

Study of a model for the folding of a small protein

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2006 J. Phys.: Condens. Matter 18 5687

(<http://iopscience.iop.org/0953-8984/18/24/009>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 129.252.86.83

The article was downloaded on 28/05/2010 at 11:50

Please note that [terms and conditions apply](#).

Study of a model for the folding of a small protein

Andrea Nobile and Federico Rapuano

Dipartimento di Fisica, Università di Milano-Bicocca and INFN, Sezione di Milano, Italy

Received 13 December 2005, in final form 11 May 2006

Published 2 June 2006

Online at stacks.iop.org/JPhysCM/18/5687

Abstract

We describe the results obtained from an improved model for protein folding. We find that a good agreement with the native structure of a 46-residue-long, five-letter protein segment is obtained by carefully tuning the parameters of the self-avoiding energy. In particular we find an improved free-energy profile. We also compare the efficiency of the multidimensional replica exchange method with the widely used parallel tempering.

1. Introduction

In this paper we report on the results obtained by modifying the model proposed in [1]. We show that the free-energy profile strongly depends on small changes in the self-avoid interaction. Despite the success of the original model in reproducing the native structure of a small three-helix protein (the 10–55 fragment of the B domain of staphylococcal Protein A), a difficulty arises in distinguishing between two quasi-specular topologies of the native structure. The third helix can be either in front of or behind the structure formed by the first and second helices. The tuning of the self-avoid interaction solves this situation. To simulate the thermodynamic properties of the system, we use a multidimensional version of parallel tempering in which both the temperature and other parameters of the model become dynamical variables. This method allows for a much deeper search in the configurational space with respect to standard tempering algorithms.

2. The model

In this section we briefly describe the model originally proposed in [1] which is the starting point of our study.

Geometrical structures

There are three simplified geometrical representations for the amino acids: one for proline, one for glycine and one for all other amino acids. The configuration of the protein is determined for each residue by the two Ramachandran angles ϕ , ψ [2]. Bond lengths and other bond angles are fixed. The two Ramachandran angles ϕ , ψ are the only configurational degrees of freedom

of the model. By keeping the model simple and considering only two degrees of freedom per amino acid we are deliberately reducing the configurational space. It is important to notice that bond lengths and bond angles that we keep fixed in the model are typically subject to very small changes. The three representations of the amino acids have the following characteristics:

- for all amino acids the backbone is represented by an N, C_α and C' chain. The O and H atoms form the hydrogen bonds and are attached to the C' and N atoms of the backbone. The side chains are represented by a big C_β atom bonded to the C_α atom;
- glycine has the same representation, except that the C_β atom is missing;
- proline also has the same representation, except that the H atom is replaced by C_δ and the Ramachandran angle ϕ is fixed. Thus the position of the C_δ atom is fixed with respect to N, C_α and C_β. This constraint also implies that the position of the whole proline except the O atom is fixed with respect to the C' of the previous residue.

The Hamiltonian

The different types of amino acids present in the protein fragment are subdivided into three hydrophobicity classes: hydrophobic (H), polar (P) and intermediate (A). Amino acids Leu, Ile and Phe belong to class H, Ala belongs to class A, while Arg, Asn, Asp, Gln, Glu, His, Lys, Pro, Ser and Tyr belong to class P.

The Hamiltonian \mathcal{H} of the model is written as a sum of four terms:

$$\mathcal{H} = E_{\phi\psi} + E_{sa} + E_{hb} + E_{hf}. \quad (1)$$

Here $E_{\phi\psi}$ depends only on the Ramachandran angles ϕ and ψ . All other terms account for interactions between pairs of atoms. In particular E_{sa} , E_{hb} , E_{hf} correspond to the self-avoiding, hydrogen bond and hydrophobic interaction, respectively.

The term depending on the Ramachandran angles is given by

$$E_{\phi\psi} = \frac{\epsilon_{\phi}}{2} \sum_i (1 + \cos 3\phi_i) + \frac{\epsilon_{\psi}}{2} \sum_i (1 + \cos 3\psi_i), \quad (2)$$

where the sum runs over all amino acids. All the parameters of the model can be found in [1]

All other terms have the general form

$$E_A = \frac{\epsilon_A}{2} \sum_{i \neq j} F_A^{ij}(r_{ij}) \theta(r_A^c - r_{ij}), \quad A = \{sa, hb, hf\}, \quad (3)$$

$$F_A(r) = f_A(r) - f_A(r_A^c) - (r - r_A^c) f'_A(r_A^c),$$

where the sum runs over all the atoms, r_{ij} is the distance between atom i and j , r_A^c is the cutoff radius, the function $F_A(r)$ is constructed in such a way that the interaction vanishes with its derivative at the cutoff radius.

