A minimal proteinlike lattice model: An alpha-helix motif

Piotr Pokarowski, Karol Droste, and Andrzej Kolinski

Citation: J. Chem. Phys. **122**, 214915 (2005); doi: 10.1063/1.1924601
View online: http://dx.doi.org/10.1063/1.1924601
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v122/i21
Published by the American Institute of Physics.

A minimal proteinlike lattice model: An alpha-helix motif

Piotr Pokarowski[a)]
Institute of Applied Mathematics and Mechanics, Warsaw University, Banacha 2, 02-097 Warsaw, Poland

Karol Droste
Institute of Informatics, Warsaw University, Banacha 2, 02-097 Warsaw, Poland

Andrzej Kolinski[b)]
Faculty of Chemistry, Warsaw University, Pasteura 1, 02-093 Warsaw, Poland

(Received 7 February 2005; accepted 5 April 2005; published online 8 June 2005)

A simple protein model of a four-helix bundle motif on a face-centered cubic lattice has been studied. Total energy of a conformation includes attractive interactions between hydrophobic residues, repulsive interactions between hydrophobic and polar residues, and a potential that favors helical turns. Using replica exchange Monte Carlo simulations we have estimated a set of parameters for which the native structure is a global minimum of conformational energy. Then we have shown that all the above types of interactions are necessary to guarantee the cooperativity of folding transition and to satisfy the thermodynamic hypothesis. © *2005 American Institute of Physics*. [DOI: 10.1063/1.1924601]

I. INTRODUCTION

Small globular proteins fold to a well-defined native structure in a way which resembles the first-order phase transition. Reduced lattice models give rigorous account for the physical principles of this cooperative folding transition (see reviews[1–7] and references therein). In this paper we attempt to design a minimal model of protein folding that in a qualitative way reproduces the most pronounced features of globular proteins. On the level of protein geometry, the model should mimic a proteinlike secondary structure, a well-defined hydrophobic core, and a unique topology. On the level of protein energetics, several basic principles should be satisfied. First, the model should address the physics of protein interactions, allowing non-native contacts. Second, a definition of the force field should be universal, at least for the most frequent protein folds. Third, the folding transition temperature should be well defined by the maximum of heat capacity. This condition means that folding should be cooperative. Fourth, the native structure should be a global minimum of free energy for all temperatures below the folding transition. This is a rigorous formulation of Anfinsen's thermodynamic hypothesis.[8] Finally, the model should be a minimal one, i.e., all introduced potentials should be necessary to fulfill the above assumptions.

The protein is modeled as a face-centered cubic (fcc) lattice chain where each residue is reduced to a single united atom located in a lattice point. The native state has a unique topology of a four-helix bundle, which is the simplest and the most frequent α-helical domain (page 38 of Ref. 9). The chain's conformational energy has only three components: attractive interactions between hydrophobic (H) residues, repulsive interactions between hydrophobic and polar (P) residues, and a potential that favors helical turns. We have shown that all components of the force field are necessary to satisfy the requirements of proteinlike thermodynamics. Recently we have designed a complementary β motif,[10] then this work may be regarded as the second step toward designing a minimal force field, universal for the most popular folds.

Our model goes significantly beyond simple cubic lattice models[11–15] and is similar in spirit to several fine-grained reduced models,[16–25] but differs from them in the two crucial ways. First, the tradeoff between resolution and tractability seems to be better. Indeed, we have only three to four parameters as in models on a diamond lattice[16,17] but geometry is almost such realistic as in models on the "210" lattice[19,20] or in off-lattice representations.[18,21–23,25] Second, we present a new approach to the analysis of the folding thermodynamics. (i) Using replica exchange Monte Carlo (REMC) simulations we look for *ruling* structures which could be global minima of conformational energy for certain parameters. (ii) Then we divide space of parameters on *countries* of these structures. The country is a set of interaction parameters for which the ruling structure is of the lowest energy. (iii) Next, we maximize cooperativity in the country of the native structure and (iv) we check if the native is the global minimum of free energy at the folding temperature.

