LU TP 97-23
December 21, 1997

Simulations of Toy Proteins

Anders Irbäck[1]

Complex Systems Group, Department of Theoretical Physics
University of Lund, Sölvegatan 14A, S-223 62 Lund, Sweden
http://thep.lu.se/tf2/complex/

To appear in
The Proceedings of the Workshop Field Theoretical Tools for Polymer and Particle Physics
Wuppertal, Germany, June 17–19, 1997

Abstract:
Folding properties of two simple off-lattice protein models in two and three dimensions, respectively, are analyzed numerically by using the simulated-tempering method. Both models have two types of "amino acids", hydrophobic and hydrophilic. In the two-dimensional model, a total of 300 randomly selected sequences with 20 monomers are studied. About 10% of these meet criteria for good folders. A statistical analysis of the distribution of hydrophobic monomers along the chains is performed, both for the good folders in this model and for functional proteins. This analysis convincingly show that the hydrophobicity distribution is nonrandom for functional proteins. Furthermore, qualitatively similar deviations from randomness are found for good folders in the model. The study of the three-dimensional model demonstrates the importance of local interactions.

[1] irback@thep.lu.se

1 Introduction

A protein is a linear chain of amino acids, often consisting of 100 to 500 amino acids. A major class of proteins, globular proteins, fold into compact native structures. Each globular protein has a unique native structure, determined by its amino-acid sequence (Anfinsen 1973, Creighton 1993). Understanding the mechanism by which this structure is selected among the huge number of possible conformations is one of the great challenges in molecular biology. The task of predicting the three-dimensional native structure from knowledge of the linear amino-acid sequence is often referred to as the protein folding problem.

There are two major theoretical approaches to protein folding. One is molecular dynamics simulations of fairly detailed models, in which each interaction center represents one or a few atoms (for a review, see e.g. Karplus and Šali 1995). Such simulations can be used for exploring details of the energy landscape near the native structure, and for studying unfolding from the native state. However, the evolution of the system is much too slow for a proper exploration of the full conformational space.

In this situation, simplified "coarse-grained" models have become an increasingly popular complementary tool. These models are often lattice-based and the basic entities are amino acids rather than atoms. Their motivation is roughly analogous to that behind the Ising model for magnetism. They are used to study general characteristics of possible folding mechanisms. Needless to say, they are not intended to reproduce the detailed behavior of specific amino-acid sequences.

Simplified models for protein folding have been studied both analytically and numerically. Useful insights into the phase diagram of such models have been obtained by using ideas from the theory of spin glasses (for a recent review, see Garel et al. 1997). However, in this approach one calculates averages over the form of the disorder, which in this case is the sequence information. This is a limitation since the behavior of sequences that fold well may be different from the average. By using numerical simulations averages over sequence can be avoided.

Most of the simplified models that have been studied numerically are lattice-based with nearest-neighbor contact interactions only. Such simple models have been found to exhibit interesting and nontrivial properties, which is encouraging (for reviews, see e.g. Karplus and Šali 1995 and Dill et al. 1995). However, this does not imply that the approximations involved are understood. Although these models have advantages in terms of computational efficiency, it is therefore important to pursue the study of alternative models.

In this paper I discuss numerical results obtained in two simple off-lattice models in two (2D) and three (3D) dimensions, respectively. The simulations were performed by using the simulated-tempering method (Marinari and Parisi 1992). Both models have only two types of "amino acids", hydrophobic

and hydrophilic, instead of the 20 that occur naturally. The monomers interact through sequence-dependent Lennard-Jones potentials that favor the formation of a hydrophobic core. In addition, the models contain sequence-independent local interactions.

This paper is organized as follows. In Sect. 2 I define the two models and give a brief description of the simulation-tempering method. Section 3 deals with a study of the folding properties of 300 randomly selected sequences in the 2D model. In Sect. 4 the statistical distribution of hydrophobic monomers along chains is examined, both for functional proteins and for good folding sequences in the 2D model. The 3D model is discussed in Sect. 5, focusing on the effects of the local interactions. A brief summary is given in Sect. 6.