We now recall the form of the function $f_A(r)$ for the various interactions:

- The self avoiding case ($A = sa$). The sum in (3) in this case involves all atoms except the C_βC_β pair that interact through the term E_{hf}

$$f_{sa} = \left(\frac{\sigma_i + \sigma_j + \Delta\sigma_{ij}}{r_{ij}} \right)^{12}. \quad (4)$$

σ_i is the effective radius of the atom i . $\Delta\sigma_{ij}$ vanishes for all pairs except for C_βC', C_βN and C_βO. For contiguous residues in the chain this term produces a local interaction that modifies the angular dependent part of the potential restricting the space of allowed configurations.

- The hydrogen bond term ($A = \text{hb}$). The sum in (3) runs over ij where i and j label, respectively, H and O atoms. The function $f_{\text{hb}}(r)$ is given by

$$f_{\text{hb}}(r_{ij}) = 2u(r_{ij})v(\alpha_{ij}, \beta_{ij}), \quad (5)$$

where α_{ij} and β_{ij} are the NHO angles and HOC', respectively, and

$$u(r) = 5 \left(\frac{\sigma_{\text{hb}}}{r} \right)^{12} - 6 \left(\frac{\sigma_{\text{hb}}}{r} \right)^{10},$$

$$v(\alpha_{ij}, \beta_{ij}) = \begin{cases} \cos^2 \alpha_{ij} \cos^2 \beta_{ij} & \text{if } \alpha_{ij}, \beta_{ij} > \pi/2 \\ 0 & \text{elsewhere} \end{cases} \quad (6)$$

- The hydrophobicity energy term ($A = \text{hf}$). The sum in (3) is over all C_β atoms belonging to the classes HH, HA, AH. The function $f_{\text{hf}}(r)$ is

$$f_{\text{hf}}(r) = \left(\frac{\sigma_{\text{hf}}}{r} \right)^{12} - 2 \left(\frac{\sigma_{\text{hf}}}{r} \right)^6. \quad (7)$$

3. Improvement of the model

After having reproduced all results obtained in [1], we improved the model in two respects. First, by using a more elaborate algorithm we have explored a larger region of the parameter space in order to test the stability of the results. Second, by modifying the self-avoid interaction we solve the problem of the quasi-specular degeneracy. We discuss both items in the following.

Computational methods

To simulate the model we use a multidimensional extension of parallel tempering (multidimensional replica exchange method). In standard parallel tempering [3, 4] various copies of the system are simulated simultaneously with different temperatures β for a fixed number of steps before an exchange of the temperatures between systems is proposed. In the multidimensional version [5], the copies of the system are evolved with different temperatures and with different values of the parameters. In our study we have used as dynamical variables the temperature and the parameter ϵ_{hf} .

We denote by β_1, \dots, β_m and $\epsilon_1, \dots, \epsilon_m$ the set of temperatures and of ϵ_{hf} considered and C_{ij} the corresponding configurations weighted with the Boltzmann–Gibbs factor $e^{-\beta_i \mathcal{H}_j(C_{ij})}$ (here \mathcal{H}_j is the Hamiltonian with parameter $\epsilon_{\text{hf}} = \epsilon_j$).

After a certain number of Monte Carlo steps to be determined by the dynamics, one proposes a sequence of exchanges between two pairs of parameters $\{\beta_i, \epsilon_j\}$ and $\{\beta_{i'}, \epsilon_{j'}\}$ and the two corresponding systems C_{ij} and $C_{i'j'}$

$$\begin{aligned} \{\beta_i, \epsilon_j, C_{ij}\} &\rightarrow \{\beta_i, \epsilon_j, C_{i'j'}\}, \\ \{\beta_{i'}, \epsilon_{j'}, C_{i'j'}\} &\rightarrow \{\beta_{i'}, \epsilon_{j'}, C_{ij}\}, \end{aligned} \quad (8)$$

with probability

$$P = \min(1, e^{-R}),$$

$$R = \beta_i H_j(C_{i'j'}) + \beta_{i'} H_{j'}(C_{ij}) - \beta_i H_j(C_{ij}) - \beta_{i'} H_{j'}(C_{i'j'}). \quad (9)$$

Each new system, which is already thermalized, runs again for the same number of Monte Carlo steps before undergoing a new exchange.