II. MODEL

There are 12 fcc vectors, which form the base set of the lattice, base$=\{\mathbf{e}_0,\mathbf{e}_1,\ldots,\mathbf{e}_9,\mathbf{e}_\mathbf{A},\mathbf{e}_\mathbf{B}\}$, where

$$\mathbf{e}_0=(1,1,0),\quad \mathbf{e}_1=(1,-1,0),\quad \mathbf{e}_2=(1,0,1),$$

$$\mathbf{e}_3=(1,0,-1),\quad \mathbf{e}_4=(0,1,1),\quad \mathbf{e}_5=(0,1,-1),$$

$$\mathbf{e}_6=(0,-1,1),\quad \mathbf{e}_7=(0,-1,-1),\quad \mathbf{e}_8=(-1,0,1),$$

[a)]Electronic mail: pokar@mimuw.edu.pl
[b)]Electronic mail: kolinski@chem.uw.edu.pl

$$\mathbf{e}_9 = (-1,0,-1), \quad \mathbf{e}_\mathbf{A} = (-1,1,0), \quad \mathbf{e}_\mathbf{B} = (-1,-1,0).$$

Points $\mathbf{x}_1, \mathbf{x}_2$ are neighbors on the lattice ($\mathbf{x}_1 \sim \mathbf{x}_2$) if there exists $\mathbf{e} \in$ base such that $\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{e}$. Let chain $= \{1, \ldots, N\}$ be a set of residues in a polypeptide chain. A structure of a polypeptide is represented on the lattice by a one-to-one function $\mathbf{s}$: chain $\rightarrow$ fcc such that $\mathbf{s}(1) = (0,0,0)$ and $\mathbf{s}(i) \sim \mathbf{s}(i+1)$ for $i < N$.

The sequence of a polypeptide chain is defined by its hydrophobic pattern Pat: chain $\rightarrow \{H, P\}$ and its secondary structure Sec: chain $\rightarrow \{\alpha, \gamma\}$, where α denotes a helical residue and γ denotes a flexible loop residue, respectively. Consequently, there are four types of residues, characterized by their hydrophobicity and flexibility.

Let Qua$(\mathbf{s}, i)$ denotes a quadruple of consecutive fcc vectors among points $\mathbf{s}(i-2)$, $\mathbf{s}(i-1)$, $\mathbf{s}(i)$, $\mathbf{s}(i+1)$, and $\mathbf{s}(i+2)$ of a given structure $\mathbf{s}$. We say that Qua$(\mathbf{s}, i)$ forms an *α turn*, if it can be superimposed onto one from the following quadruples of the base vectors: 1287, 2871, 8712, 7128, *A*439, 439*A*, 39*A*4, and 9*A*43. A set of all α turns on fcc is denoted by TRN. The *α helix* in $\mathbf{s}$ is, by definition, a maximal sequence of consecutive residues $\mathbf{s}(i), \ldots, \mathbf{s}(j)$ such that Qua$(\mathbf{s}, i), \ldots,$ Qua$(\mathbf{s}, j)$ form α turns. There are two different helices running in the given direction of fcc. Moreover each helix contains eight different α turns. Hence we have $2 \times 12 \times 8 = 192$ possible α turns on fcc (equivalently #TRN $= 192$).

The above definition of a helix is quite realistic—the model helices are right handed, and their length is equivalent to 1.6 Å/residue, which is close to 1.5 Å seen in real proteins. It is also essential that the helices modeled that way can be packed in a parallel orientation with the β strands introduced in our previous paper.[10] Therefore, it is possible to design reasonable α/β motifs.

Now we can define three types of molecular interactions.

$$K_{HH}(\mathbf{s}) = \#\{\{i,j\}: |i-j| > 3, \mathbf{s}(i) \sim \mathbf{s}(j), \mathrm{Pat}(i) = \mathrm{Pat}(j) = H\} \tag{1a}$$

$$K_{HP}(\mathbf{s}) = \#\{\{i,j\}: |i-j| > 1, \mathbf{s}(i) \sim \mathbf{s}(j), \mathrm{Pat}(i) \neq \mathrm{Pat}(j)\} \tag{1b}$$

$$K_\alpha(\mathbf{s}) = \#\{i: \mathrm{Sec(i)} = \alpha, \mathrm{Qua}(\mathbf{s}, i) \in \mathrm{TRN}\}, \tag{1c}$$

where the symbol # denotes the number of elements in a set. Let $\mathbf{K}(\mathbf{s}) = [K_{HH}(\mathbf{s}), K_{HP}(\mathbf{s}), K_\alpha(\mathbf{s})]$ be the vector defining the numbers of various interactions and $\varepsilon = (\varepsilon_{HH}, \varepsilon_{HP}, \varepsilon_\alpha)$ be the vector of the force field parameters, which mimic physiological conditions. Thus, the conformational energy of a structure $\mathbf{s}$ is defined as a linear combination, $E(\mathbf{s}) \equiv E_\varepsilon(\mathbf{s}) = \varepsilon^T \mathbf{K}(\mathbf{s})$.

