABSTRACT

The presence of disulfide bonds in proteins has very important implications on the three-dimensional structure and folding of proteins. An adequate treatment of disulfide bonds in de-novo protein simulations is therefore very important. Here we present a computational study of a set of small disulfide-bridged proteins using an all-atom stochastic search approach and including various constraining potentials to describe the disulfide bonds. The proposed potentials can easily be implemented in any code based on all-atom force fields and employed in simulations to achieve an improved prediction of protein structure. Exploring different potential parameters and comparing the structures to those from unconstrained simulations and to experimental structures by means of a scoring function we demonstrate that the inclusion of constraining potentials improves the quality of final structures significantly. For some proteins (1KVG and 1PG1) the native conformation is visited only in simulations in presence of constraints. Overall, we found that the Morse potential has optimal performance, in particular for the  $\beta$ -sheet proteins.

## 1. Introduction

Computational prediction of tertiary structure of proteins with high accuracy on the basis of primary structure, i.e. the amino-acid sequence, requires development of transferable protein force fields as well as powerful optimization methods. In particular, proteins containing disulfide bridges represent major challenge for computational biophysics. The folding process of several proteins without disulfide bridges could be studied in detail in simulations (Simmerling et al., 2002; Hubner et al., 2005; Snow et al., 2002b; Garcia and Onuchic, 2003) using all-atom protein force fields (Duan and Kollman, 1998), molecular dynamics (Simmerling et al., 2002; Snow et al., 2002a), replica exchange (Garcia and Onuchic, 2003), Monte Carlo (Hubner et al., 2005; Yang et al., 2007) or non-equilibrium methods (Herges and Wenzel, 2004; Herges and Wenzel, 2005; Verma et al., 2007; Verma et al., 2006). Recently, folding has been also studied using coarse-grained modelling and rate constant analysis (Carr and Wales, 2008; Carr and Wales, 2009). For a long time, the molecular dynamics simulations with atomic resolution had been limited to nano- and microsecond times scales and thus could not be employed for structure prediction purposes. Only recently, Shaw et al. (2010) succeeded to fold proteins, such as FIP35 and villin, from completely extended conformations and

studied the dynamics of the disulfide-bridged BPTI in all-atom molecular dynamics simulations on the millisecond time scale conducted on a special-purpose supercomputer.

Due to the long timescale of the folding process all-atom simulations have faced difficulties for proteins containing disulfide bridges (Mart-Renom et al., 1998; Qin et al., 2006). This is why there were only few works that studied the folding behavior of proteins with disulfide bonds by means of molecular dynamics (Mart-Renom et al., 1998; Qin et al., 2006), conformational space annealing with a united-residue force field (Czaplewski et al., 2004), lattice models (Abkevich and Shakhnovich, 2000), topology-based approach (Micheletti et al., 2003), distance geometry (Huang et al., 1999), neural networks (Muskal et al., 1990) and the island model (Kobayashi et al., 1992; Watanabe et al., 1991). In other works (Chen and Hwang, 2005; Fiser et al., 1992; Fiser and Simon, 2000; Mucchielli-Giorgi et al., 2002; O'Connor and Yeates, 2004) methods and services for qualitative prediction of occurrence and location of disulfide bonds have been proposed. In the recent two years however, due to rapid development of molecular dynamics methods as well as increase of computing power, there was an increased number of all-atom molecular dynamics studies of disulfide-bonded proteins (Allison et al., 2010; Aschi et al., 2010; Laghaei et al., 2010; Sharadadevi and Nagaraj, 2010; Zhang et al., 2009). In particular, the impact of disulfide bonds on the secondary structure has been demonstrated for Hepsidin-20 and Hepsidin-25 (Aschi et al., 2010), for the human Amylin (Laghaei et al., 2010), and for mammalian Defensins HBD-1 and HNP-3 (Sharadadevi and Nagaraj, 2010). For instance, in the latter work it was found that for both HBD-1 and HNP-3 the N-terminal  $\beta$ -strands unfold in the absence of disulfide

\* Corresponding author.

E-mail address: [ivan.kondov@kit.edu](mailto:ivan.kondov@kit.edu) (I. Kondov).

<sup>1</sup> Present address: Centro de Investigaciones Biológicas Spanish Research Council (CSIC), Madrid, Spain.

bonds. In a replica-exchange study of human Amylin (Laghaei et al., 2010) it was shown that the absence of the disulfide bond greatly decreases the extent of helix formed throughout residues 5–9 in favor of random coil and  $\beta$ -sheet structure thus indicating a helix-stabilization due to the disulfide bond.

While disulfide bonds stabilize the protein in the thermodynamic sense, they may impede the folding process by stabilizing misfolded intermediates (Rey and Skolnick, 1994), especially in early folding stages (Abkevich and Shakhnovich, 2000). In the latter study it was found that stiffness of polypeptide chain due to constrained motion results in smaller folding rates. Specifically, disulfide bonds that accelerate folding seem to be mostly formed in early stages of 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.

Nevertheless, studies based on molecular dynamics often do not account for the dynamic formation of disulfide bonds during the folding process, i.e. they assume fixed disulfide bonds during the course of simulated folding. Moreover, atomistic treatment of disulfide-bond formation has been also insufficient in algorithms for protein-structure prediction. In a recent study, employing the conformational space annealing method and an united-atom force field (Czaplewski et al., 2004) various proteins with disulfide bridges were investigated in which 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.

Using transferable all-atom protein force fields and stochastic optimization methods the native conformations of a number of proteins have been predicted within experimental resolution (Herges and Wenzel, 2004; Herges and Wenzel, 2005; Verma et al., 2006). Recently, the approach has been extended to take into account different secondary structures (Strunk et al., 2009; Verma and Wenzel, 2009; Wenzel, 2006). A further extension of this efficient all-atom structure prediction technique aiming to treat and correctly predict disulfide-bonded proteins motivated us to carry out the present study.

We have recently investigated the potassium channel blocker 1WQE containing two disulfide bonds using a combination of the free-energy protein force field PFF01 and molecular dynamics simulations (Quintilla et al., 2007; Quintilla and Wenzel, 2007) under reducing conditions. While failing to stabilize the native configuration, the protein exhibited significant native secondary structure content and even visited near-native configurations. Later, we extended our approach to study folding of disulfide-bridged proteins under oxidizing conditions, i.e. including the long-timescale process of disulfide bond formation. In a more recent study (Kondov et al., 2009) we have investigated the folding mechanism of two disulfide-bonded proteins, the 1WQE and the 18-residue antimicrobial peptide protegrin-1 1PG1, in all-atom basin hopping simulations starting from completely extended conformations. The minimal-energy conformations deviated 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 revealed that the folding mechanism of disulfide-bridged proteins can vary dramatically from Levinthal's single-path scenario to a cooperative process more consistent with the funnel paradigm of protein folding.

In this paper, we report on all-atom stochastic-search simulations of different proteins containing disulfide bonds (DSB) with inclusion of an additional constraining potential as energy term.

In particular, we will focus on the efficiency assessment of different types of constraining potentials by comparing to experimental structures and structures found in absence of DSB constraints. The paper is organized as follows. Section 2 will provide an introduction the physical model and the computational methods employed followed by Section 3, in which, exploring DSB proteins of different types and size in a presence of different DSB constraining potentials, we will assess the applicability and the efficiency of the constraining potentials and their implications on the computationally predicted structures. In the last section we will conclude with a summary. We will show that inclusion of a constraining potential improves the best final structures significantly compared with those from a constraint-free simulations and, moreover, it is essential for correct prediction of the native structure of two of the studied proteins.