2 The Models and the Algorithm

The models studied contain two kinds of monomers, A (hydrophobic) and B (hydrophilic). These are linked by rigid bonds of unit length to form linear chains. A sequence is specified by a choice of monomer types at each position along the chain, $\{\sigma_i\}$, where σ_i takes the values A and B and i is a monomer index. The structure of a chain with N monomers is described by $N-1$ unit bond vectors, $\{\mathbf{b}_i\}$, where $\mathbf{b}_i$ connects monomers i and $i+1$.

The energy function consists of sequence-independent local interactions and sequence-dependent Lennard-Jones potentials. The latter are responsible for the compactification of the chain, and are chosen so as to favor the formation of a core of A monomers.

In the 2D model, introduced by Stillinger et al. (1993), the energy function is given by

$$E = \frac{1}{4}\sum_{i=1}^{N-2}(1-\mathbf{b}_i\cdot\mathbf{b}_{i+1}) + 4\sum_{i=1}^{N-2}\sum_{j=i+2}^{N}\left(\frac{1}{r_{ij}^{12}} - \frac{C(\sigma_i,\sigma_j)}{r_{ij}^{6}}\right) \quad , \qquad (1)$$

where r_{ij} is the distance between sites i and j. The first term in (1) consists of bend potentials that favor alignment of adjacent bond vectors. The species-dependent coefficient $C(\sigma_i,\sigma_j)$ in the Lennard-Jones potential is taken to be 1 for an AA pair, 1/2 for a BB pair, and -1/2 for an AB pair.

The energy function of the 3D model (Irbäck et al. 1997b) is given by

$$E = -\kappa_1\sum_{i=1}^{N-2}\mathbf{b}_i\cdot\mathbf{b}_{i+1} - \kappa_2\sum_{i=1}^{N-3}\mathbf{b}_i\cdot\mathbf{b}_{i+2} + 4\sum_{i=1}^{N-2}\sum_{j=i+2}^{N}\epsilon(\sigma_i,\sigma_j)\left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}}\right) \quad , \qquad (2)$$

where $\epsilon(\sigma_i,\sigma_j)$ is 1 for an AA pair, and 1/2 for BB and AB pairs. The parameters κ_1 and κ_2 determine the strength of the local interactions. The model will be studied for the three choices $(\kappa_1,\kappa_2)=(0,0)$, $(-1,0)$ and $(-1,0.5)$.

Thermodynamic simulations of these models have been performed for different sequences and temperatures, by using the simulated-tempering method

(Marinari and Parisi 1992). At low temperatures conventional Monte Carlo methods tend to become extremely time-consuming, due to the presence of high free-energy barriers. In simulated tempering one tries to overcome this problem by allowing the system to visit higher temperatures where the barriers are lower. More precisely, rather than simulating the Boltzmann distribution at a fixed temperature, one simulates the joint probability distribution

$$P(\mathbf{b}, k) \propto \exp(-g_k - E/T_k) \ , \tag{3}$$

where T_k, $k = 1, \ldots, K$, are the allowed temperatures. The g_k's are tunable parameters that govern the probabilities of visiting the different temperatures. Hence, they must be chosen carefully, which is done by means of trial runs. The joint distribution (3) is then simulated by using separate, ordinary updates of $\{\mathbf{b}_i\}$ and k. Such a simulation directly generates Boltzmann-weighted configurations for each T_k.

Simulated tempering turns out be of great help in simulating these models. For the 2D model speedup factors of 10^3–10^4 were observed, compared with conventional methods (Irbäck and Potthast 1995).

In simulated tempering the temperature becomes a dynamical parameter. It is, of course, possible to apply this idea also to other parameters of the model. An algorithm where the sequence is treated as a dynamical parameter was tested for the 2D model with some success (Irbäck and Potthast 1995). Other dynamical-parameter algorithms have been successfully applied to, for example, the Potts model (Kerler and Weber 1993) and U(1) gauge theory (Kerler et al. 1995).