K_{HH} and K_{HP} are the numbers of long-range interactions. The condition $|i-j| > 3$ in (1a) is needed to assess that two different helices, running in a given direction, have the same number of intrahelical *HH* contacts. K_α counts the local conformational stiffness of the chain. Usually,[17,19,21] such short-range propensity is defined by three consecutive backbone vectors, but the present potential (four vectors dependent) incorporates also helical hydrogen bonds. For comparison, hydrogen bonds in β sheets have the long-range character, then in our model of the β motif, we have included these interactions in specific geometry of contacts between polar residues.[10] Hence in both our models hydrogen bonds are only implicitly included in the conformational energy.

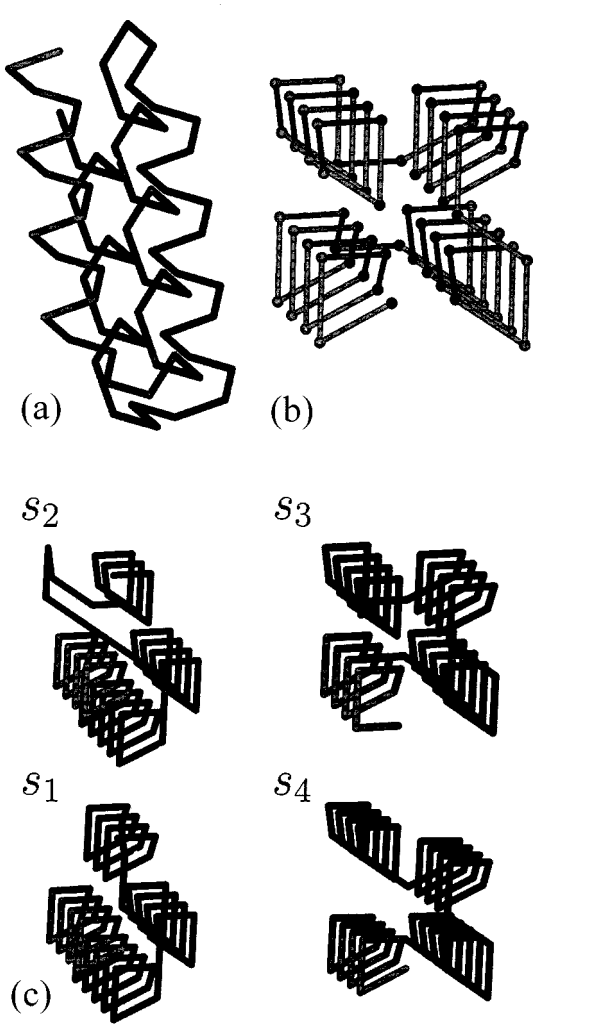

FIG. 1. [(A) and (B)] Model four-helix bundle in two alternative projections. Picture (B) illustrates the well-defined hydrophobic core of the bundle. Hydrophobic residues are shown in black. Note that two helices running in the same direction have different geometry. (C) Snapshots of the four ruling structures which are neighboring to $\mathbf{s}_0$ (definitions are given in the text).

The target structure $\mathbf{s}_0$ is a 69-mer ideal four-helix bundle [Figs. 1(a) and 1(b)], which, using the numbers representing the base vectors, could be abbreviated, as shown in Table I. The following sequence has been designed to be consistent with the above structure: Sec(chain) $= \gamma_2\alpha_{11}\gamma_5\alpha_{13}\gamma_4\alpha_{15}\gamma_4\alpha_{13}\gamma_2$ and Pat(chain) $= (PPHH)_3P(PPHH)_{10}(HPPH)_4$.

III. RESULTS

A. Ruling structures and their countries

A structure $\mathbf{s}^\star$ is called the *ruler* if there exist ε, such that $E_\varepsilon(\mathbf{s}^\star) < E_\varepsilon(\mathbf{s})$ for all $\mathbf{s}$, provided $\mathbf{K}(\mathbf{s}^\star) \neq \mathbf{K}(\mathbf{s})$. The *country* of $\mathbf{s}^\star$, denoted by $C(\mathbf{s}^\star)$, is defined to be a set of all ε for which $\mathbf{s}^\star$ is a ruler. We say that two ruling structures are

TABLE I. Strings of the base vectors and numbers of particular interactions for the native and ruling structures $\mathbf{s}_1,\ldots,\mathbf{s}_4$. To shorten notation, we define four quadruples of the base vectors: $w=8712$, $x=5980$, $y=439A$, and $z=B326$.

