# Predictive folding of a tryptophane zipper hairpin with a free-energy forcefield

Konstantin Klenin and Wolfgang Wenzel

***Abstract:***—De-novo folding of even small proteins from completely extended conformations remains a grand-computational challenge for all-atom simulation methods. We have recently developed a novel, free-energy based simulation approach, that permits folding of several helical proteins with an all-atom forcefield. He we demonstrate that it is also possible to fold that beta-hairpins. We reproducibly and predictively fold a monomeric stable tryptophane-zipper (pdb-code 1LE0) from unfolded starting conformations. We find that the entire population of simulated structures converges to a near-native ensemble, which differs by just 1-2 Å from the native conformation.

***Key-Words:***—Protein folding, Monte Carlo, numerical optimization

## I. INTRODUCTION

Ab-initio protein tertiary structure prediction and the elucidation of the mechanism of the folding process are among the important outstanding problems of biophysical chemistry.

All-atom protein structure prediction with free-energy models is based on the thermodynamic hypothesis [1], which stipulates that proteins in their native conformation are in thermodynamic equilibrium with their environment. Following this hypothesis, the native conformation of a protein can be predicted as the global optimum of its free-energy surface.

According to the funnel paradigm for protein folding [2,3] the free-energy landscape guides the protein on average towards the native conformation. Stochastic optimization methods speed up the search process [4] on this landscape by sacrificing the kinetics of the folding process, but still permit a characterization of the folding landscape.

We developed an all-atom free-energy forcefield for helical proteins (PFF01) [5] and demonstrated the reproducible and predictive folding of several helical proteins of 20-60 amino-acids in length [6-11]. A recent modification of our forcefield stabilizes several β-hairpins [12, 13] and also the helical proteins investigated previously.

Predictive hairpin folding remains a computational challenge [14-18], because the phase space of possible β-backbone hydrogen-bond topologies is much larger than that of α-helices. Here we investigate a stable monomeric 12 amino-acid trp-zipper protein (pdb code 1LE0) [19].

A recent landmark molecular dynamics simulation comprising more than 22 ms total simulation time ($O(10^{12})$ energy evaluations) succeeded to fold three trp-zippers [17], including 1LE0, and to characterize their free-energy landscape under physiological conditions.

Manuscript received June 9, 2007: Revised received December 1, 2008.
Konstantin Klenin and Wolfgang Wenzel are with the Institut für Nanotechnologie, Forschungszentrum Karlsruhe PO Box 2640, 76021 Karlsruhe, GERMANY, wenzel@int.fzk.de http://www.fzk.de/biostruct

Here we use a free-energy model to demonstrate the predictive reproducible folding of the trp-zipper to within experimental resolution using a simulation method which requires only $O(10^8)$ energy evaluations.

## II. METHODS

### A. Forcefield

Forcefield PFF02 [5, 13] parameterizes the internal free-energy of proteins, excluding backbone entropy, in an all-atom model (except apolar $CH_n$ groups) and contains the following non-bonded interactions:

$$F_{\text{int}} = \sum_{i,j} \frac{q_i\, q_j}{\varepsilon_{g(i)g(j)}\, r_{ij}} + \sum_i \sigma_i A_i + V_{LJ} + V_{\text{hb}} + V_{\text{tor}}\,.$$

The non-trivial electrostatic interactions in proteins are represented via group-specific dielectric constants [5, 20]. Interactions with the solvent were first fit and adapted in a minimal solvent accessible surface model [21] ($\sigma_i$ is the free-energy of solvation per unit area and $A_i$ the area of atom $i$ that is in contact with a fictitious solvent). $V_{\text{LJ}}$ is the Lennard-Jones potential. Hydrogen bonds are described via dipole-dipole interactions included in the electrostatic terms and an additional short range correction ($V_{\text{hb}}$) for backbone-backbone hydrogen bonds [5].