A method closely related to simulated tempering is the method of multiple Markov chains (Geyer and Thompson 1994, Tesi et al. 1996), also called parallel tempering. Here a set of K allowed temperatures are simulated using K copies of the system, which evolve in parallel. This method has the advantage that there are no g_k parameters to be tuned. The efficiency of this method was tested for the 3D model, with results very similar to those for simulated tempering.

3 Good Folding Sequences

In this section I discuss a study of 300 randomly selected sequences with 14 A and 6 B monomers in the 2D model defined in Sect. 2 (Irbäck et al. 1997a). Both thermodynamic and kinetic simulations were performed for each sequence. The thermodynamic properties are found to be strongly sequence dependent in contrast to the kinetic ones. Hence, criteria for good folding sequences are formulated entirely in terms of thermodynamic properties.

The thermodynamic simulations were performed by using simulated tempering with 13 allowed temperatures, ranging from 0.15 to 0.60. As the temperature is decreased from 0.60 to 0.15, the size of the chains, as measured by e.g. the radius of gyration, decreases substantially. This compactification

is found to take place gradually. The variations in size among the different sequences studied, at a fixed temperature, are found to be small, which is expected since they have the same composition (14 A and 6 B).

The low-temperature behavior shows, nevertheless, a strong sequence dependence. This can be clearly seen from the probability distribution of the mean-square distance between different configurations, δ^2. For two configurations with monomer positions $\{\mathbf{x}_i^{(a)}\}$ and $\{\mathbf{x}_i^{(b)}\}$, respectively, this is defined by

$$\delta^2 = \delta_{ab}^2 = \min \frac{1}{N} \sum_{i=1}^{N} |\mathbf{x}_i^{(a)} - \mathbf{x}_i^{(b)}|^2 \ , \tag{4}$$

where the minimum is taken over translations, rotations and reflections. Figure 1 shows three examples of δ^2 distributions at $T = 0.15$ and 0.60. The corresponding three sequences are given in Table 1. At $T = 0.60$ the three distributions are similar and very broad, showing that the fluctuations in shape are large. At $T = 0.15$ the δ^2 distribution is, by contrast, strongly sequence dependent. One of the three chains (sequence 81) has a well-defined shape at this temperature.

In particular, this implies that these sequences have different folding temperatures T_f. The folding temperature is the temperature where the dominance of a single state sets in. Sequence 81 has highest T_f (> 0.15) among the three sequences in Fig. 1.

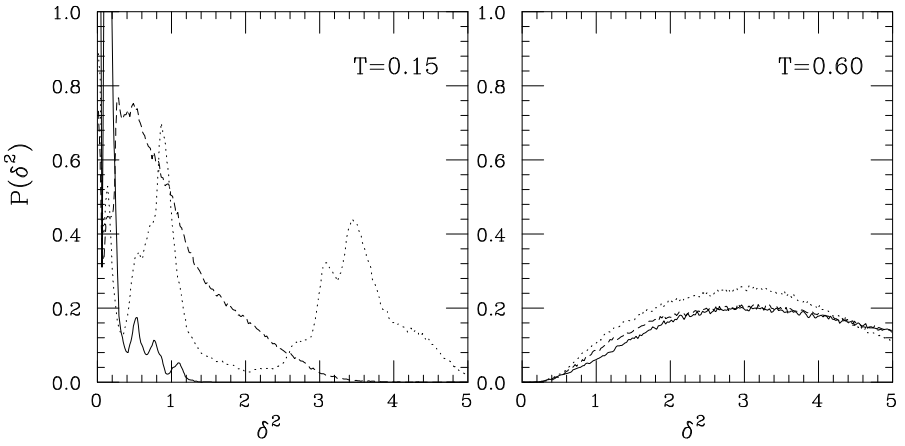

Fig. 1. δ^2 distributions in the 2D model at $T = 0.15$ and 0.60 for the sequences in Table 1; 81 (solid line), 10 (dashed line) and 50 (dotted line). At $T = 0.15$ the distribution for sequence 81 has two narrow peaks at small δ^2 which extend outside the figure with a maximum value of around 20.