## 2. Methodology

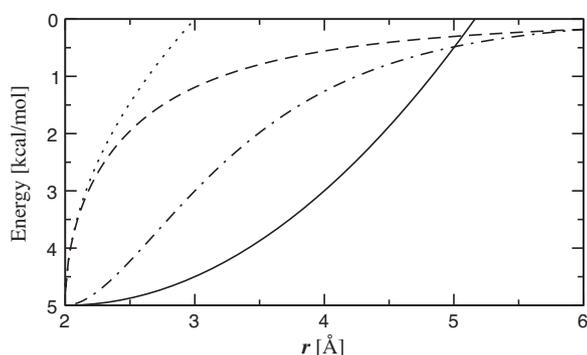
Recently we have followed a free-energy approach to protein simulation and structure prediction and developed methods based on an all-atom forcefield including the most important physical interactions in proteins. This method (Verma and Wenzel, 2009) is based on Anfinsen's thermodynamic hypothesis (Anfinsen, 1973) suggesting that many proteins in their native conformations are in thermodynamic equilibrium with their environment and the native conformation. As a consequence, the native conformation of a protein corresponds to the global minimum of its free-energy surface. Within the solvent accessible surface area (SASA) model (Eisenberg and McLachlan, 1986) each protein backbone conformation is assigned an "internal free energy", such that the 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. Thus, we can use any sampling technique, including non-equilibrium methods, such as the basin hopping technique (BHT), to simulate a protein conformational ensemble, until the low-energy region of the free-energy surface including the native conformation is sufficiently sampled. Most of the methods outlined below, in particular the force field, the simulated annealing and the DSB constraining potentials, have been implemented in the program code POEM (Herges and Wenzel, 2004) that is developed at the Karlsruhe Institute of Technology. A more detailed description of the methodology used in this work has been done recently by Strunk et al. (2009).

### 2.1. Force field

The free-energy protein force fields PFF01 (Herges and Wenzel, 2004; Herges and Wenzel, 2005) and PFF02 (Verma and Wenzel, 2009; Wenzel, 2006) have been developed to describe the free-energy surface. The force field PFF01 comprises following non-bonding interactions:

$$E = \sum_{ij} V_{ij} \left[ \left( \frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}}{r_{ij}} \right)^6 \right] + \sum_{ij} \frac{q_i q_j}{\epsilon_{g_i g_j} r_{ij}} + \sum_i \sigma_i \cdot A_i + \sum_{\text{H-bonds}} V_{\text{hb}} \quad (1)$$

where  $r_{ij}$  denote the inter-atomic 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 database. The second, electrostatic term depends on the partial atomic charges  $q_i$  and pairwise dielectric constants  $\epsilon_{g_i g_j}$ , with  $g_i$  being the type of amino acid residue pertaining to atom  $i$ . The last two terms provide implicit description of solvent interaction



**Fig. 1.** Constraining potentials for disulfide bonds as functions of the distance  $r$  between involved sulfur atoms: square-root potential  $V_{SS}(r) = E_0(\sqrt{|r-r_0|} - 1)$  (Sqrt, dotted line), harmonic potential  $V_{SS}(r) = -E_0 + 0.5\beta(r-r_0)^2$  (solid line), the hyperbolic tangent potential  $V_{SS}(r) = E_0 \left[ \tanh \sqrt{|r-r_0|} - 1 \right]$  (Tanh, dashed line), the Morse potential  $V_{SS}(r) = E_0 \left[ (1 - e^{-\beta(r-r_0)})^2 - 1 \right]$  (dashed-dotted line). For all shown potentials  $\beta = 1 \text{ \AA}^{-1}$  ( $\text{kcal mol}^{-1} \text{ \AA}^{-2}$ ),  $r_0 = 2 \text{ \AA}$  and  $E_0 = 5 \text{ kcal/mol}$ .

(the SASA model (Eisenberg and McLachlan, 1986)), where  $A_i$  is the contact area for atom  $i$ , and the short-range backbone-backbone hydrogen bonding  $V_{hb}$ , respectively.

While PFF01 was parameterized for helical proteins, its extension PFF02 (Verma and Wenzel, 2007; Verma and Wenzel, 2009; Wenzel, 2006) includes terms that allow thermodynamically correct description of  $\beta$ -sheets and proteins of mixed secondary structure. The first additional term differentiates between the backbone dipole alignments found in different secondary structure elements included in the electrostatic potential between atoms  $i$  and  $j$  belonging to the backbone NH or CO groups via the dielectric constants  $\epsilon_{g_i, g_j}$ . The second, torsional potential term for backbone dihedral angles gives a small contribution (about 0.3 kcal/mol) to stabilize conformations with dihedral angles in the  $\beta$ -sheet region of the Ramachandran plot. To date, over 20 proteins (Verma and Wenzel, 2007) with 20–60 amino acids could be simulated starting from extended conformations to structures of an average backbone root-mean-square deviation (RMSD) of 2.8 Å. Thus, the PFF02 provides an unbiased and transferable description of the free-energy landscape for proteins containing both  $\beta$ -sheet and  $\alpha$ -helix secondary structures. In this work, the force field PFF02 will be used.

## 2.2. Constraining potentials

To promote the formation of disulfide bridges we employ constraining potentials of different types as plotted in Fig. 1. In a recent study, Czaplewski et al. (2004) employed a harmonic potential biased by 5.5 kcal/mol to describe the disulfide bond with equilibrium distance of 4.7 Å defined between the centers of mass of pertinent cysteine residues. In our study we also use harmonic potential but we define the equilibrium distance of  $r_0 = 2.0 \text{ \AA}$  between the sulfur atoms of pertinent cysteine residues. Because the potential increases unlimitedly with distance  $r$  between sulfur atoms we defined a cut-off for all  $r$  for which the potential is positive.

Morse potential  $V_{SS}(r) = E_0 \left[ (1 - e^{-\beta(r-r_0)})^2 - 1 \right]$  also was considered, where  $r_0$  is the equilibrium distance between the sulfur atoms forming a disulfide bridge and  $-E_0$  is the energy corresponding to  $r_0$ . The spacial extent of the potential  $\beta$  has the meaning of the inverse distance at which the potential well is one half. For smaller  $\beta$  values the potential is more extended and for larger  $\beta$  more compact. Morse potential has been widely used to describe chemical bonds in classical forcefields, as well as in studies of atomic clusters (Miller et al., 1999; Wales, 2010).

In addition, two other potentials, square root (denoted as Sqrt in the following) and hyperbolic tangent containing a square root (denoted as Tanh in the following). The Sqrt and Tanh are characterized by very small spatial extents (about 3 Å) and singularity at the equilibrium distance  $r_0$ . On the other hand, the harmonic and Morse potentials have a smooth minimum at  $r_0$  and have broader wells extending to different ranges depending on  $\beta$ , e.g. to 4–5 Å for  $\beta = 1$  and  $E_0 = 5 \text{ kcal/mol}$  (cf. Fig. 1). These properties imply larger binding forces around the minima of the Sqrt and Tanh potentials compared to the harmonic and Morse potentials. Another important property of the binding potentials is their asymptotic behavior. The Morse and Tanh potentials decay exponentially to zero and remain negative for long distances. In contrast, the Sqrt and the harmonic potentials become infinitely large for large distances and thus have to be cut off at some value. In this study, we cut off the harmonic potential for distances for which it becomes positive.

In the simulations presented here, we varied the parameters  $E_0$  and  $\beta$  for all constraining potentials and  $r_0$  was kept fixed at 2.0 Å.

## 2.3. Simulation method

The BHT (Li and Scheraga, 1987; Nayeem et al., 1991; Verma et al., 2006; Wales and Doye, 1997) employs a relatively straightforward approach to eliminate high-energy transition states of the free-energy surface: The original free-energy surface is simplified by replacing the energy of each conformation with the energy of an adjacent 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 (Abagyan and Totrov, 1994) have been used previously to study the free-energy surface of model proteins (Wales and Dewbury, 2004) and poly-alanines using all-atom models (Mortenson et al., 2002; Mortenson and Wales, 2004), as well as for protein structure prediction (Prentiss et al., 2008) and protein folding studies (Carr and Wales, 2005).