Structure	String	K_{HH}	K_{HP}	K_α
$\mathbf{s}_0$	$12w_3820x_35984w_4879x_4$	53	22	52
$\mathbf{s}_1$	$26z_3A8y_48z_4B35x_4$	66	39	52
$\mathbf{s}_2$	$2x_28\ 971\ 844y_48z_4B35x_4$	69	37	50
$\mathbf{s}_3$	$12w_3820x_35984w_4879x_35982$	54	21	51
$\mathbf{s}_4$	$12w_3840x_35984w_4879x_4$	44	19	52

neighboring rulers if their countries are different and have a common border. Anfinsen's hypothesis implies that the native structure $\mathbf{s}_0$ is a ruler. We employed the REMC method to check this fact and to find (eventually) the native's neighboring rulers.

We performed a large number of simulations and this way we have found the native $\mathbf{s}_0$, its neighboring ruling structures $\mathbf{s}_1,\ldots,\mathbf{s}_4$, and another structures $\mathbf{s}_5,\ldots,\mathbf{s}_k$ as global minima of E in the very broad range of ε. Figure 1(c) shows $\mathbf{s}_1,\ldots,\mathbf{s}_4$, Table I gives their string formulas and interaction patterns. Next, we have estimated $C(\mathbf{s}_i)$, $i=0,\ldots,4$ by polygons C_i, which are given by inequalities $E(\mathbf{s}_i)<E(\mathbf{s}j)$, $i=0,\ldots,4$, $j=0,\ldots,k$, $i\neq j$. It is evident that C_i is an upper bound for $C(\mathbf{s}_i)$. On the other hand, we are not able to find any nonempty lower bound for $C(\mathbf{s}_i)$. However, a credible confirmation that C_i is a good estimator of $C(\mathbf{s}_i)$ has been obtained in a large number of independent Monte Carlo simulations.

To compute C_0 we used four inequalities $E(\mathbf{s}_0)<E(\mathbf{s}_i)$, $i=1,\ldots,4$. While neighboring rulers $\mathbf{s}_1,\ldots,\mathbf{s}_4$ were found many times in various simulations, no other ruler was ever recorded in the neighborhood of $\mathbf{s}_0$. Hence, by the interaction patterns (Table I) we obtain

$$-13\varepsilon_{HH}<17\varepsilon_{HP}, \tag{2a}$$

$$-16\varepsilon_{HH}<15\varepsilon_{HP}-2\varepsilon_\alpha, \tag{2b}$$

$$-\varepsilon_{HH}<-\varepsilon_{HP}-\varepsilon_\alpha, \tag{2c}$$

$$-3\varepsilon_{HH}>\varepsilon_{HP}. \tag{2d}$$

Now it is easy to see that our force field is a minimal one. Indeed, (2a) and (2d) give $-\varepsilon_{HH},\varepsilon_{HP}>0$. Then, from (2c) we obtain $-\varepsilon_\alpha>0$. Therefore, we can assume that $-\varepsilon_\alpha=1$ and identify C_i with their projections on $(\varepsilon_{HP},-\varepsilon_{HH})$ which leads to a *thermodynamic map* drawn in Fig. 2. The terms of ruler, country, and thermodynamic map are maybe somewhat unusual but they are very convenient for illustration of analysis of the model parameters. There are only a few structures, which can be in a global minimum of E in a range of realistic values of ε.