Table 1. Three of the sequences studied in the 2D model.

81 AAABAAABAAABBAABBAAA
10 ABAAABBAABAAAAAAAABB
50 BABAAAAAAABAAAABAABB

The energy level spectrum was studied by using a quenching procedure; for a large number of different initial configurations, obtained from the simulated-tempering run, the system was brought to a nearby local energy minimum by means of a conjugate gradient method.

In the subsequent kinetic simulations the mean-square distance δ_0^2 to the minimum energy configuration was monitored. As a criterion of successful folding the condition $\delta_0^2 < 0.3$ was used. The probability of successful folding within a given number of Monte Carlo steps, called the foldicity, was estimated using 25 random initial configurations for each sequence and temperature. Figure 2 shows the foldicity against temperature for the three sequences in Table 1. As expected, the foldicity is found to be low both at high temperatures, where the search is random, and at low temperatures, where the search is hindered by the ruggedness of the free-energy landscape. The freezing temperature where the slow low-temperature behavior sets in, is found to have a weak sequence dependence compared to the folding temperature T_f.

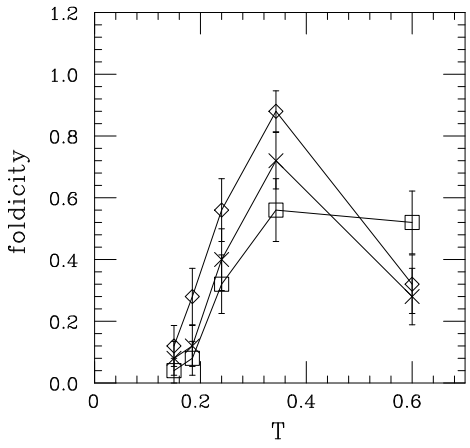

Fig. 2. Foldicity against T in the 2D model for the sequences 81 ($\times$), 10 ($\diamond$) and 50 ($\Box$); see Table 1.

A good folder is a sequence that, at some temperature, exists in a unique and kinetically accessible state with well-defined shape. Having found that the foldicity shows a relatively weak sequence dependence, simplified criteria for good folders were formulated, based entirely on the δ^2 distribution. With these criteria, 37 of the 300 sequences were classified as good folders.

4 Nonrandom Hydrophobicity Patterns

Hydrophobicity plays a central role in the formation of protein structures. To understand the statistical distribution of hydrophobic amino acids along proteins is therefore of great interest. In this section I discuss a study of this distribution (Irbäck et al. 1996) based on the SWISS-PROT protein sequence data bank (Bairoch and Boeckmann 1994) and binary hydrophobicity assignments (± 1). The same analysis is carried out for good folding sequences in the 2D model (see Sect. 3).

The hydrophobicity distribution is analyzed by two different methods, based on block and Fourier variables, respectively. By using these variables rather than the raw sequence variables σ_i, the analysis becomes more sensitive to long-range correlations along the sequence.

Analyzing the behavior of block variables is a widely used and fruitful technique in statistical mechanics, and this application turns out to be no exception. As a first step the sequence is divided into blocks of size s; the block variable $\sigma_k^{(s)}$ is then defined as the sum of the s σ_i's in the block labeled k. The behavior of the block variables will be studied using the normalized mean-square fluctuation $\psi^{(s)}$, defined as

$$\psi^{(s)} = \frac{1}{K}\frac{s}{N}\sum_{k=1}^{N/s}(\sigma_k^{(s)} - s\bar{\sigma})^2 \ , \tag{5}$$

where N is the length of the sequence, and $\bar{\sigma}$ denotes the average of σ_i over the whole sequence. The normalization factor K is chosen such that the average of $\psi^{(s)}$ over all possible sequences with N_+ positive σ_i's and length N, is given by

$$\langle \psi^{(s)} \rangle_{N,N_+} = s \ , \tag{6}$$

independent of N and N_+.