We have chosen seven values for the parameter β and six values for the parameter ϵ_{hf} and thus have 42 systems running simultaneously.

The temperatures are assigned by the rule

$$T_i = T_{\min} \left(\frac{T_{\max}}{T_{\min}} \right)^{\frac{i}{N_T-1}} \quad (10)$$

where N_T is the number of values. The values of the parameter ϵ_{hf} are chosen using the rule

$$\epsilon_i = \epsilon_{\min} + \frac{(\epsilon_{\max} - \epsilon_{\min})}{N_\epsilon - 1} i. \quad (11)$$

We have taken the following values:

$$\begin{aligned} \beta^{-1} &= 0.44, 0.486, 0.0537, 0.593, 0.655, 0.724, 0.8, \\ \epsilon_{\text{hf}} &= 1.8, 2.02, 2.24, 2.46, 2.68, 2.9. \end{aligned} \quad (12)$$

We have implemented this algorithm on a cluster of 42 processing nodes each one simulating the entire run with fixed parameters $\{\beta_i, \epsilon_j\}$ while the configurations $C_{i'j'}$ are possibly exchanged. The exchanges are proposed approximately every 7000 standard Monte Carlo updates; this value allows a high number of exchanges.

The simulations consist of 250×10^6 Monte Carlo steps for each processor and take about 8 hours with the processor used (Athlon 2200+ at 1800 MHz).

A great advantage of the multidimensional replica exchange method is that we can study the system for various values of the promoted constants in a single run while improving the sampling for each value at the same time. This is particularly useful for frustrated models where the system easily gets trapped in local minima. Allowing the variation of ϵ_{hf} we can lower specific energy barriers and improve the exploration of the configurational space.

The self-avoid term

The main idea behind this simplified model is that the geometry of the amino acids plays a fundamental role in determining the geometry of the allowed conformations. In this case a simple hydrophobic term added to a detailed representation of the backbone and hydrogen bonds is sufficient to reproduce correctly the fold of a small protein. In order to explore the sensitivity of the results to variations of the geometry we have modified the self-avoid term and the cutoff function.

The smooth cutoff is changed to a simple discontinuous cutoff for the term E_{sa}

$$f(r) \rightarrow \tilde{f}(r) = \begin{cases} 0 & \text{if } r > r_c \\ f(r) & \text{elsewhere.} \end{cases} \quad (13)$$

The parameter $\Delta\sigma_{ij}$ takes the value 0.425 Å instead of 0.625 Å. We use this parameter for all $C_\beta C'$, $C_\beta N$ and $C_\beta O$ couples, while in [1] it is used only for $C_\beta C'$, $C_\beta N$ and $C_\beta O$ separated by three covalent bonds.

4. Results

As stated in [1] we find that the free-energy of the model is characterized by two minima corresponding to the two topologies that a three-helix bundle can assume. The correct topology is the favoured one but the wrong one is present with a non-negligible probability. The difficulty in distinguishing between the two topologies arises from the fact that the contact patterns between helices are very similar in the two states.

The behaviour of the model with respect to the variation of the coefficient ϵ_{hf} that controls the strength of the hydrophobicity forces is very simple: weak interaction constants correspond to free-energy profiles dominated by a totally extended helix while high values of the constant