Here we replace the gradient-based minimization step with a simulated annealing run (Kirkpatrick et al., 1983), because local minimization generates only very small steps on the free-energy surface. In addition, the calculation of gradients for the employed SASA model (Eisenberg and McLachlan, 1986) is computationally prohibitive. Within each annealing simulation, new configurations are accepted according to the Metropolis criterion, while the temperature is decreased geometrically from its starting 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 parameterized by the starting temperature ( $T_S$ ), the final temperature ( $T_F = 2 \text{ K}$ ) and the number of steps. We investigated various choices for the numerical parameters of the method, but have always used a geometric cooling schedule (Verma et al., 2006). In every cycle the initial temperature  $T_S$  was chosen from an exponential distribution and the number of simulated annealing steps was increased following (Verma et al., 2006).

At the end of one annealing cycle the new conformation is accepted if its energy difference to the current configuration is not higher than a given threshold energy. Throughout this study we use a threshold acceptance criterion of 1 kcal/mol. For each disulfide parameter set we performed 30 independent runs starting the extended conformation (i.e. all backbone dihedral angles, except for proline residues, were set to 180°) each containing 300, 300, 600 and 800 basin hopping cycles for 1KVG, 1WQE, 1PG1, and 1HD6 respectively. For the three-helical 1HD6 we used a slightly mod-

ified simulation protocol in which we performed two annealing cycles per BHT cycle.

In addition, 30 simulations of length 300 cycles have been performed for each protein starting from the NMR structure of each protein under the conditions outlined above. The first structures in the PDB records of the respective proteins were used as references. The sets of accepted structures of all runs were used in the further analysis in which, e.g. the root mean square deviation (RMSD) was calculated for the backbone atoms. The structures in this work were visualized with PyMOL (Delano, 2002).

#### 2.4. Secondary structure

To determine secondary structure we used the DSSP program (Kabsch and Sander, 1983). Then we applied a simple per-residue scoring function to calculate the match between the secondary structure of the simulated conformations and that from the PDB. In this function, every symbol from the DSSP output string is compared with the corresponding symbol in the experimental structure (see the caption of Table 1 for explanation of symbols). If the symbols match then a unity, otherwise zero, is added to the score. At the end, the secondary structure score (SSS) is normalized to unity (100%).

#### 2.5. Native contacts

For the lowest-energy conformation from each simulation the fraction of native contacts (FNC) was computed. While the backbone RMSD for backbone atoms is characteristic for the quality of the secondary structure and for extent of mutual alignment of helical regions, native contacts characterize the mutual alignment of side chains in the tertiary native structure. Therefore, the contact order and the FNC have been used as folding reaction coordinates (Kondov et al., 2009; Paci et al., 2005; Plaxco et al., 1998). In this study, the FNC will be used to estimate the quality of the simulated structures with respect to the alignment of side chains. We have used the MMTSB tool set (Feig et al., 2004) to calculate the residue-residue native contacts basing on distances between heavy atoms. A threshold distance of 4.2 Å between side-chain heavy atoms was set.

#### 2.6. Performance assessment

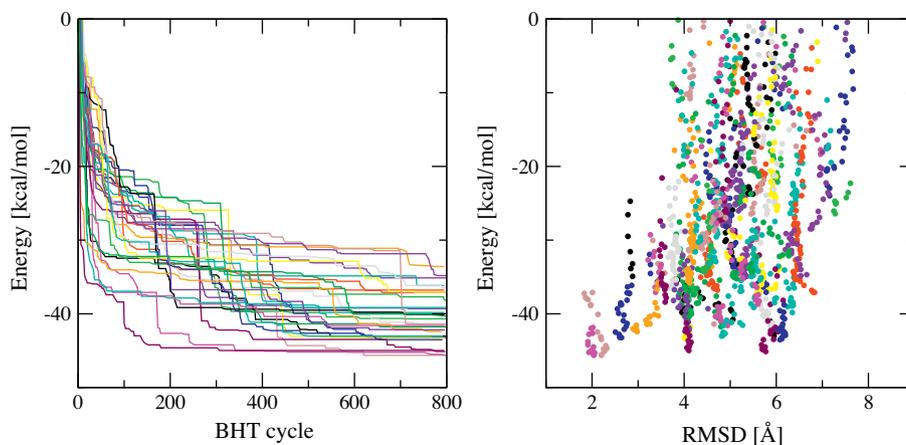
To assess the performance of every different potential for different parameters we introduce an unbiased metric (total score) defined basing on the pure PFF02 energy, i.e. the total energy subtracting the contributions of DSB constraining potentials,  $E_{DSB}$ . For each simulation set (for given potential type and parameter set) we select the structure with lowest total energy for which we calculate a total score

$$\text{Score} = \frac{E_{sc} \times SSS \times FNC}{\text{RMSD} \times \sum_i d_i/n} \times 1000 \quad (2)$$

where  $E_{sc} = |E_{PFF02}/E_{\min}, PFF02|$ ,  $d_i$  is the length for the  $i$ th DSB out of  $n$  DSBs. According to this definition, the higher values of  $E_{sc}$  (being optimal at unity) correspond to better protein structures. This definition of the total score allows comparison across simulations for different proteins as well as with different constraining potential type or parameters. Additionally, it captures all quality information including the RMSD, SSS, FNC and the DSB distances into one single quantity. For each constraining potential type the set with the highest total score was selected and used in further analysis.

**Table 1** Characteristics of the best final conformations obtained from simulations starting from the extended conformations with different DSB constraining potentials. Disulfide bond distances are given in Å. All energies are given in kcal/mol. We used following notation for secondary structure: C = coil, E =  $\beta$ -sheet, T = turn, S = bend, H =  $\alpha$ -helix, B =  $\beta$ -bridge.