The Fourier analysis is based on a random walk representation of the sequences,

$$\rho_n = \sum_{i=1}^{n}\sigma_i - n\frac{2N_+ - N}{N} \ , \tag{7}$$

which is defined so that $\rho_0 = \rho_N = 0$. With these boundary conditions, the sine transform of ρ_n, to be called f_k, is considered. The average of f_k^2 over all sequences with fixed N and N_+ is found to be

$$\langle f_k^2 \rangle_{N,N_+} = \frac{2N_+(N-N_+)}{N-1} \left(2 \sin \frac{\pi k}{2N} \right)^{-2} \quad . \tag{8}$$

Using this, the normalized quantity $\tilde{f}_k^2 = f_k^2 / \langle f_k^2 \rangle_{N,N_+}$ is formed.

Before applying these methods to protein sequences, there are two important observations to be made. First, the data originating from the ends of the sequences display a different behavior than the data from the rest of the sequences. Therefore, the analysis is performed using the central 70% of the sequences only. Second, sequences with different fractions of hydrophobic amino acids tend to behave in different ways. Because of this, the analysis is performed for a fixed range of the variable

$$X = \frac{N_+ - Np}{\sqrt{Np(1-p)}} \quad , \tag{9}$$

where p is the average of N_+/N over all sequences.

Figure 3a shows the results of the blocking analysis for three different regions in X: $|X| < 0.5$, $|X| > 3$ and all X. The straight line represents random sequences, (6). The results for large $|X|$ lie above this line, while those for small $|X|$ show the opposite behavior. The stability of the results was tested by performing the same analysis both for different intervals in N and for a selected set of nonredundant sequences. Figure 3b shows the results for the good folders in the 2D model. These data show deviations from randomness that are qualitatively similar to those for functional proteins with small $|X|$.

Figure 4a shows the results of the Fourier analysis for $|X| < 0.5$. As one might have expected, there is a peak around the wavelength corresponding to α-helix structure, $2N/k = 3.6$. In addition, there are also clear deviations from randomness at long wavelengths. Such components are suppressed. The results for the good folders in the 2D model are shown in Fig. 4b. Although the statistical errors are somewhat large, there is a clear suppression of the long-wavelength components in this case also.

Here the results have been compared to those for an uncorrelated system. As an illustration of the effects of correlations on the observables $\psi^{(s)}$ and $\tilde{f}_k^2$ one may consider the one-dimensional Ising model. For antiferromagnetic coupling one finds, as for functional proteins with $|X| < 0.5$, that $\psi^{(s)} < s$ and that $\tilde{f}_k^2$ is suppressed for small k. For ferromagnetic coupling the behavior is the opposite, which is what one observes for $|X| > 3$.

The analysis of good folding sequences in the 2D model shows that there exist substantial correlations between certain sequence variables and the folding properties. This suggests that, to some extent, it should be possible to predict folding properties given the sequence. To test this, a feedforward artificial neural network was trained to predict the mean of the δ^2 distribution, at $T = 0.15$, with sequence information only as input (Irbäck 1997a). The results were promising and are likely to improve with a larger data set.