Let us note that (i) all $\mathbf{s}_0,\ldots,\mathbf{s}_4$ have a similar secondary structure, a hydrophobic core, and a topology which may mimic conformational mobility of the native structure of real proteins (page 104 of Ref. 9). (ii) The second loop of the designed sequence is one residue longer than the other loops, i.e., there is an amino acid i for which geometrical conditions

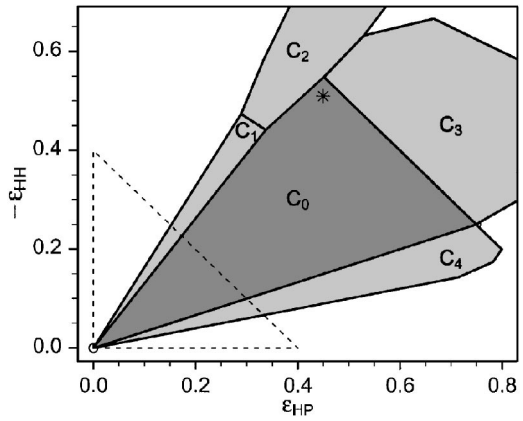

FIG. 2. A thermodynamic map. C_0 and $C_1,\ldots,C_4$ are estimators of the countries of the native structure and its neighboring rulers, respectively. A dashed triangle denotes the area, where folding has not been observed. The star symbol indicates parameters for the analysis of the system's thermodynamics.

of the α propensity in the native structure are fulfilled, $\mathrm{Qua}(\mathbf{s}_0,i)\in\mathrm{TRN}$, but $\mathrm{Sec}(i)=\gamma$. This is the reason why there are two structures $\mathbf{s}_1,\mathbf{s}_4$ which have maximal K_α, but exhibit different packing patterns (denser and looser, respectively) than the native one. This detail is essential for the minimality of our force field. (iii) Similar to the model of the β motif,[10] the native structure has maximal K_α and we are not able to separate $-\varepsilon_{HH}$ and ε_{HP} from 0.

B. Thermodynamics

REMC sampling provided data for the analysis of thermodynamic properties of the model.[26,27] Each experiment employed 12 replicas with 6.5×10^6 attempts to exchange replicas and 10^3 local moves between the exchanges, per replica. The temperatures of replicas were linearly distributed around the estimated (in preliminary simulations) transition temperature. A modified multihistogram method was used for the analysis of data from REMC simulations.[26]

The thermodynamics of the model system is analyzed in terms of the density of states $w(E')=\#\{\mathbf{s}:E(\mathbf{s})=E'\}$. This enables us to define the Boltzmann distribution of states for the model system. $p_T(E')=Z_T^{-1}w(E')\exp(-E'/k_\mathrm{B}T)$, where Z_T is the partition function. We assume without loss of generality that $k_\mathrm{B}=1$. This leads to the definitions of entropy and free energy: $S(E')=k_\mathrm{B}\log[w(E')]$ and $F_T(E')=E'-TS(E')=-k_\mathrm{B}T[\log[p_T(E')]+\log(Z_T)]$. At an infinite temperature the mean system energy can be estimated as $\langle E\rangle_\infty=\Sigma_{E'}E'w(E')/\Sigma_{E'}w(E')$. Subsequently, an equivalent of the system calorimetric enthalpy could be defined as $\Delta E_\mathrm{cal}=\langle E\rangle_\infty-E_\mathrm{native}$. The ratio of van't Hoff and calorimetric enthalpy is a conventional way to measure the transition cooperativity, $\kappa=2T_\mathrm{max}[k_\mathrm{B}C_v(T_\mathrm{max})]^{1/2}/\Delta E_\mathrm{cal}$.

Similar to the previous work, the highest cooperativity has been observed on the edge of C_0, where the contributions from long-range interactions are the largest one. Our results are summarized in Fig. 3.

Figure 3(a) shows that the folding transition temperature $T_f=0.48$ is well defined by the maximum of C_v. In our model the cooperativity is smaller than that observed by others[21,24]