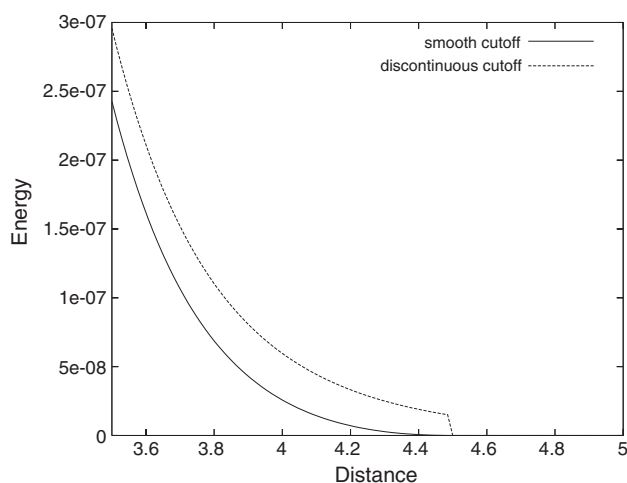


Figure 1. Comparison between the smooth cutoff and the discontinuous cutoff used in our simulations for the case of the self-avoid term. The cutoff radius r_c is 4.5 Å.

correspond to disordered collapsed states. In the collapsed phase, changes in the configuration of the protein occur with great difficulty because of the enormous number of rejected moves due to steric collisions. This effectively reduces the sampling efficiency. The standard parallel tempering or simulated tempering algorithms try to solve the problem by raising and lowering the temperature, thus effectively increasing the volume of the sampled conformational space. The main drawback of these methods is that the increase in temperature easily destroys segments of secondary structure that are difficult to recreate either in an uncollapsed high temperature phase or in a collapsed low temperature phase. These are the main reasons that led us to choose the parameter ϵ_{hf} as a dynamic variable.

The indicator Q of similarity with the native structure is defined by

$$Q = \exp(-\delta^2/100 \text{ \AA}^2) \quad (14)$$

where δ is the root mean square deviation (rmsd), and the free-energy

$$F(Q) = -kT \ln P(Q) \quad (15)$$

where $P(Q)$ is the probability distribution of Q .

In figure 2 we show the free-energy profile $F(Q)$ and the distribution $P(Q)$ of the system at the lowest simulated temperature $kT = 0.44$ and $\epsilon_{\text{hf}} = 2.9$. The two most important minima at $Q \approx 0.81$ and $Q \approx 0.91$ correspond to the native topology. The structure at $Q \approx 0.81$ differs from the native configuration in having the loop region between the second and the third helix partly helical and in the relative positions of the three helices. The first helix tends to stay more aligned with the others than in the native configuration. The minimum corresponding to the wrong three-helix bundle topology that was present in the original model at $Q \approx 0.5$ is not present or is negligible in the simulations with the modified model. The minima at $Q \approx 0.23$ and $Q \approx 0.47$ correspond to disordered structures of various shapes. The free-energy profiles taken at lower ϵ_{hf} are characterized by a deeper minimum in correspondence with the configurations folded in a unique long helix and by a homogeneous reduction in the population of the other states. A possible explanation for the suppression of the wrong topology is that the small difference in the self-avoid potential is particularly important for the residues not folded into a helix in the native state. The difference causes the chain to fold during the collapsing phase with higher probability in the correct topology.