Protein	Potential	$E_0$	$\beta$	Secondary structure	$E_{tot}$	RMSD	SSS	FNC	$d_1$	$d_2$	$d_3$	$E_{DSB}$	$E_{PFF02}$	$E_{sc}$	Score
1WQE	Harmonic	2	0.5	CHHHHHHHHSCCHHHHHHHHC	-44.5	2.1	0.88	0.33	3.0	3.4		-3.3	-41.2	0.94	41
	Morse	5	0.5	CHHHHHHHHSHCHHHHHHHHC	-47.4	2.2	0.88	0.50	2.9	3.0		-8.5	-38.9	0.89	61
	Sqrt	5		CHHHHHHHHSCCHHHHHHHHC	-35.8	2.5	0.83	0.25	2.6	2.7		-1.9	-33.9	0.77	24
	Tanh	5		CHHHHHHHHHTCCHHHHHHHHC	-44.5	2.0	0.96	0.58	2.8	2.8		-2.9	-41.6	0.95	97
	None			CHHHHHHHHHTCCHHHHHHHHC	-42.6	2.1	0.96	0.58	3.3	5.5		0.0	-42.6	0.99	59
1HD6	Harmonic	5	1	CHHHHHHHHSCBSSSCHHHHHHSCSCCC	-62.5	5.6	0.50	0.19	2.8	2.8	2.9	-13.9	-48.6	0.85	5
	Morse	5	1	CHHHHHHHHSCBSSSCHHHHHHSCSCCC	-60.1	7.8	0.32	0.26	8.3	2.7	2.7	-7.5	-52.6	0.92	2
	Sqrt	2		CEESSITCEBECCHHHHSCSCBSCSCCC	-49.2	6.5	0.13	0.26	3.2	2.9	3.3	0.3	-49.5	0.86	1
	Tanh	5		CCSCHHHHHCHHHHHHHHHHHHHHHTTTC	-57.9	4.8	0.63	0.10	2.8	12.3	4.6	-1.9	-56.0	0.98	2
	None			CHHHHHHSCCHHHHHHHHCBTTTBTSCSCSCCC	-57.3	5.1	0.39	0.19	8.4	6.0	5.3	0.0	-57.3	1.00	2
1PG1	Harmonic	2	1	CTTSCSEETEEESSCC	-42.1	6.8	0.58	0.44	2.9	3.0		-3.2	-39.0	0.89	11
	Morse	5	1	CCSSEEEETEEEECC	-48.2	3.8	0.89	0.72	2.7	2.7		-7.6	-40.6	0.93	59
	Sqrt	5		CCSSEEEETEEEECC	-42.1	3.8	0.89	0.72	2.7	2.6		-2.0	-40.1	0.92	58
	Tanh	5		CCSEEESSCEEECC	-39.8	4.3	0.74	0.72	5.9	2.7		-1.7	-38.1	0.87	25
	None			CTTCHHHHHTCCSSCC	-39.5	7.3	0.21	0.33	8.5	3.2		0.0	-39.5	0.90	1
1KVG	Harmonic	2	2	CEEEETEEEECC	-8.5	2.1	1.00	0.64	2.7			-1.5	-7.1	0.79	86
	Morse	10	1	CEEEETEEEECC	-14.4	2.1	1.00	0.73	2.6			-8.2	-6.2	0.70	93
	Sqrt	10		CEEEETEEEECC	-8.9	2.1	1.00	0.55	2.4			-3.3	-5.6	0.62	67
	Tanh	5		CBCBTTTBSCBC	-9.9	2.3	0.38	0.64	2.7			-1.5	-8.4	0.94	36
	None			CBCBTTTBSCBC	-8.8	2.4	0.38	0.64	3.1			0.0	-8.8	0.99	33



**Fig. 2.** Best-energy decrease with basin hopping step (left) and a population diagram (right) of the energy versus backbone RMSD for 30 independent simulations for protein 1WQE using the Morse potential with  $E_0 = 2$  kcal/mol and  $\beta = 0.5 \text{ \AA}^{-1}$ .

### 3. Results and discussion

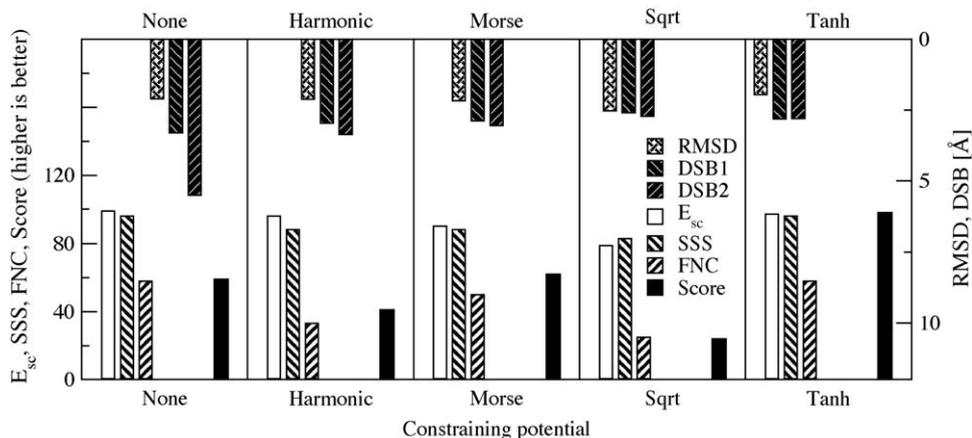
#### 3.1. $\alpha$ -Helical proteins

##### 3.1.1. 1WQE

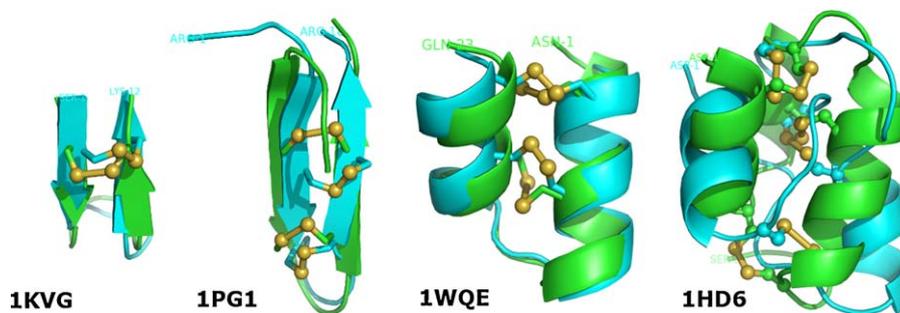
The 23-amino-acid potassium channel blocker 1WQE (Chagot et al., 2005) comprises two  $\alpha$ -helices (PRO3–HIS11 and VAL15–CYS22) connected by a loop region. The native conformation is characterized by two disulfide bonds between the residues CYS4 and CYS22 (DSB1), and CYS8 and CYS18 (DSB2), respectively. Previously, this protein has been investigated in a combination of free-energy and molecular dynamics simulations (Quintilla et al., 2007; Quintilla and Wenzel, 2007). These simulations were performed under reducing conditions, i.e. they did not employ a DSB potential. Therein, the free-energy protocol yielded a low-energy conformation of 1WQE with correct secondary structure, while failing to fully satisfy the disulfide constraints. The sulfur atoms in the lowest-energy conformation had a distance of  $d_1 = 3.3 \text{ \AA}$  and  $d_2 = 5.5 \text{ \AA}$ , compared to  $2 \text{ \AA}$  in the native conformation for both. Subsequent molecular dynamics simulations (Quintilla et al., 2007; Quintilla and Wenzel, 2007) started from this low-energy conformation suggested that the secondary structure elements are stable. The tertiary arrangement of the helices fluctuated significantly, but visited conformations consistent with the correct disulfide bonding pattern on a short ( $\sim 50$  ns) timescale. Later, it was shown (Kondov et al., 2007, 2009)

that near-native conformations of 1WQE could be simulated with and without constraining Morse potential. Moreover, it was found (Kondov et al., 2009) that 1WQE follows the Anfinsen’s funnel folding scenario (Dill and Chan, 1997; Onuchic et al., 1997), i.e. the folding pathway has only local low barriers and a global slope leading to the native state. The DSB formation in such scenario follows continuously the build-up of secondary and tertiary structures and the protein reaches its native conformation in the absence of DSB constraints. In the case with constraints, intermediates with near-native conformations with partially closed DSBs could be identified.

In the following, we study the protein in the presence of the disulfide bridge potential described above in independent basin hopping simulations. All independent simulations start from the completely extended conformation at very high energy (cf. Fig. 2). After every BHT cycle the new structure is accepted either if the energy is decreasing leading the simulation to exploit the improved structure or if the energy is increased up to a certain threshold (see Section 2.3) allowing exploration of larger areas of the conformation space. On average, best energies decrease for all simulations as seen in Fig. 2. At the end of the simulation four simulations have reached lowest values indicating convergence. However, not all of these four best-energy conformations can be attributed to the global minimum. As seen in Fig. 2 (right panel) the first two conformations with lowest energies have RMSD values of around  $2 \text{ \AA}$ . The next two have RMSDs about  $4$  and  $6$ , respectively, being far



**Fig. 3.** Performance analysis based on best-energy conformations for the protein 1WQE. The bars on the bottom (left axis) depict the  $E_{sc}$ , SSS, FNC and the Score according to Eq. (2). The bars descending from the top (right axis) denote the RMSD and the DSB distances.