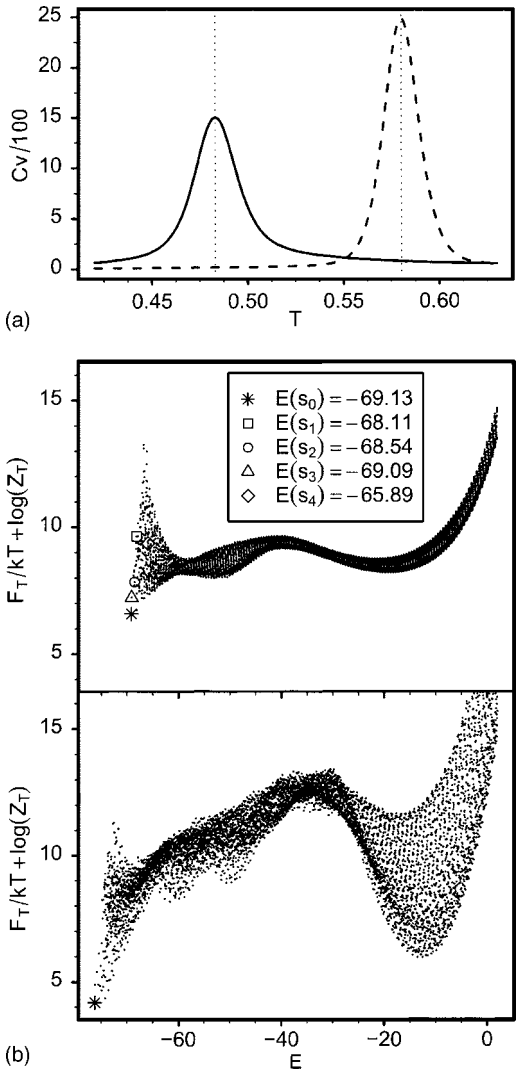

FIG. 3. The thermodynamics of the α motif for $(\varepsilon_{HH},\varepsilon_{HP},\varepsilon_{\alpha})=(-0.51,0.45,-1)$ in comparison with the β-motif (Ref. 10) for $(\varepsilon_{HH},\varepsilon_{HP},\varepsilon_{PP},\varepsilon_{\beta})=(-0.79,0.65,-0.81,-1)$. (A) Heat capacity vs temperature for the α motif (solid line) and the β motif (dashed line). Transition temperatures (vertical dotted lines) and cooperativity coefficients are $T=0.48$, $\kappa=0.47$ for the α motif and $T=0.58$, $\kappa=0.71$ for the β motif (B). Free energy vs conformational energy. The inset gives the values of E for the native structure and its neighboring rulers.

for somewhat similar α motifs, but their force field contained more parameters (7 and 14, respectively). We can easily improve the cooperativity by augmenting specificity of long-range interactions (e.g., we can expand hydrophobic pattern to five types of residues: *H*, *P*, positively charged, negatively charged, and neutral). Another way of enhancing folding cooperativity is to include many-body interactions or a local-nonlocal coupling (reviewed recently by Chan *et al.*[7]). However, such changes lead to increase of the number of parameters and are not consistent with minimality of the model. Let us note that recent off-lattice Go models[28] of chymotrypsin inhibitor 2 exhibit roughly the same cooperativity as our models. Moreover, as explained in that paper,[28] only the base line-subtracted ratio $\kappa^{(s)}$ is a proper measure of the experimentally observed two-state cooperativity. We have obtained $\kappa^{(s)}$ equal to 0.81 and 0.996 for the α and β motif, respectively. The smaller value of κ obtained here for the α motif is probably not a serious drawback of our minimal model, because there are some real differences in cooperativity of α and β motifs. Indeed, unfolded α structures have more native contacts and more native hydrogen bonds, that may lead to a lower energetic barrier and weaker calorimetric cooperativity of the folding transition.

The free-energy plots at the transition temperature are shown in Fig. 3(b). Due to the discrete character of the model there is no single line and the free energy can assume various scattered values for almost the entire range of the system energies. In spite of this, the free-energy barrier between the low- and the high-energy states is well pronounced and provides the signature an all-or-none folding transition. Moreover Fig. 3(b) shows that $\mathbf{s}_0$ is the global minimum of F_T in T_f. Because the thermodynamic map confirms that $\mathbf{s}_0$ is the global minimum of $E=\lim_{T\to 0}F_T$, then it follows easily that the native structure is the global minimum of F_T for every $0\leqslant T\leqslant T_f$. Thus, the thermodynamical hypothesis is satisfied.

IV. DISCUSSION

The present model has the well-defined concept of secondary structure, hydrophobic core, and fold topology, fundamental features of protein geometry. The ground state of the model is similar to the native state of a large class of globular proteins—the four-helix bundle. Similar idealized helical bundles were designed in experimental laboratories[29,30] usually employing only six different types of amino acids. The four-letter code for the model sequence captures fundamental properties of amino acids: hydrophobicity and secondary structure preferences. The amphipatic pattern of our model sequence $(PPHHPPHH)_n$ differs only slightly from the ideal helical pattern of real proteins $(PPHPPHH)_n$. This difference is due to the fcc lattice representation.