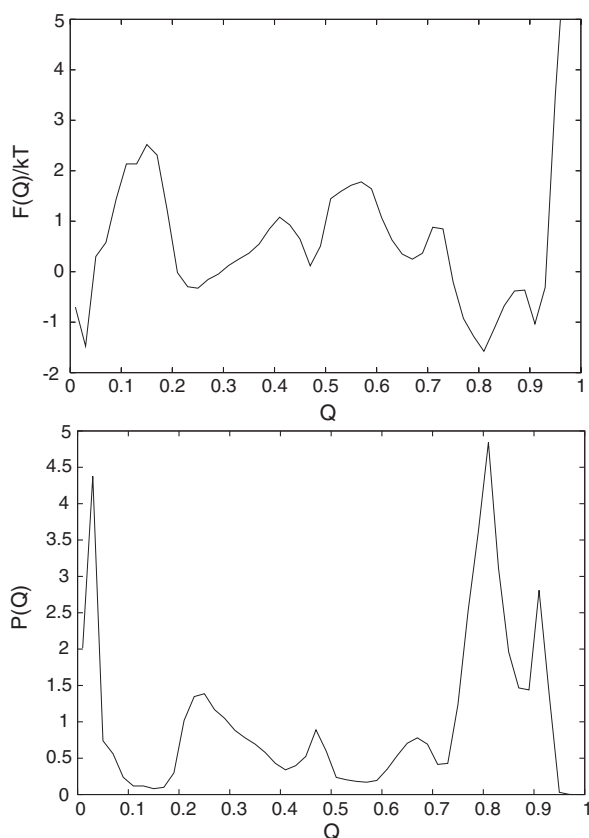


Figure 2. Free-energy $F(Q)/kT$ at $kT = 0.44$ and $\epsilon_{\text{hf}} = 2.9$ (top). The distribution $P(Q)$ (bottom). The two minima in the free-energy located at $Q \approx 0.81$ and $Q \approx 0.91$ correspond to the native topology. The minima at $Q \approx 0.23$ and $Q \approx 0.47$ correspond to states composed mainly of disordered structures. The minimum at $Q \approx 0.03$ is composed by structures folded in a unique long helix.

In figure 3 we show the free-energy profile $F(Q)$ of the unmodified model at $kT = 0.54$ and $\epsilon_{\text{hf}} = 2.4$. This profile was obtained in [1] using simulated tempering. Our modifications suppress the wrong topology but we get a larger amount of disordered structure. The changes in the self-avoid term increase the repulsion between atoms as shown in figure 1; in order to compensate for this effect we have increased the maximum value of ϵ_{hf} to 2.9 and sampled at lower temperatures.

In figure 4 we compare the distribution obtained by the model using parallel tempering with the distribution obtained using the multidimensional replica exchange method. The parallel tempering simulation was done using the same two-dimensional method but setting all the allowed values for ϵ_{hf} to 2.9. The model used in these simulations is characterized by a different geometry of the amino acids responsible for the high probability density peaked at $Q \approx 0.23$. Here we used $\Delta\sigma_{ij} = 0.625$ and no cutoff for the self-avoid potential. This peak corresponds to a variety of different disordered states including four-helix bundles and partially disordered structures with the quasi-specular wrong topology. Our attention focuses on the peak at $Q \approx 0.03$, which is present in the simulation with the multidimensional replica exchange method but nearly absent in the other. Increasing the temperature causes a reduction

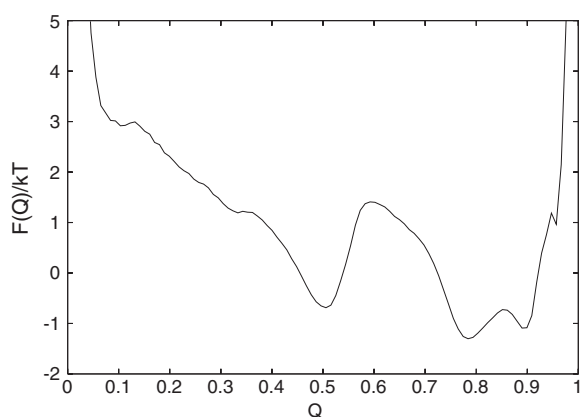


Figure 3. Free-energy $F(Q)/kT$ at $kT = 0.54$ and $\epsilon_{\text{hf}} = 2.4$ obtained from the unmodified model in [1] using simulated tempering.

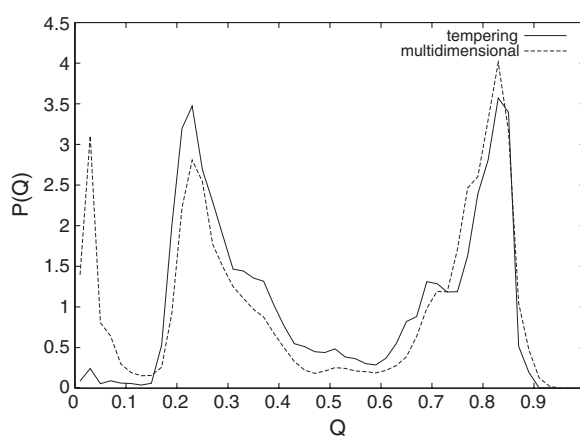


Figure 4. The distribution $P(Q)kT = 0.44$ and $\epsilon_{\text{hf}} = 2.9$ using the parallel tempering and multidimensional replica exchange method. The peak at $Q \approx 0.03$ composed of structures folded in a unique long helix is nearly absent in the parallel tempering simulation.

of the peak at $Q \approx 0.03$ in all cases. This can be easily understood in terms of entropy. Higher temperatures favour more entropic states and the unique long helix cannot compete with more flexible configurations.