**Fig. 4.** Cartoons of the best simulated protein structures starting from extended conformations (shown in green) superimposed on the corresponding native structures from PDB (shown in cyan). Sulfur atoms of cysteine residues forming disulfide bridges are depicted as spheres in golden color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

from the native conformation. Therefore, it is always important to carry out more than one independent simulation in order to capture the whole ensemble of conformation with energies close to that of the native state. Moreover, the backbone RMSD in respect to the experimental structure should be considered, besides the energy, in order to more properly assess the quality of the final structures.

Analyzing all accepted structures from all simulations and for each DSB potential the best structures were determined and characterized. These structures and their characteristics are summarized in Table 1 and depicted in Fig. 3.

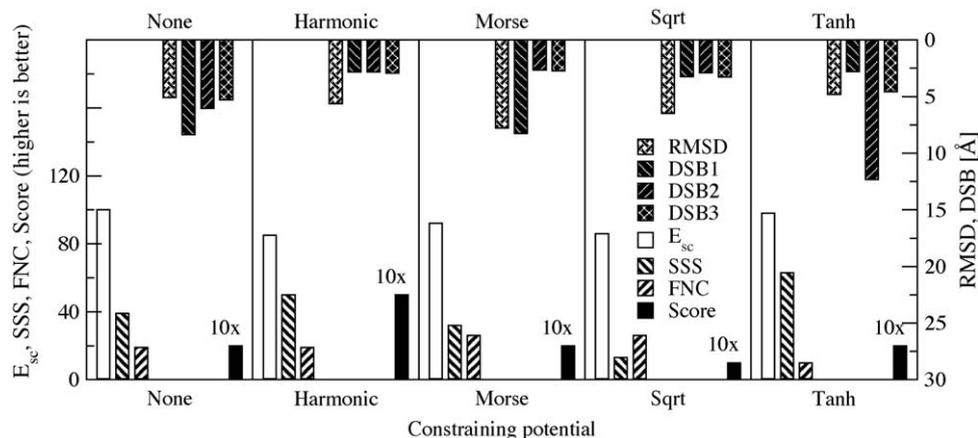
For the protein 1WQE we found a lowest-energy conformation with RMSD 2.0 Å with the Tanh constraining potential that is shown in Fig. 4. The helical regions and their alignment are well reproduced for both with and without constraining potential. In particular, the obtained secondary structure scores are equal and near unity, the fractions of native contacts are the same and the RMSD of the structure from the simulation without DSB potential is only slightly larger (2.1 Å) than that with the Tanh potential (2.0 Å, cf. Table 1). Nevertheless, the DSB distances for the unconstrained structure, 3.3 and 5.5 Å, are much larger than those of the best structure with Tanh, 2.8 and 2.8 Å, respectively. This gives rise to a large total score of 97 for the Tanh potential and only 59 for the simulation without constraints. With score of 61 the Morse potential shows the second best performance after the Tanh potential. Because the Morse potential around the minimum is flat, the formed DSBs are loose, 2.9 and 3.0 Å, compared to those with the Tanh potentials. A comparison of the best-energy structures with the Morse and the Tanh potentials reveals that the tight constraining potential Tanh improves not only the DSB distances but also the secondary structure (SSS), helices alignment (RMSD) and alignment of side chains (FNC) and hence stabilizes the native state of 1WQE.

Again due to the flatness, larger DSB distances, 3.0 and 3.4 Å, were obtained with the harmonic potential. The Sqrt potential yields an overall quality of the best final structure (score of 24) which is inferior to the best structure obtained without constraints but results in best DSB distances. In summary, DSB constraining potentials, especially such with asymptotic decay, improve the efficiency of the stochastic search for the global minimum significantly. On the other hand potentials with steep increase near the minimum (“tight” potentials) improve the description of DSBs. For instance, the square-root potential (Sqrt) yields the best values for the DSB distances. However, the latter constraining potential affects all other metrics and the energy, and thus has total score even lower than the score of the best constraint-free structure (cf. Fig. 3).

### 3.1.2. 1HD6

The pheromone Er-22 isolated from *Euplotes raikovi* with PDB code 1HD6 (Liu et al., 2001) consists of an anti-parallel bundle of three helices comprising residues ILE2–GLN9, LEU12–CYS18 and THR21–SER31 with nearly parallel axes (cf. Fig. 4). As was already indicated by (Liu et al., 2001) the first and the third helix are regular  $\alpha$ -helices, whereas the second helix is a continuous stretch of distorted  $3_{10}$ -helix turns. The distortion of the middle helix ensures that CYS15 and CYS18 are properly oriented to form two DSBs with CYS24 (DSB3) and CYS3 (DSB1), respectively, of the other two helices. Another disulfide bond, DSB2, between CYS10–CYS32 connects the end of the first with the end of the third helix.

For all DSB potentials (each with  $E_0 = 2$  and  $E_0 = 5$  kcal/mol) and without DSB potential we performed 100 independent simulations of 1HD6 starting from the extended conformation each 800 BHT cycles long. For each parameter set we also carried out 30 independent BHT relaxations of the NMR structure from the PDB with



**Fig. 5.** Performance analysis based on best-energy conformations for the protein 1HD6. The bars on the bottom (left axis) depict the  $E_{sc}$ , SSS, FNC and the Score according to Eq. (2). The bars descending from the top (right axis) denote the RMSD and the DSB distances.

**Table 2** Characteristics of the best final conformations of 1HD6 obtained from simulations starting from the native conformation with different DSB constraining potentials. Disulfide bond distances are given in Å. All energies are given in kcal/mol. See Table 1 for secondary structure notation.

Potential	$E_0$	$\beta$	Secondary structure	$E_{\text{tot}}$	RMSD	SSS	FNC	$d_1$	$d_2$	$d_3$	$E_{\text{DSB}}$	$E_{\text{PFF02}}$	$E_{\text{sc}}$	Score
Harmonic	5	1	CCCHHHHTTCHHHHHSSCHHHHHHHHHHHSSCSCC	-67.7	3.5	0.79	0.61	2.8	2.8	3.0	-4.7	-63.0	1.00	49
Morse	5	1	CHHHHHHTTCHHHHHSSCHHHHHHHHHHHSSSTTC	-63.0	2.1	0.79	0.74	2.7	2.7	2.7	-11.4	-51.6	0.82	85
Sqrt	5		CHHHHHHTTCHHHHHSSCHHHHHHHHHHHSSSTTC	-48.4	1.8	0.79	0.71	2.7	2.7	2.8	-2.2	-46.2	0.73	84
Tanh	5		CHHHHHHTTCHHHHHSSCHHHHHHHHHHHSSCTTTC	-59.8	6.7	0.76	0.68	3.0	11.3	3.0	-2.5	-57.3	0.91	12
None			CHHHHHHTTCHHHHHSSCHHHHHHHHHHHSSCTTTC	-59.7	6.9	0.74	0.45	3.5	11.3	4.6	0.0	-59.7	0.95	7

length 300 cycles. All simulations of 1HD6 have consumed approximately 38 CPU years on a cluster with Intel Xeon E5430 processors. Again, for 1HD6 we detected and analyzed the best structures as shown in Fig. 5.

In absence of constraining potential an incorrect secondary structure consisting of two extended helical regions is obtained. Only 39% of the secondary structure is consistent with that of the native structure. The inclusion of DSB constraints does not improve significantly the best final structures. With the Tanh potential we obtained a structure with lowest unbiased energy  $E_{\text{PFF02}}$ , the best RMSD below 5 Å and high SSS (cf. Table 1, Fig. 5). However, the DSB2 in this structure is not present (DSB distance of 12.3 Å) and the fraction of native contacts is very small (10%). The structure with the harmonic potential has higher total score because all three DSB distances are smaller than 3 Å and the fraction of native contacts is higher. The latter structure is shown in Fig. 4. The total score of the best structure for 1HD6 is an order of magnitude smaller than the scores for the other three proteins indicating the overall failure to find the native conformation of this protein starting from the extended conformation.