The forcefield PFF02 contains an additional term that differentiates between the backbone dipole alignments found in different secondary structure elements (included in the electrostatic potential between atoms $i$ and $j$ via the dielectric constants $\varepsilon_{g(i)g(j)}$) [22] and a torsional potential $V_{\text{tor}}$ for backbone dihedral angles [13]. In our simulation we consider only moves around the sidechain and backbone dihedral angles as detailed in Ref. 11.

### B. Folding Simulations

The low-energy part of the free-energy landscape of proteins is quite rugged due to the comparatively close packing of the atoms in the native structure. The basin hopping technique [23,24] (BHT) eliminates high-energy barriers in the stochastic search by replacing the energy of each conformation with the energy of a nearby local minimum.

The basin hopping technique and derivatives [25] have been used previously to study the potential energy surface of model proteins [26] and polyalanines using all-atom models [27, 28].

Recently, we have generalized this method to a population of size $N$, which is iteratively improved by $P$ concurrent dynamical processes [29, 30]. The whole population is guided

towards the optimum of the free energy surface with a simple evolutionary strategy in which members of the population are drawn and then subjected to a basin hopping cycle. At the end of each cycle, the resulting conformation either replaces a member of the active population or is discarded.

This algorithm was implemented on a distributed master-client model in which idle clients request a task from the master. The master maintains a list of open tasks comprising the active conformations of the population. The client then performs a Monte Carlo run of 50000 steps at the fixed temperature of 400 K followed by a simulated annealing of 50000 steps.

| Number | Energy | bRMSD | CEEECSSSEEEC |
|---|---|---|---|
| 1 | –30.05 | 1.50 | CEEEETTEEEEC |
| 2 | –29.98 | 1.41 | CEEEETTEEEEC |
| 3 | –29.34 | 1.56 | CEEEETTEEEEC |
| 4 | –28.65 | 1.48 | CEEEETTEEEEC |
| 5 | –28.07 | 1.17 | CEEEETTEEEEC |
| 6 | –27.42 | 2.23 | CEEETTTEEECC |
| 7 | –27.40 | 2.10 | CEEEETTEEEEC |
| 8 | –26.09 | 2.12 | CEEETTTEEECC |
| 9 | –26.02 | 2.92 | CEEETTTEEECC |
| 10 | –25.99 | 2.92 | CEEETTTEEECC |

Table 1. Optimal internal free-energy (in kcal/mol), bRMSD (in Å) and secondary structure, as computed by DSSP [31] for the best ten final conformations. The top row shows the secondary structure of the native conformation. The five energetically best simulations predict iso-energetic conformations that agree with the native conformation (first model) to experimental resolution.

The simulated annealing runs used a geometric cooling schedule reducing the temperature from 400 to 50 K. The number of clients was $P = 99$, the number of conformation in the population was $N = 50$, and the total number of cycles was 15. The individual conformations from the final population were further optimized by BHT of 5 cycles. Each cycle consisted of the annealing run of 400000 steps with the temperature decreasing from 200 to 5 K.

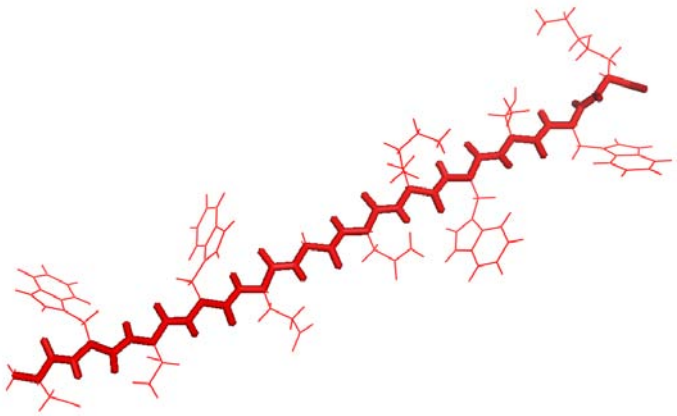

Figure 1. Totally extended structure used as a starting conformation.

