



## Reduction to the Simplest – The Complexity of Minimalistic Heteropolymer Models

Michael Bachmann, Wolfhard Janke

published in

*NIC Workshop 2006,*  
*From Computational Biophysics to Systems Biology,*  
Jan Meinke, Olav Zimmermann,  
Sandipan Mohanty, Ulrich H.E. Hansmann (Editors)  
John von Neumann Institute for Computing, Jülich,  
NIC Series, Vol. 34, ISBN-10: 3-9810843-0-6,  
ISBN-13: 978-3-9810843-0-6, pp. 59-62, 2006.

© 2006 by John von Neumann Institute for Computing  
Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume34>

# Reduction to the Simplest – The Complexity of Minimalistic Heteropolymer Models

Michael Bachmann and Wolfhard Janke

Institut für Theoretische Physik and Centre for Theoretical Sciences (NTZ),  
Universität Leipzig, 04109 Leipzig, Germany  
*E-mail:* {bachmann, janke}@itp.uni-leipzig.de

Simple coarse-grained hydrophobic-polar models for heteropolymers as the lattice HP and the off-lattice AB model allow a general classification of characteristic behaviors for hydrophobic-core based tertiary folding. The strongly reduced computational efforts enable one to reveal systematically the thermodynamic properties of comparatively long sequences in a wide temperature range of conformational activity. Based on a suitable cooperativity parameter, characteristic folding channels and free-energy landscapes, which have strong similarities with realistic folding paths, can be analysed.

## 1 Introduction

The understanding of the sequence-dependent stability of native protein structures, the interplay of sequence and conformation, and the classification of conformational phases are essential aspects in protein folding studies. By means of computer simulations of detailed microscopic models with modern algorithms, great progress has been achieved in the past few years in predicting native structures for exemplified amino acid sequences. Also, the relation between conformational cooperativity and thermodynamic activity in the folding process has been subject of numerous selective studies. Nonetheless, a general framework, which would allow the classification of characteristic global folding events on the basis of an appropriate “order” or cooperativity parameter, is still lacking. Another difficult problem is the systematic analysis of the influence of the amino acid order in the sequence on folding kinetics and native fold. Since these questions cannot or only partially be answered with all-atom protein models, it is necessary to increase the level of abstraction in appropriate modelling. It is known that most of these problems are related to the so-called tertiary folding process which concerns almost all amino acid residues in a cooperative manner, in contrast to the much more localised secondary structure segments. The main idea is that tertiary folding is strongly related to the hydrophobic effect. Many proteins fold spontaneously in an aqueous solution and the hydrophobicity of the distinguished side chains is mainly responsible for the position of the side chains in the native fold. Polar residues favor contact with the polar environment and screen more hydrophobic monomers, which typically form a compact core in the interior of the protein structure. In connection with volume exclusion, it is the hydrophobic force that enters into minimalistic hydrophobic-polar models.

## 2 The Hydrophobic-Polar (HP) Lattice Model

The simplest hydrophobic-core based approach is the HP lattice model<sup>1</sup>, where in the original form only nearest-neighbor hydrophobic monomers, being nonadjacent along the self-

avoiding chain, attract each other. A remarkable result obtained from exactly enumerating the entire sequence and conformation space of HP lattice proteins in this parameterization<sup>2-4</sup> is that only a small fraction of possible sequences is nondegenerate, i.e., for these only one conformation representing the “native” state exists. This is interpreted as indication of stability in the funnel-like free-energy landscape: There is an energy gap towards higher-energetic states. Such sequences are called *designing*. On the other hand, it is interesting that the number of designable conformations, i.e., structures being native folds of designing sequences, is also very small, typically even smaller than the number of designing sequences.<sup>4</sup> This means that several designing sequences fold into the same native conformation. In studies of thermodynamic activity accompanying the folding process, it turned out that, depending on the sequence, the folding trajectory can pass highly compact intermediate globular states. Favoring hydrophobic contacts energetically, however, hydrophobic-core folds are not necessarily maximally compact.<sup>4,5</sup> Identifying low-lying energetic states and analyzing thermodynamic properties of HP lattice proteins is not straightforward and requires sophisticated computational methods.<sup>5,6</sup> Although the tendency of the HP model to favor folding paths with intermediary conformational phases is seen as sign of too little cooperativity, its power lies in the fact that sequences with far more than 100 monomers can be analyzed. These sequences are much longer than what can be studied by means of other models which are not knowledge-based, i.e., not biased towards a certain target structure.