A reasonable explanation for the difference between the two simulations is that, when simulating at constant ϵ_{hf} , the probability of getting a totally helical configuration is negligible because of the most entropically favoured collapsed conformations, thus leaving that region of the phase space unexplored. In the multidimensional replica exchange simulation, totally helical states are produced at low ϵ_{hf} values and can reach high ϵ_{hf} values through paths characterized by low temperatures. Promoting ϵ_{hf} to the role of dynamical variable allows for a deeper search in phase space.

Trying to suppress the totally helical state by increasing the value of ϵ_{hf} results in free-energy profiles populated by disordered and collapsed states with poor or not well-defined secondary structure. This is an indication that the effect of the solvent and the electrostatic interaction lack detail, or at least that a more detailed hydrophobic interaction is needed.

In figure 5 we show a sample of a *permanence histogram* relative to a structure started from a Hamiltonian at high temperature and high ϵ_{hf} obtained with the multidimensional replica exchange method. The histogram is nearly flat and ensures that the algorithm is working properly. This means that the structure walked through all the Hamiltonians with flat probability.

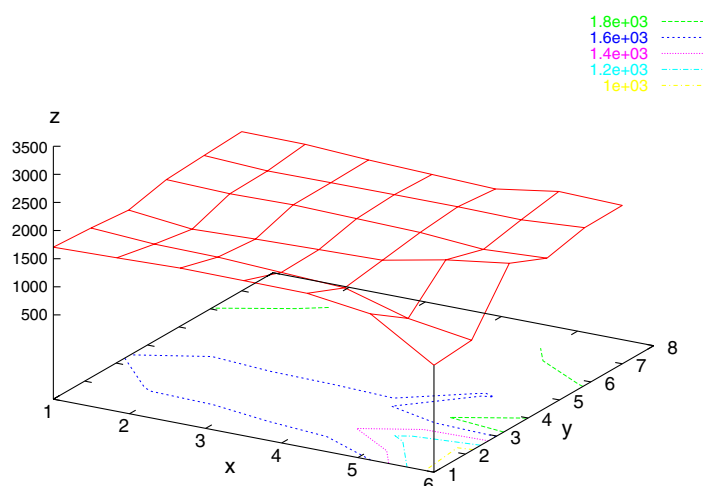


Figure 5. The permanence histogram of a structure obtained using the multidimensional replica exchange method. The number of samples (z axis) is plotted against the index of temperature (y -axis) and of ϵ_{hf} (x -axis).

(This figure is in colour only in the electronic version)

5. Conclusions

We have explored the properties of a modified version of the model presented in [1] with the use of the multidimensional replica exchange method. We have shown that a new self-avoiding term in the Hamiltonian and a different sampling of the configurational space lead to a different free-energy profile in which the population of the quasi-specular topology is lowered.

This indicates that the model is sensitive to small changes in geometry, and thus the self-avoid interaction not only plays a very important role in determining the local conformation (the secondary structure) but also has strong influences on the tertiary structure.

This work also validates the multidimensional replica exchange method as a very efficient tool for the exploration of rugged energy landscapes.

Acknowledgments

We thank Anders Irbäck, Guido Tiana, Giuseppe Marchesini and Claudio Destri for fruitful discussions.

References

- [1] Favrin G *et al* 2001 *J. Chem. Phys.* **114** 8154
Favrin G *et al* 2002 *Protein: Struct. Funct. Bioinf.* **47** 99
- Irbäck A, Sjunnesson F and Wallin S 2000 Three-helix-bundle protein in a Ramachandran model *Proc. Natl Acad. Sci. USA* **97** 13614–8
- [2] Branden C and Tooze J 2001 *Introduction to Protein Structure* (New York: Garland)
- [3] Marinari E and Parisi G 1992 Simulated tempering: a new Monte Carlo scheme *Europhys. Lett.* **19** 451
- [4] Hukushima K and Nemoto K 1995 Exchange Monte Carlo method and application to spin glass simulation *Preprint cond-mat/9512035*
- [5] Sugita Y, Kitao A and Okamoto Y 2000 Multidimensional replica-exchange method for free-energy calculations *J. Chem. Phys.* **113** 6042