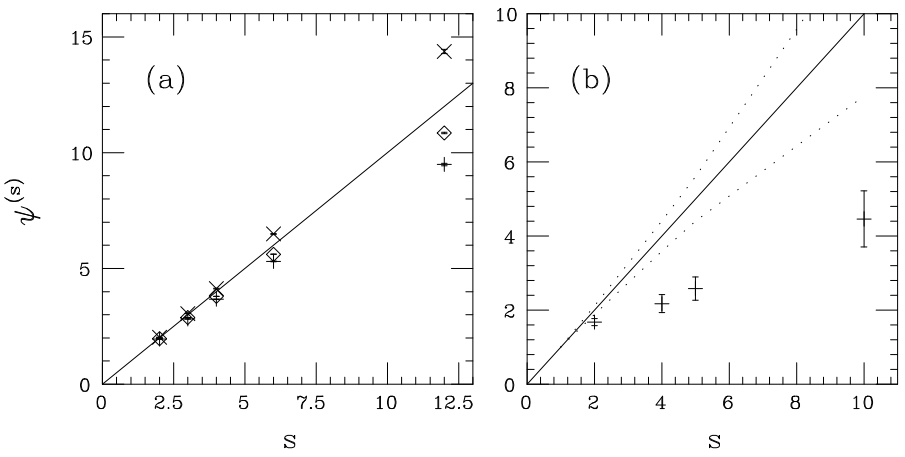

Fig. 3. $\psi^{(s)}$ against block size s. (a) Functional proteins for $|X| < 0.5$ (+; 10154 qualifying proteins), $|X| > 3$ ($\times$; 4928), and all X ($\diamond$; 36765). The straight line is the result for random sequences. (b) Good folders in the 2D model. Also shown are the mean s (full line) and the $s \pm \sigma$ band (bounded by dotted lines) for random sequences.

5 Local Interactions

I now turn to the 3D model defined in Sect. 2, focusing on the effects of the local interactions. Both the local structure of the chains and the δ^2 distributions are discussed, for the three parameter choices $(\kappa_1, \kappa_2) = (0, 0)$, $(-1, 0)$ and $(-1, 0.5)$. The numerical results were obtained using six different sequences of length $N = 20$ (Irbäck et al. 1997b), which were deliberately chosen to represent different types of behavior.

In a 3D off-lattice model, it is possible to check the local properties against those for functional proteins in a direct way. To this end, I consider the structure defined by the protein backbone of C^α atoms. The shape of a chain with N monomers can be specified by $N - 2$ (virtual) bond angles τ_i and $N-3$ torsional angles α_i. In Fig. 5 the distributions of the angles τ_i and α_i are shown, using data from the Protein Data Bank (Bernstein et al. 1977). From these distributions it is evident that there are strong regularities in the local structure. There are two favored regions in the (τ_i, α_i) plane. One of these corresponds to right-handed α-helix, $\tau_i \in [85°, 100°]$ and $\alpha_i \in [35°, 70°]$, and the other to β-sheet, $\tau_i \in [105°, 145°]$ and $\alpha_i \in [170°, 250°]$. Needless to say, the model is not intended to reproduce the precise form of these distributions. It turns out, however, that there are substantial differences in the behaviors for the different (κ_1, κ_2). In particular, the α_i distribution is much flatter, and less protein-like, for $(\kappa_1, \kappa_2) = (0, 0)$ than for $(\kappa_1, \kappa_2) = (-1, 0)$ and $(-1, 0.5)$. Also, two-bond correlations were found to decay very fast with separation along the chain for $(\kappa_1, \kappa_2) = (0, 0)$. This shows that the pure

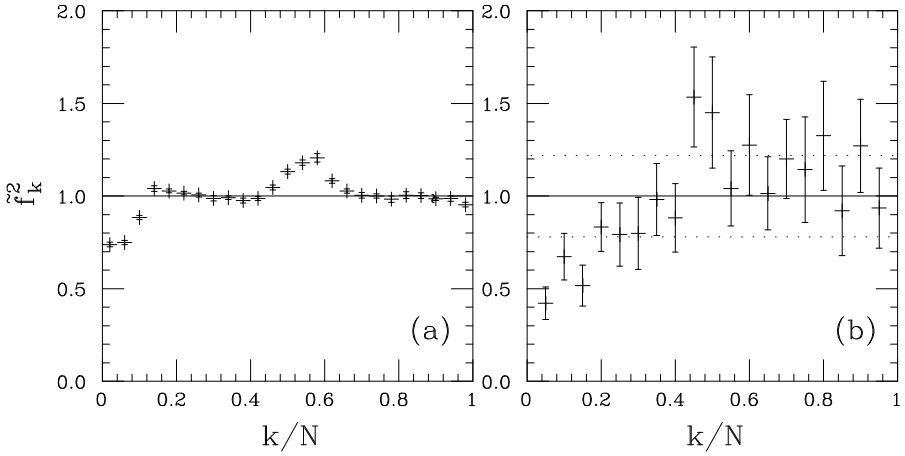

Fig. 4. $\tilde{f}_k^2$ against k/N. (a) Functional proteins for $|X| < 0.5$. (b) Good folders in the 2D model. The full line and dots are as in Fig. 3b.