## III. RESULTS

We have performed evolutionary algorithm optimization starting from a totally extended structure (over 12 Å backbone root mean square deviation (bRMSD) to the native conformation) as shown in Fig. 1. After the free energy optimization the five energetically lowest simulations (see Table 1) had converged to nearly iso-energetic conformations, each agreed with the native conformation within the resolution of the experiment (1.2-1.6 Å bRMSD). The agreement between the experimental and the model structure is illustrated in Fig. 2. As can be seen from the secondary structure analysis ($4^{th}$ column of the table) all conformations correctly predicted the location of the turn in the sequence. Also the assignment of the beta-sheet regions agrees very well with that found in experiment. We note, however, variations in the stacking of the trp side-chains, which differ among the members of the population.

Considering the energy differences between the various members of the population, many of these conformations can be expected to be populated under physiological conditions in the experiment. Indeed the designed tryptophan zipper investigated here is one of the few beta-hairpin structures, which are stable in isolation under physiological conditions. The terminal fragment of protein G, for instance, which also forms a beta-hairpin structure, populates and the folded confirmation only about 40% of the time. Therefore the designed trp-zipper mini-proteins are well suited to study the basic ingredients of the protein folding mechanism.

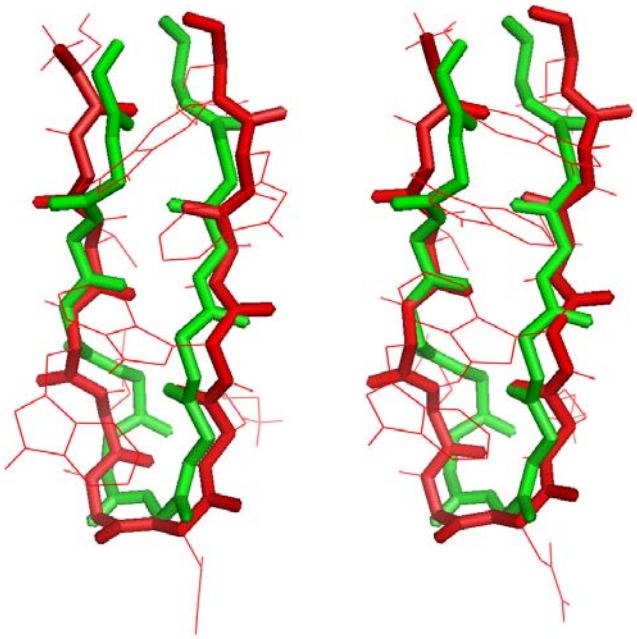

Figure 2. Two typical representatives of the five energetically best conformations obtained in the simulations (red) overlaid with the experimental structure (first NMR model) (green).

## IV. CONCLUSIONS

Using an all-atom free-energy model we have predictively folded the trp-zipper to within experimental resolution. Our results demonstrate that free-energy based methods can

optimize the internal free-energy of proteins orders of magnitude faster than methods that construct folding trajectories [17], even for proteins with glassy free-energy surfaces such as the trp-zipper. Free-energy methods complement molecular dynamics based studies into the folding process, but presently lack the ability to incorporate backbone conformational entropy. Such data, as well as the temperature dependence of the model parameters, would be required to fully characterize the folding process as a function of temperature. Ultimately, the free-energy approach offers prospects for *de-novo* protein design, because it is possible to investigate the relative stability of competing structural ensembles as a function of the amino-acid sequence.

## ACKNOWLEDGMENT

We thank the volunteers of POEM@HOME for the support of this project. This work was funded by the German National Science Foundation (DFG WE 1863/10-2). We also acknowledge the use of computational facilities at the Computational Science Center of the Korean Institute of Science.