### 3 Getting Rid of the Lattice: A Minimal Continuum Model

A straightforward off-lattice generalization of the HP model is the AB model (*A*: hydrophobic, *B*: polar), which was originally proposed for heteropolymers in two dimensions.<sup>7</sup> Volume exclusion and effective inter-residue interactions are modeled by Lennard-Jones potentials, and a “ferromagnetic” bending energy models effectively stiffness and torsional barriers in the three-dimensional representation.<sup>8,9</sup> For studies of folding channels, we define an angular overlap parameter of torsional and bond angles. Comparing a conformation  $\mathbf{X}$  with the native conformation  $\mathbf{X}_0$ , this similarity parameter is defined by  $Q(\mathbf{X}) = 1 - d(\mathbf{X})$ , where  $d(\mathbf{X})$  is a measure of the angular deviation of  $\mathbf{X}$  from  $\mathbf{X}_0$ .<sup>8,9</sup> Restricting the canonical partition function at temperature  $T$  to the ensemble of conformations with overlap  $Q$ ,  $Z(Q) = \int \mathcal{D}\mathbf{X} \delta(Q - Q(\mathbf{X})) \exp\{-E(\mathbf{X})/k_B T\}$ , the overlap free energy is given by  $F(Q) = -k_B T \ln Z(Q)$ . Figure 1 shows the overlap free-energy landscapes for the 20-mer  $BA_6BA_4BA_2BA_2B_2$ , which behaves like a typical two-state folder. For permuted sequences, we also find characteristic folding through intermediate conformations and metastability.<sup>9</sup>

### 4 Concluding Remarks

Protein folding is a highly complex process where microscopic details are of essential importance. Nonetheless, the effective hydrophobic force which is responsible for the tertiary fold is of rather mesoscopic nature, since it is due to the interaction with the bath of surrounding solvent molecules. For this reason, appropriate effective coarse-grained models should allow a qualitative description of characteristic features accompanying conformational folding transitions. By means of minimalistic standard hydrophobic-polar models

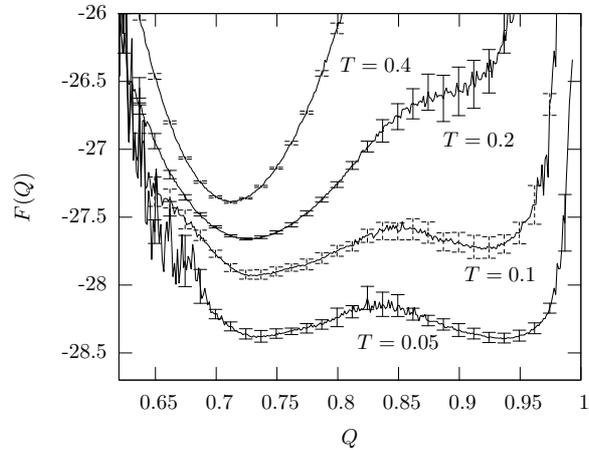


Figure 1. Free-energy landscapes  $F(Q)$  at different temperatures for the sequence  $BA_6BA_4BA_2BA_2B_2$ , exhibiting typical two-state folding behavior.<sup>9</sup>

we have shown that even at this level of abstraction typical properties of native folds and cooperative effects in the folding process can be described qualitatively.

## Acknowledgments

We thank H. Arkin, R. Schiemann, S. Schnabel, T. Vogel, A. Kallias, J. Schluttig, and C. Junghans for cooperation. This work is partially supported by the DFG (German Science Foundation) under Grant No. JA 483/24-1 and the computer time Grant No. h1z11 of the John von Neumann Institute for Computing (NIC), Forschungszentrum Jülich.

## References

1. K. A. Dill, *Biochemistry* **24**, 1501 (1985).
2. C. Tang, *Physica A* **288**, 31 (2000).
3. H. Li, R. Helling, C. Tang, and N. Wingreen, *Science* **273**, 666 (1996); H. Cejtin, J. Edler, A. Gottlieb, R. Helling, H. Li, J. Philbin, C. Tang, and N. Wingreen, *J. Chem. Phys.* **116**, 352 (2002).
4. R. Schiemann, M. Bachmann, and W. Janke, *J. Chem. Phys.* **122**, 114705 (2005); *Comp. Phys. Commun.* **166**, 8 (2005).
5. M. Bachmann and W. Janke, *Phys. Rev. Lett.* **91**, 208105 (2003); *J. Chem. Phys.* **120**, 6779 (2004), and references therein.
6. H.-P. Hsu, V. Mehra, W. Nadler, and P. Grassberger, *J. Chem. Phys.* **118**, 444 (2003).
7. F. H. Stillinger, T. Head-Gordon, and C. L. Hirshfeld, *Phys. Rev. E* **48**, 1469 (1993).
8. M. Bachmann, H. Arkin, and W. Janke, *Phys. Rev. E* **71**, 031906 (2005).
9. S. Schnabel, M. Bachmann, and W. Janke, to be published (2006).