Lennard-Jones potential is not sufficient to produce the strong regularities in the local structure observed for proteins.

Next I turn to the question of how the choice of (κ_1, κ_2) affects the overall stability of the chains. To study this the δ^2 distributions were computed (see Sect. 3). In contrast to the local quantities discussed above, the δ^2 distribution is strongly sequence dependent. Nevertheless, some general trends were observed. First, the pure Lennard-Jones potential yields very broad δ^2 distributions. Second, the stability tends to be somewhat higher for $(\kappa_1, \kappa_2) = (-1, 0.5)$ than for $(\kappa_1, \kappa_2) = (-1, 0)$. Formation of a compact and well-defined structure was observed for two of the six sequences for $(\kappa_1, \kappa_2) = (-1, 0.5)$.

These results show that local interactions are necessary for structural stability in the 3D model. In the 2D model, where the movements are hampered by compressing one dimension, the local interactions were found to be less crucial. Simulations without the local interaction term were performed for 15 of the 300 sequences, and relatively small changes in the δ^2 distributions were observed.

It should be stressed that the sequences studied contain only two types of monomers. In the work by Yue et al. (1995) a number of two-letter sequences were studied in a 3D lattice model with contact interactions only. The two-letter code was found to be insufficient in the sense that these sequences did not have unique native structures. In models with a larger alphabet more of the possible sequences have unique native structures (Shakhnovich 1994). In the model we have discussed here, with sequence-independent local interactions, it is possible to find two-letter sequences that have unique native structures (see also Irbäck and Sandelin 1997).

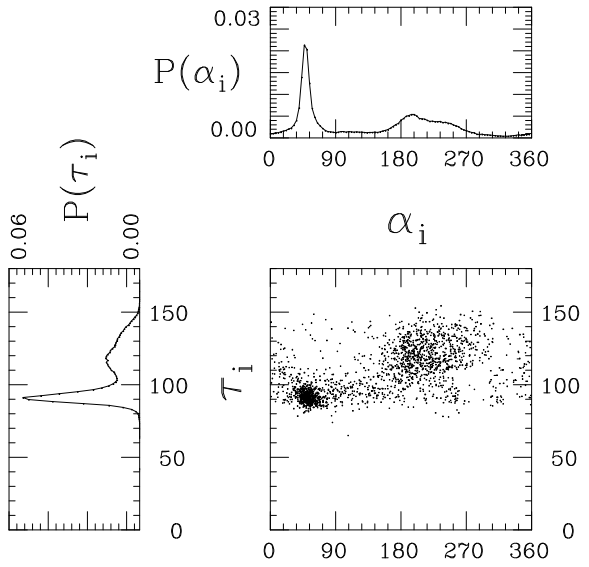

Fig. 5. Virtual bond (τ_i) and torsional (α_i) angle distributions for functional proteins.

6 Summary

Two simple off-lattice models for protein folding have been studied. In the 2D model the termodynamic δ^2 distribution was found to have a much stronger sequence dependence than the kinetic foldicity. Hence criteria for good folding sequences were devised solely in terms of the δ^2 distribution. Approximately 10% of the 300 sequences studied meet these criteria.

Using blocking and Fourier analysis, it was shown that the hydrophobicity distribution of good folders in the 2D model is nonrandom. Qualitatively similar deviations from randomness were observed for low-$|X|$ functional proteins.