## REFERENCES

[1] C. B. Anfinsen, Science, Vol.181, 1973, pp.223-230.
[2] K. A. Dill and H. S. Chan. *Nature Structural Biology* Vol. 4, 1997, pp. 10–19.
[3] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, *Annu. Rev. Phys. Chem.*, Vol. 48, 1997, pp. 545–600.
[4] U. H. E. Hansmann and Y. Okamoto, *Curr. Opin. Struc. Biol.*, Vol. 9, 1999, p. 177.
[5] T. Herges and W. Wenzel, *Biophys. J.*, Vol. 87, 2004, pp. 3100–3109.
[6] A. Schug, T. Herges, and W. Wenzel, *Phys. Rev. Letters*, Vol. 91, 2003, 158102.
[7] A. Schug and W. Wenzel, *Europhys. Lett.*, Vol. 67, 2004, pp. 307–313.
[8] T. Herges and W. Wenzel, *Phys. Rev. Lett.*, Vol. 94, 2005, 018101.
[9] A. Schug, T. Herges, and W. Wenzel, *Proteins*, Vol. 57, 2004, pp. 792–798.
[10] A. Schug and W. Wenzel, J. *Am. Chem. Soc.*, Vol. 126, 2004, pp. 16736–16737.
[11] A. Schug and W. Wenzel, *Biophysical J.*, Vol. 90, 2006, pp. 4273–4280.
[12] W. Wenzel, *Europhys. Letters*, Vol. 76, 2006, pp.156–162.
[13] A. Verma and W. Wenzel, *J. Phys: Cond. Matter*, Vol. 19, 2007, 285213.
[14] B. D. Bursulaya and C. L. Brooks, *J. Am. Chem. Soc.*, Vol. 121, 1999, pp. 9947–9951.
[15] P. Ferrara and A. Caflisch, *Proc. Nat. Acad. (USA)*, Vol. 97, 2000, pp. 10780–10785.
[16] E. Paci, A. Cavalli, M. Vendruscolo and A. Caflisch, *Proc. Nat. Acad. (USA)*, Vol. 100, 2003, pp. 8271–8222.
[17] C. D. Snow, L. Qiu, D. Du, F. Gai, S. J. Hagen, and V. S. Pande, *Proc. Nat. Acad. (USA)*, Vol. 101, 2004, pp. 4077–4082.
[18] W. Y. Yang, J. W. Pitera, W. C. Swope, and M. Gruebele, *J. Mol. Biol.*, Vol. 336, 2004, pp. 241–251.
[19] A. G. Cochran, N. J. Skelton, and M. A. Starovasnik, *Proc. Natl. Acad. Sci. (USA)*, Vol. 98, 2001, pp. 5578–5583.
[20] F. Avbelj, *Biochemistry*, Vol. 31, 1992, pp. 6290–6297.
[21] D. Eisenberg and A. D. McLachlan, *Nature*, Vol. 319, 1986, pp. 199–203.
[22] F. Avbelj and J. Moult, *Biochemistry*, Vol. 34, 1995, pp. 755–764.
[23] A. Nayeem, J. Vila and H. A. Scheraga, *J. Comp. Chem.*, Vol. 12, 1991, pp. 594–605.
[24] D. J. Wales and J. P. K. Doye, *J. Phys. Chem.*, Vol. 101, 1997, p. 5111.
[25] R. A. Abagyan and M. Totrov, *J. Mol. Biol.*, Vol. 235, 1994, pp. 983–1002.
[26] J. M. Carr and D. J. Wales, *J. Chem. Phys.*, Vol. 123, 2005, 234901.
[27] P. N. Mortenson and D. J. Wales, *J. Chem. Phys.*, Vol. 114, 2004, pp. 6443–6454.
[28] P. N. Mortenson, D. A. Evans, and D. J. Wales, *J. Chem. Phys.*, Vol. 117, 2002, pp. 1363–1376.
[29] A. Verma, A. Schug, K. H. Lee, and W. Wenzel, *J. Chem. Phys.*, Vol. 124, 2006, 044515.
[30] A. Verma, S. M. Gopal, J. S. Ooh, K. H. Lee, and W. Wenzel, *J. Comp. Chem.*, Vol. 28, 2007, pp. 2552–2558.
[31] W. Kabsch and C. Sander, *Biopolymers*, Vol. 22, 1983, pp. 2577–2637.