The model has no target interaction of Go type, and therefore accounts for yet another important feature of proteins, a competition between various interactions. The short-range conformational preferences of the model account for the hydrogen bonds, excluded volume, and other physical interactions controlling local stiffness of real polypeptides. The long-range interactions of the model mimic the average effect of hydrophobic interactions in proteins. We cannot rigorously derive these interactions as potentials of mean force. However, proteinlike thermodynamics of the model provides strong evidence that its general physics is qualitatively correct.

For these reasons the present model differs qualitatively from the "toy" *HP* models where the ground-state conformation is usually assumed in the form of a maximally compact cube on the simple cubic lattice.[11–15] These toy models satisfy at most just one of the geometrical requirements postulated here for a more realistic proteinlike model. The simple *HP* models may have a hydrophobic core, however, the notion of secondary structure is essentially ignored and the topology of the target ground-state structures is far from being proteinlike.

Our model is similar in spirit to the models studied in the

past[16–25] and bases on the same principles as the model of the β-type structure studied in our previous paper.[10] The set of interactions of the present model closely resembles the force field of the continuous models investigated by Thirumalai and co-workers and by Head-Gordon and co-workers.[18,21–25] All these models are based on a similar reduced representation and alphabet of amino acids. Their conformational energy is composed of short-range propensity and *HP*-type interactions. The models stress upon the interplay between short-range and long-range interactions, which seems to be one of the most fundamental features of the protein's physics.

What then are the specific properties of the model described here that distinguish it from the models studied previously? The following are the most important features of the present approach:

(1) When compared with the simple cubic or diamond lattice protein models,[11–17,24] the present models have a more realistic geometry. In both our models there exist structures with the same secondary structure propensities and topology as in the native structure, but exhibiting denser packing order. Minimal models should mimic conformational mobility of real proteins.

(2) In respect to the continuous models,[18,21–23,25] the present model allows for more precise analysis of its properties. In particular, (i) the native structure is well defined, with unique topology and unique pattern of interactions. (ii) Then it was possible to show in a rigorous fashion, starting from the thermodynamic hypothesis, that the proposed force field is indeed the minimal one. Nonzero values of all the components of the interaction scheme are the consequences of the requirement that the energy of the native structure has to be lower than the energy of all the competitive structures. The minimal character of the force field has been proved here for a single specific sequence. However, it should be pointed out that the findings for the analogous β-type structure[10] were similar and both sequences were designed to mimic the most common sequence patterns seen in small globular proteins. Of course, it cannot be excluded that for some peculiar sequences the minimal requirements for the force field would be different. What our works show is that for the "realistic" sequences the three to four types of interactions are necessary in order to reproduce cooperative, all-or-none, folding transition into a unique nativelike globular structure.

(3) There is one more qualitative difference between the present models and the majority of simple idealized *HP* models studied by others. The simple realizations of the *HP* models assumed attractive interactions only between the hydrophobic residues (with the exception of the model by Socci and Onuchic[11] in which interactions between the polar residues are also attractive). In our previous paper describing the β motif, we have shown that the *PP* type of interactions for the "parallel" on-surface contacts has to be attractive. In the helical model studied here these parallel *PP* interactions are incorporated in the helical turn propensity. Interestingly, a recent comprehensive analysis of various statistical contact potentials (CPs) for globular protein modeling and fold recognition also indicated that interactions between the polar side groups are attractive.[31] We have analyzed 29 matrices of CPs, mostly new and used by groups that were the most successful in predicting protein structure from the amino acid sequence in recent CASP experiments. Almost all studied CPs can be approximated with a correlation greater than 0.8 by the following formula:

$$e_{ij} = c_0 - h_i h_j, \quad 1 \leqslant i,j \leqslant 20, \tag{3}$$

where c_0 is a constant and the residue-type dependent vector $\mathbf{h}=(h_i)$ is highly (0.9) correlated to one of the one-body hydrophobicity scales. In our model of the β motif, long-range interactions can be exactly rewritten in the form (3), where $c_0=-0.075$, $h_H=-0.85$, and $h_P=0.86$. Similarly, in the Socci and Onuchic model[11] we can put $c_0=-2$, $h_H=-1$, and $h_P=1$. Hence the formula (3) describes the dominant property of amino acid interactions leading to attraction between hydrophobic/polar like-type residues and repulsions between unlike-type residues that gives the spatial segregation between a protein's hydrophobic interior and polar surface. In several very successful knowledge-based force fields for protein structure predictions the interactions between the polar groups in the parallel orientation are on average stronger, and more specific, than the interactions between hydrophobic side groups.[32]

V. CONCLUSIONS

The work shows that the proposed model of intraprotein interactions is a minimal one. All components of its force field are necessary for the proteinlike uniqueness of the native structure (Anfinsen's thermodynamic hypothesis) and the two-state folding transition. Therefore, the model provides a very plausible picture of the interplay between the short- and long-range interactions and explains the basic physics of the two-state folding transition.