In the extension to 3D of the 2D model, the sequence-independent local interactions were found to play a crucial role. It was shown that, in the presence of these, it is possible to find two-letter sequences that have compact, unique native structures. This is in contrast to the results of an earlier study of a 3D lattice model with two-letter code and contact interactions only, where no such sequences were found.

The simulations of these systems were carried out by using simulated tempering, which gives a dramatic improvement of the efficiency compared to conventional methods. Still, as it stands, the results for each $N = 20$ chain in the 3D model require 70 CPU hours on a DEC Alpha 200, so there is room for further improvement.

References

Anfinsen, C.B. (1973): Principles that Govern the Folding of Protein Chains. Science **181**, 223–230

Bairoch, A., Boeckmann, B. (1994): The SWISS-PROT Protein Sequence Data Bank: Current Status. Nucleic Acids Res. **22**, 3578–3580

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. (1977): The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structure. J. Mol. Biol. **112**, 535–542

Creighton, T.E. (1993): *Proteins: Structures and Molecular Properties* (W.H. Freeman, New York)

Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P. Thomas, P.D., Chan, H.S. (1995): Principles of Protein Folding — A Perspective from Simple Exact Models. Protein Science **4**, 561–602

Garel, T., Orland, H., Pitard, E. (1997): Saclay preprint T97-003, to appear in *Spin Glasses and Random Fields*, A.P. Young (ed.) (World Scientific, Singapore)

Geyer, C.J., Thompson, E.A. (1994): Annealing Markov Chain Monte Carlo with Applications to Pedigree Analysis. Preprint, University of Minnesota

Irbäck, A., Potthast, F. (1995): Studies of an Off-Lattice Model for Protein Folding: Sequence Dependence and Improved Sampling at Finite Temperature. J. Chem. Phys. **103**, 10298–10305

Irbäck, A., Peterson, C., Potthast, F. (1996): Evidence for Nonrandom Hydrophobicity Structures in Protein Chains. Proc. Natl. Acad. Sci. USA **93**, 9533–9538

Irbäck, A., Peterson, C., Potthast, F. (1997a): Identification of Amino Acid Sequences with Good Folding Properties in an Off-Lattice Model. Phys. Rev. E **55**, 860–867

Irbäck, A., Peterson, C., Potthast, F., Sommelius, O. (1997b): Local Interactions and Protein Folding: A 3D Off-Lattice Approach. J. Chem. Phys. **107**, 273–282

Irbäck, A., Sandelin, E. (1997): Local Interactions and Protein Folding: A Model Study on the Square and Triangular Lattices. LU TP 97-17, cond-mat/9708049

Karplus, M., Šali, A. (1995): Theoretical Studies of Protein Folding and Unfolding. Curr. Opin. Struct. Biol. **5**, 58–73

Kerler, W., Weber, A. (1993): Cluster Dynamics for First-Order Phase Transitions in the Potts Model. Phys. Rev. B **47**, 11563–11566

Kerler, W., Rebbi, C., Weber, A. (1995): Phase Transition and Dynamical-Parameter Method in U(1) Gauge Theory. Nucl. Phys. B **450**, 452–460

Marinari, E., Parisi, G. (1992): Simulated Tempering: A New Monte Carlo Scheme. Europhys. Lett. **19**, 451–458

Shakhnovich, E.I. (1994): Proteins with Selected Sequences Fold into Unique Native Conformations. Phys. Rev. Lett. **72**, 3907–3910

Tesi, M.C., Janse van Rensburg, E.J., Orlandini, E., Whittington, S.G. (1996): Monte Carlo Study of the Interacting Self-Avoiding Walk Model in Three Dimensions. J. Stat. Phys. **82**, 155–181

Stillinger, F.H., Head-Gordon, T., Hirschfeld, C.L. (1993): Toy Model for Protein Folding. Phys. Rev. E **48**, 1469–1477

Yue, K., Fiebig, K.M., Thomas, P.D., Chan, H.S., Shakhnovich, E.I., Dill, K.A. (1995): A Test of Lattice Protein Folding Algorithms. Proc. Natl. Acad. Sci. USA **92**, 325–329