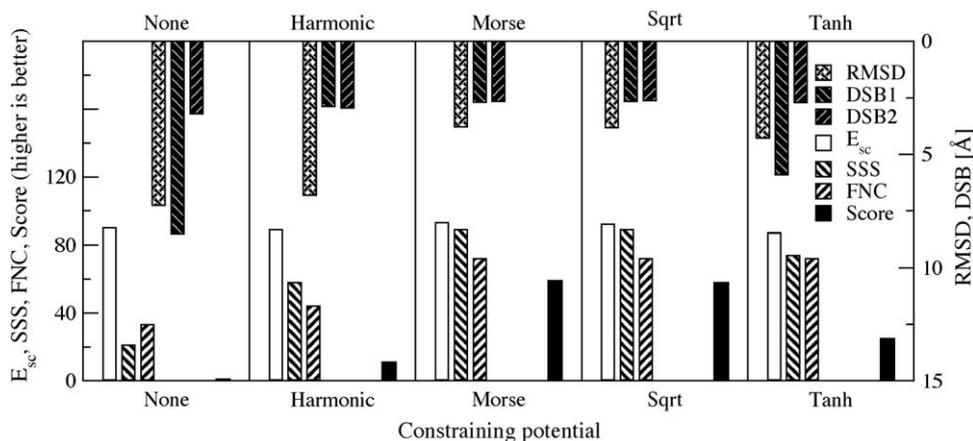
All simulations starting from the NMR structure yielded best structures that, apart from the unstructured proline-dominated region PRO33-SER37, agree well with the input structure preserving the three helical regions and their relative orientation. These structures are characterized by SSS from 0.74 up to 0.79 (cf. Table 2) and total scores always higher than for those obtained from the extended conformation. Also, the best energies (with and without constraining potentials) from these simulations are lower than those in the simulations from the extended conformation. The inclusion of DSB constraints improves the overall quality of the final structures significantly. Further increase of the DSB potential bias (e.g.  $E_0 = 10$  kcal/mol and larger, data not shown) additionally improves the final structure because the bias becomes much larger than the BHT move acceptance threshold, i.e. the structure becomes more rigid for moves and stays closer to the initial structure.

### 3.2. $\beta$ -Sheet proteins

#### 3.2.1. 1PG1

We investigated a  $\beta$ -sheet protein, the antimicrobial protegrin-1, PDB code 1PG1 (Fahrner et al., 1996) isolated from porcine leukocytes. It has a native conformation with two  $\beta$ -sheet strands connected by a turn backbone region and two disulfide bridges near the backbone termini and the turn located between CYS6 and CYS15, and CYS8 and CYS13, respectively. In a previous study (Kondov et al., 2009) we have carried out stochastic-search simulations of 1PG1 in presence of Morse potential and without DSB constraints. Therein, we found that 1PG1, in contrast to 1WQE, cannot be correctly described without DSB constraining potential. Particularly, the secondary structure could not be correctly reproduced. In the presence of the disulfide potential the native-like conformation is significantly stabilized. On the basis of a connectivity graph we identified only a small energy gap between the native conformation and competing unstructured and helical non-native conformations.

In the following, for each DSB potential and parameter set we have performed 30 independent BHT simulations with duration of 600 cycles starting from the completely extended conformation, which differed by  $\approx 18$  Å from the native conformation, has no secondary structure and also no disulfide bonds. We have selected the best structures for each DSB constraint type and characterized their quality using the same metrics as for the other proteins and depicted them in Fig. 6. The scaled energies differ insignificantly in all simulations. 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



**Fig. 6.** Performance analysis based on best-energy conformations for the protein 1PG1. The bars on the bottom (left axis) depict the  $E_{sc}$ , SSS, FNC and the Score according to Eq. (2). The bars descending from the top (right axis) denote the RMSD and the DSB distances.

the PDB set), as is illustrated in Fig. 4. The highest total score of 59 was found for the structure with the Morse potential followed by the only slightly different structure with the Sqrt DSB potential. For the whole peptide the large overall backbone RMSD of 3.8 Å arises from the outlying 4-residue fragment ARG-GLY-GLY-ARG at the N-terminus that is apparently fluctuating under physiological conditions as evidenced by the 20 experimental structures in the protein data bank (PDB). Excluding this region the backbone RMSD between the PDB and the simulated structure is just 1.26 Å, indicating that the simulation reproduces the native structure within the experimental resolution.

In absence of DSB constraints structures with helical content are close in energy (about 0.2 kcal/mol) to native-like  $\beta$ -sheet structures. The lowest-energy structure (cf. Table 1) has one helical region and thus SSS of only 21%. In contrast, the SSS in the unconstrained 1WQE simulation is nearly unity (see the previous subsection). As a consequence, the score of the constraint-free calculation for 1PG1 is much smaller than that for 1WQE. In the simulation with the DSB potential, the protein is stable in its native conformation, whereas in the simulation without DSB constraints it diverges by as much as 7.3 Å from the native structure. This shows that the disulfide bridges are essential for stabilizing this protein, which is structurally well reproduced in their presence.

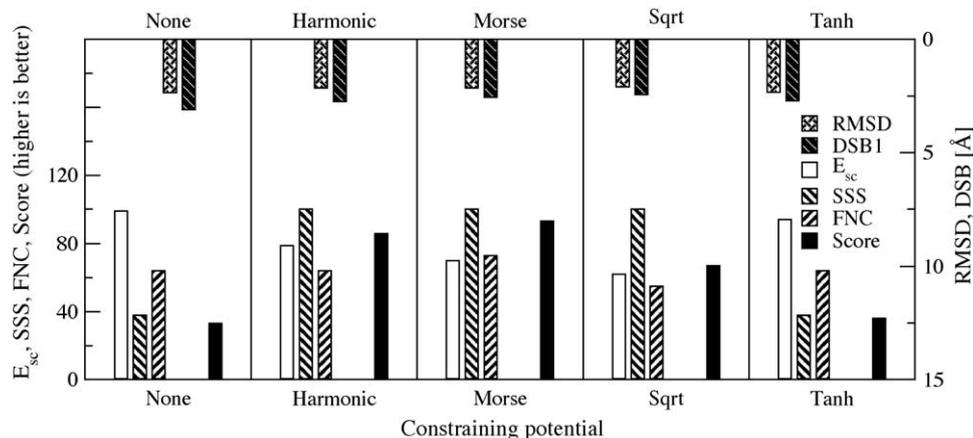
### 3.2.2. 1KVG

The  $\beta$ -haipin 1KVG is a 12-residue artificially designed polypeptide (Skeltan et al., 2002). It has one disulfide bond between CYS2

and CYS11. We performed 30 simulations each taking 300 BHT cycles starting from the extended conformation. Again, for the best-energy conformation from each parameter set we calculated the quantities used as performance metrics and then plotted the best-performing sets for every potential in Fig. 7.

1KVG could not be found in near-native conformations without constraining potential. In particular, it was found in an unstructured coil conformation at the end of the simulation (cf. Table 1). The inclusion of constraining potential of any of the studied types stabilizes the native  $\beta$ -sheet secondary structure and yields near-native conformations. Especially improved are the DSB lengths, the secondary structure and the fraction of native contacts are significantly improved only with the harmonic, Morse and square root potentials. The RMSD is not notably influenced by the presence and the type of constraints that is expected for the small size of the protein for which the RMSD for a random coil is not much larger than the experimental error.