We believe that this model may provide a universal basement for designing more realistic simplified lattice models. The previous[10] and the present works focused on β and α motifs, respectively. Studies of minimal α/β model polypeptides are now in progress.

ACKNOWLEDGMENTS

This work was supported by the Polish Research Council KBN under Grant Nos. 7 T11F 016 21 and PZB-KBN-088/P04/2003.

[1] M. H. Hao and H. A. Scheraga, J. Mol. Biol. **277**, 973 (1998).
[2] D. Thirumalai and D. K. Klimov, Curr. Opin. Struct. Biol. **9**, 197 (1999).
[3] A. R. Dinner, A. Sali, L. J. Smith, C. M. Dobson, and M. Karplus, Trends Biochem. Sci. **25**, 331 (2000).
[4] J. N. Onuchic, H. Nymeyer, A. E. Garcia, J. Chahine, and N. D. Socci, Adv. Protein Chem. **53**, 87 (2000).
[5] T. Head-Gordon and S. Brown, Curr. Opin. Struct. Biol. **13**, 160 (2003).
[6] A. Kolinski and J. Skolnick, Polymer **45**, 511 (2004).
[7] H. S. Chan, S. Shimizu, and H. Kaya, Methods Enzymol. **380**, 350 (2004).

[8]C. B. Anfinsen, Science **181**, 223 (1973).
[9]C. Branden and J. Tooze, *Introduction to Protein Structure*, 2nd ed. (Garland, New York, 1999).
[10]P. Pokarowski, A. Kolinski, and J. Skolnick, Biophys. J. **84**, 1518 (2003).
[11]N. D. Socci and J. N. Onuchic, J. Chem. Phys. **101**, 1519 (1994).
[12]A. Sali, E. Shakhnovich, and M. Karplus, J. Mol. Biol. **235**, 1614 (1994).
[13]A. Sali, E. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).
[14]K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, Protein Sci. **4**, 561 (1995).
[15]A. R. Dinner, A. Sali, and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **93**, 8356 (1996).
[16]J. Skolnick, A. Kolinski, and R. Yaris, Proc. Natl. Acad. Sci. U.S.A. **86**, 1229 (1989).
[17]A. Sikorski and J. Skolnick, Proc. Natl. Acad. Sci. U.S.A. **86**, 2668 (1989).
[18]J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **87**, 3526 (1990).
[19]J. Skolnick and A. Kolinski, J. Mol. Biol. **221**, 499 (1991).
[20]M. H. Hao and H. A. Scheraga, J. Chem. Phys. **102**, 1334 (1995).
[21]Z. Guo and D. Thirumalai, J. Mol. Biol. **263**, 323 (1996).
[22]J. M. Sorenson and T. Head-Gordon, Proteins **37**, 582 (1999).
[23]J. M. Sorenson and T. Head-Gordon, J. Comput. Biol. **7**, 469 (2000).
[24]H. Kaya and H. S. Chan, Phys. Rev. Lett. **85**, 4823 (2000).
[25]S. Brown, N. J. Fawzi, and T. Head-Gordon, Proc. Natl. Acad. Sci. U.S.A. **100**, 10712 (2003).
[26]M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon, Oxford, 1999).
[27]U. H. Hansmann and Y. Okamoto, Curr. Opin. Struct. Biol. **9**, 177 (1999).
[28]H. Kaya and H. S. Chan, J. Mol. Biol. **326**, 911 (2003).
[29]S. P. Ho and W. F. DeGrado, J. Am. Chem. Soc. **109**, 6751 (1987).
[30]D. P. Raleigh, S. F. Betz, and W. F. DeGrado, J. Am. Chem. Soc. **117**, 7558 (1995).
[31]P. Pokarowski, A. Kloczkowski, R. L. Jernigan, N. Kothari, M. Pokarowska, and A. Kolinski, Proteins **59**, 49 (2005).
[32]A. Kolinski, Acta Biochim. Pol. **51**, 349 (2004).