Overall, the simulations with Morse DSB potential result in structures with the highest total score. However, the energy with the Morse potential is approx. 30% (Fig. 7 is lower) that is the largest difference for all studied proteins. One trivial source for this bias is the small size and hence small total energy of the protein compared to the DSB energy  $E_{DSB}$  (cf. Table 1). On the other hand, the formation of highly stable non-native conformations contribute to this behavior. Thus, the stabilization effect of the DSBs is essential for the formation of the correct secondary structure.



**Fig. 7.** Performance analysis based on best-energy conformations for the protein 1KVG. The bars on the bottom (left axis) depict the  $E_{sc}$ , SSS, FNC and the Score according to Eq. (2). The bars descending from the top (right axis) denote the RMSD and the DSB distances.

At the end of this section, let us discuss the performance of the different constraining potentials comparing the simulations of all four proteins. The structure of the two-helical 1WQE is simulated at best quality without constraining potentials (total score of 59) compared to the other proteins within this study (33 for 1KVG, 1 for 1PG1 and 2 for 1HD6). This implies that, for DSB proteins in general, constraints improve the simulation efficiency and quality of the final structures significantly. While the best-performing constraining potential varies for the different proteins the Morse potential is sufficient for simulation of all of studied proteins and is hence most transferable from the set of studied types of constraints. In addition, the Morse potential shows optimal performance for the  $\beta$ -sheet proteins 1KVG and 1PG1.

In all simulations, the DSB lengths for the best final structures are somewhat overestimated (cf. Table 1). The natural bond length in most DSB proteins is between 2.0 and 2.1 Å which implies a systematic error of the order of 20–30%. Because the constraining potentials have minimum value at 2.0 Å and the conformational space is sufficiently sampled in the vicinity of the final structure we assume that this is due to a steric hindrance caused by the force field not allowing the sulfur atoms to come closer to each other. Future work can improve the length of the disulfide bonds in the final structures.

#### 4. Conclusions

In this work, four proteins with varying secondary structure, size and number of disulfide bonds were studied by a stochastic simulation technique employing an all-atom force field in presence and without DSB constraining potentials. Different types of constraining potentials have been tested—harmonic, Morse, square root and hyperbolic tangent. For all proteins studied, inclusion of the constraining potential resulted in structures closer to the native conformation as compared to the best structures obtained in constraint-free simulations. The improvement of the structures has been demonstrated by means of metrics, such as the RMSD, DSSP secondary structure (score), fraction of native contacts and the distance between sulfur atoms of involved cysteine residues. All these metrics were computed using the relevant experimental structures from the PDB as reference.

One of the proteins, the  $\alpha$ -helical 1WQE was simulated starting from the completely extended conformation to its native conformation without inclusion of DSB constraints whereas the  $\beta$ -sheet proteins 1KVG and 1PG1 could be found in low-energy near-native conformations only in presence of DSB constraints.

Moreover, the introduction of a total score allowed for direct unbiased comparison of the efficiency of different types of DSB potentials across all simulations performed for different proteins and with variation of the parameters. It turned out that none of the studied constraining potentials is universally surpassing. The Morse potential is, in average, the best performing for the two  $\beta$ -sheet proteins and the second best for the helical proteins. This is why, we expect that it can be transferable and efficient for a larger sets of DSB proteins.

#### Acknowledgements

Support by program “Supercomputing” of the Helmholtz Association and by the Baden-Württemberg-Stiftung (project HPC-5) is gratefully acknowledged.

#### References

Abagyan, R.A., Totrov, M., 1994. Biased probability Monte Carlo conformation searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* 235, 983–1002.

Abkevich, V.I., Shakhnovich, E.I., 2000. What can disulfide bonds tell us about protein energetics, function and folding: simulations and bioinformatics analysis. *J. Mol. Biol.* 300 (4), 975–985.

Allison, J.R., Moll, G.-P., van Gunsteren, W.F., 2010. Investigation of stability and disulfide bond shuffling of lipid transfer proteins by molecular dynamics simulation. *Biochemistry* 49 (32), 6916–6927.

Anfinsen, C.B., 1973. Principles that govern the folding of protein chains. *Science* 181, 223–230.

Aschi, M., Bozzi, A., Di Bartolomeo, R., Petruzzelli, R., 2010. The role of disulfide bonds and N-terminus in the structural properties of hepcidins insights from molecular dynamics simulations. *Biopolymers* 93 (10), 917–926.

Carr, J.M., Wales, D.J., 2005. Global optimization and folding pathways of selected  $\alpha$ -helical proteins. *J. Chem. Phys.* 123, 234901.

Carr, J.M., Wales, D.J., 2008. Folding pathways and rates for the three-stranded  $\beta$ -sheet peptide beta3s using discrete path sampling. *J. Phys. Chem. B* 112 (29), 8760–8769.

Carr, J.M., Wales, D.J., 2009. Refined kinetic transition networks for the gb1 hairpin peptide. *Phys. Chem. Chem. Phys.* 11, 3341–3354.

Chagot, B., Pimentel, C., Dai, L., Pil, J., Tytgat, J., Nakajima, T., Corzo, G., Darbon, H., 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.

Chen, Y.-C., Hwang, J.-K., 2005. Prediction of disulfide connectivity from protein sequences. *Proteins* 61, 507–512.

Czaplewski, C., Oldziej, S., Liwo, A., 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 (1), 29–36.

Delano, W.L., 2002. The PyMOL Molecular Graphics System.

Dill, K.A., Chan, H.S., 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4, 10–19.

Duan, Y., Kollman, P.A., 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282, 740–744.

Eisenberg, D., McLachlan, A.D., 1986. Solvation energy in protein folding and binding. *Nature* 319, 199–203.

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

Feig, M., Karanicolas, J., Brooks III, C.L., 2004. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* 22 (5), 377–395.

Fiser, A., Cserző, M., Tüdös, E., Simon, I., 1992. Different sequence environments of cysteines and half cystines in proteins application to predict disulfide forming residues. *FEBS Lett.* 302 (2), 117–120.

Fiser, A., Simon, I., 2000. Predicting the oxidation state of cysteines by multiple sequence alignment. *Bioinformatics* 16 (3), 251–256.

Garcia, A.E., 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 (24), 13898–13903.

Herges, T., Wenzel, W., 2004. An all-atom force field for tertiary structure prediction of helical proteins. *Biophys. J.* 87, 3100–3109.

Herges, T., 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.

Huang, E.S., Samudrala, R., Ponder, J.W., 1999. Ab initio fold prediction of small helical proteins using distance geometry and knowledge-based scoring functions. *J. Mol. Biol.* 290 (1–2), 267–281.

Hubner, I.A., Deeds, E.J., Shakhnovich, E.I., 2005. High-resolution protein folding with a transferable potential. *Proc. Natl. Acad. Sci. U.S.A.* 102 (52), 18914–18919.

Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22 (12), 2577–2637.

Kirkpatrick, S., Gelatt Jr., C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. *Science* 220 (4598), 671–680.

Kobayashi, Y., Sasabe, I., Akutsu, T., Saito, N., 1992. Mechanism of protein folding. IV. Forming and breaking of disulfide bonds in bovine pancreatic trypsin inhibitor. *Biophys. Chem.* 44, 113–127.

Kondov, I., Verma, A., Wenzel, W., 2007. Folding and structure prediction of proteins containing disulfide bridges. In: Hansmann, U.H.E., Meinke, J., Mohanty, S., Zimmermann, O. (Eds.), *NIC Series: From Computational Biophysics to Systems Biology (CBSB07)*, vol. 36. John von Neumann Institute for Computing, Jülich, pp. 185–188.

Kondov, I., Verma, A., Wenzel, W., 2009. Folding path and funnel scenarios for two small disulfide-bridged proteins. *Biochemistry* 48 (34), 8195–8205.

Laghaei, R., Mousseau, N., Wei, G., 2010. Effect of the disulfide bond on the monomeric structure of human amylin studied by combined hamiltonian and temperature replica exchange molecular dynamics simulations. *J. Phys. Chem. B* 114 (20), 7071–7077.

Li, Z., Scheraga, H.A., 1987. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 84 (19), 6611–6615.

Liu, A., Luginbühl, P., Zerbe, O., Ortenzi, C., Luporini, P., Wüthrich, K., 2001. Letter to the Editor: NMR structure of the pheromone Er-22 from *Euplotes raikovi*. *J. Biomol. NMR* 19, 75–78.

Mart-Renom, M.A., Stote, R.H., Querol, E., Avilés, F.X., Karplus, M., 1998. Refolding of potato carboxypeptidase inhibitor by molecular dynamics simulations with disulfide bond constraints. *J. Mol. Biol.* 284, 145–172.

- Micheletti, C., de Filippis, V., Maritan, A., Seno, F., 2003. Elucidation of the disulfide-folding pathway of hirudin by a topology-based approach. *Proteins* 53, 720–730.
- Miller, M.A., Doye, J.P.K., Wales, D.J., 1999. Structural relaxation in atomic clusters: master equation dynamics. *Phys. Rev. E* 60 (4), 3701–3718.
- Mortenson, P.N., Evans, D.A., Wales, D.J., 2002. Energy landscapes of model polyanions. *J. Chem. Phys.* 117, 1363–1376.
- Mortenson, P.N., Wales, D.J., 2004. Energy landscapes, global optimization and dynamics of poly-alanine Ac(ala)<sub>8</sub>NHMe. *J. Chem. Phys.* 114, 6443–6454.
- Mucchielli-Giorgi, M.H., Hazout, S., Tuffry, P., 2002. Predicting the disulfide bonding state of cysteines using protein descriptors. *Cover Image Proteins: Struct. Funct. Genet.* 46 (3), 243–249.
- Muskal, S.M., Holbrook, S.R., Kim, S.-H., 1990. Prediction of the disulfide-bonding state of cysteine in proteins. *Protein Eng.* 3 (8), 667–672.
- Nayeem, A., Vila, J., Scheraga, H., 1991. 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.* 12 (5), 594–605.
- O'Connor, B.D., Yeates, T.O., 2004. GDAP: a web tool for genome-wide protein disulfide bond prediction. *Nucleic Acids Res.* 32, 360–364.
- Onuchic, J.N., Luthey-Schulten, Z., Wolynes, P.G., 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48, 545–600.
- Paci, E., Lindorff-Larsen, K., Dobson, C.M., Karplus, M., Vendruscolo, M., 2005. Transition state contact orders correlate with protein folding rates. *J. Mol. Biol.* 352, 495–500.
- Plaxco, K.W., Simons, K.T., Baker, D., 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277, 985–994.
- Prentiss, M.C., Wales, D.J., Wolynes, P.G., 2008. Protein structure prediction using basin-hopping. *J. Chem. Phys.* 128, 225106.
- Qin, M., Zhang, J., Wang, W., 2006. Effects of disulfide bonds on folding behavior and mechanism of the  $\beta$ -sheet protein tendamistat. *Biophys. J.* 90, 272–286.
- Quintilla, A., Starikov, E., Wenzel, W., 2007. De novo folding of two-helix potassium channel blockers with free-energy models and molecular dynamics. *J. Chem. Theory Comput.* 3 (3), 1183–1192.
- Quintilla, A., Wenzel, W., 2007. Folding of two helical peptide with free energy methods and molecular dynamics. In: Hansmann, U.H.E., Meinke, J., Mohanty, S., Zimmermann, O. (Eds.), *NIC Series: From Computational Biophysics to Systems Biology (CBSB07)*, vol. 36. John von Neumann Institute for Computing, Jülich, pp. 267–269.
- Rey, A., Skolnick, J., 1994. Computer simulation of the folding of coiled coils. *J. Chem. Phys.* 100 (3), 2267–2276.
- Sharadadevi, A., Nagaraj, R., 2010. A Molecular dynamics study of human defensins HBD-1 and HNP-3 in water. *J. Biomol. Struct. Dyn.* 27 (4), 541–550.
- Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., Bank, J.A., Jumper, J.M., Salmon, J.K., Shan, Y., Wriggers, W., 2010. Atomic-level characterization of the structural dynamics of proteins. *Science* 330 (6002), 341–346.
- Simmerling, C., Strockbine, B., Roitberg, A.E., 2002. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* 124 (38), 11258–11259.
- Skelton, N., Russell, S., de Sauvage, F., Cochran, A., 2002. Amino acid determinants of beta-hairpin conformation in erythropoietin receptor agonist peptides derived from a phage display library. *J. Mol. Biol.* 316 (5), 1111–1125.
- Snow, C.D., Nguyen, H., Pande, V.S., Gruebele, M., 2002a. Absolute comparison of simulated and experimental protein folding dynamics. *Nature* 420, 102–106.
- Snow, C.D., Zagrovic, B., Pande, V.S., 2002b. The Trp cage: folding kinetics and unfolded state topology via molecular dynamics simulations. *J. Am. Chem. Soc.* 124 (49), 14548–14549.
- Strunk, T., Verma, A., Gopal, S.M., Schug, A., Klenin, K., Wenzel, W., 2009. De novo protein folding with distributed computational resources. In: Grotendorst, J., Attig, N., Blügel, S., Marx, D. (Eds.), *Multiscale Simulation Methods in Molecular Sciences (Lecture Notes)*. NIC Series, vol. 42. Jülich, pp. 397–420.
- Verma, A., Gopal, S.M., Oh, J.S., Lee, K.H., Wenzel, W., 2007. All-atom de novo protein folding with a scalable evolutionary algorithm. *J. Comput. Chem.* 28 (16), 2552–2558.
- Verma, A., Schug, A., Lee, K.H., Wenzel, W., 2006. Basin hopping simulations for all-atom protein folding. *J. Chem. Phys.* 124, 044515.
- Verma, A., Wenzel, W., 2007. Predictive and reproducible de novo all-atom folding of a  $\beta$ -hairpin loop in an improved free-energy forcefield. *J. Phys.: Condens. Matter* 19, 285213.
- Verma, A., Wenzel, W., 2009. A free-energy approach for all-atom protein simulation. *Biophys. J.* 96 (9), 3483–3494.
- Wales, D.J., 2010. Energy landscapes of clusters bound by short-ranged potentials. *ChemPhysChem* 11 (12), 2491–2494.
- Wales, D.J., Dewbury, P.E.J., 2004. Effect of salt bridges on the energy landscape of a model protein. *J. Chem. Phys.* 121, 10284–10290.
- Wales, D.J., Doye, J.P.K., 1997. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *J. Phys. Chem. A* 101 (28), 5111–5116.
- Watanabe, K., Nakamura, A., Fukuda, Y., Saitō, N., 1991. Mechanism of protein folding: III. Disulfide bonding. *Biophys. Chem.* 40 (3), 293–301.
- Wenzel, W., 2006. Predictive folding of a  $\beta$ -hairpin protein in an all-atom free-energy model. *Europhys. Lett.* 76, 156–162.
- Yang, J.S., Chen, W.W., Skolnick, J., Shakhnovich, E.I., 2007. All-atom ab initio folding of a diverse set of proteins. *Structure* 15 (1), 53–63.
- Zhang, J., Li, W., Wang, J., Qin, M., Wu, L., Yan, Z., Xu, W., Zuo, G., Wang, W., 2009. Protein folding simulations: from coarse-grained model to all-atom model. *IUBMB Life* 61, 627–643.