Folding simulation of protein models on the structure-based cubo-octahedral lattice with the Contact Interactions algorithm

LUCIO TOMA¹ AND SALVATORE TOMA²

¹Dipartimento di Chimica Organica, Università di Pavia, Via Taramelli 10, 27100 Pavia, Italy

²Pharmacia & Upjohn, Via per Pogliano, 20014 Nerviano (MI), Italy

(RECEIVED June 15, 1998; ACCEPTED September 30, 1998)

Abstract

Computer simulations of protein models on lattices have been widely used as an aid in the study of protein folding process. Following the suggestion of Raghunathan and Jernigan (1997, *Protein Sci* 6:2072–2083) that the cubo-octahedral lattice can allow a more realistic representation of proteins than other lattices, we propose here the use of a new set of internal coordinates θ for the description of a protein model on this lattice. An easy procedure for the conversion of the θ coordinates to the Cartesian coordinates is also described. When the Contact Interaction approach, already proposed by us for simulations on square or cubic lattices, was applied to the cubo-octahedral lattice, the system obeyed the correct thermodynamics derived from the definition of energy. Thus, lattice simulations of protein models, in which secondary structure elements such as α -helices or β -strands can be easily identifiable, can be performed.

Keywords: cubo-octahedral lattice; hydrophobic interactions; internal coordinates; lattice models; protein folding

The understanding of the rules governing the process of folding of proteins from an ensemble of denaturated conformations to a single native structure is the object of continuous studies that require a multidisciplinary approach. A new view of protein folding has emerged (Lazaridis & Karplus, 1997) that considers funnel-like energy surfaces (Govindarajan & Goldstein, 1997; Onuchic et al., 1997; Chan & Dill, 1998) in the place of discrete intermediates for the process. Computer simulations have been widely used but, due to the enormous extension of the conformational space of proteins and to its exponential growth with increasing of chain length, several approaches have been proposed that, on one hand, reduce the proteins under study to simplified models made of residues represented by only one or two “atoms” and, on the other hand, allow these models to adopt only a discrete number of conformational states. Usually these simplified models are lattice models.

A lot of work by several groups has been done using the simple square or cubic lattices (Šali et al., 1994; Dill et al., 1995), which have been shown to be of great help in the understanding of the folding process. Other lattices have been used (Park & Levitt, 1995; Agarwala et al., 1997; Irbäck & Sandelin, 1998) and, among them, a new lattice recently proposed by Raghunathan & Jernigan (1997) deserves special attention. In fact, starting from the crystallographic structures of 161 proteins they derived an ideal model for long-

range interactions that allow a regular packing of nonbonded residues; also side-chain position with respect to sequential backbone segments conforms this packing. The experimentally derived spatial arrangement of residues and side chains is well reproduced in its essential features by the cubo-octahedral lattice, a lattice that corresponds to the maximally compact sphere packing. Moreover, contrary to square and cubic lattices, proteins modeled on this lattice have relatively good combination of conformational bond angles and torsion angles so that regular arrangements of sequential neighbors, such as α -helices, β -strands, and various turn types, can be reproduced with high fidelity.

We recently proposed a new algorithm, named the Contact Interactions (CI) method, for the conformational search in the low-energy regions and for simulations of proteins (Toma & Toma, 1996) or homopolymers (Toma et al., 1998), configured as self-avoiding walks on square or cubic lattices, based on a Monte Carlo approach relying on cooling factors associated to single residues rather than on the overall energy of molecules. It proved (Toma & Toma, 1996) quite superior to the standard Monte Carlo algorithm and to Genetic Algorithm approach in the exploration of low energy regions of the conformational space.

As the cubo-octahedral lattice has the essential characteristic to represent proteins realistically, and hence it appears as ideal for protein models simulations, we chose to test the CI algorithm on this lattice. It should allow to run more realistic three-dimensional simulations than those performed on the cubic lattice as secondary structure elements can be taken into account (Raghunathan &

Reprint requests to: Lucio Toma, Dipartimento di Chimica Organica, Università di Pavia, Via Taramelli 10, 27100 Pavia, Italy; e-mail: toma@chifis.unipv.it.

Jernigan, 1997; Kolinski et al., 1998) and the CI algorithm should allow to run dynamic simulations in quite shorter times than other approaches.

Methods

Protein models on the cubo-octahedral lattice

The cubo-octahedral lattice is a version of the face-centered cubic lattice that has connections permitted only between face-centered points and the 12 nearest corners. The 12 vectors, of length $\sqrt{2}$, that connect each lattice site to the adjacent 12 sites are: in the xy plane $(1,1,0)$, $(1,-1,0)$, $(-1,1,0)$, $(-1,-1,0)$; in the yz plane $(0,1,1)$, $(0,1,-1)$, $(0,-1,1)$, $(0,-1,-1)$; in the xz plane $(1,0,1)$, $(1,0,-1)$, $(-1,0,1)$, $(-1,0,-1)$. A protein of N residues modeled on this lattice can be described by a sequence of N connected C^α atoms, each occupying a lattice site, and N C^β atoms, representing the side chain of the residues, also occupying lattice sites. Two residues adjacent in the sequence occupy lattice sites connected by 1 of the 12 vectors defined above (so that in a scale to real proteins each $\sqrt{2}$ vector should correspond to 3.8 Å) and also each C^β is separated by the corresponding C^α by 1 of the 12 vectors. The angle corresponding to a sequence of three connected C^α can be, in principle, 60° , 90° , 120° , and 180° . To make the model consistent with structures of real proteins, the acute bond angle of 60° and the straight 180° angle are discarded for chain generation, so that only angles of 90° and 120° are allowed. The torsion angles corresponding to a sequence of four connected residues can be 0° , 55° , 70° , 109° , 125° , 180° , -125° , -109° , -70° , -55° , and no limitation does exist in this case deriving from the structure of real proteins.

The first problem in describing a protein chain model on this lattice is a suitable choice of coordinates as the absolute xyz Cartesian coordinates alone are unable to allow a good efficiency in the description and evolution of a molecule. A set of internal coordinates, which gives the relative position of a limited group of adjacent residues (i.e., a system which defines the position of one residue with respect to the preceding two or three ones), is more advisable as it allows easy movements in a portion of the structure leaving unchanged the relative positions of the residues at the two sides of the moved point; moreover, the set of internal coordinates should be chosen in such a way to allow an easy transformation into the absolute coordinates. Raghunathan and Jernigan (1997) described the conformations of protein models on the structure-based cubo-octahedral lattice through 16 states defined through the two bond angles and the torsion angle connecting four adjacent residues. We propose here a different definition of internal coordinates.

Given two connected residues (C_1^α and C_2^α) there are only six orientations allowed for the third one as one of the 12 lattice sites surrounding the second residue is occupied by the first one, and five are not available for the third one as bond angles of 60° and 180° are not allowed. The six allowed orientations of the third residue can be defined by an internal coordinate that can assume values 1 to 6. Figure 1 reports a three-dimensional plot of the six allowed orientations. $\theta(3) = 1$ and $\theta(3) = 4$ correspond to bond angles of 90° and the other four cases correspond to bond angles of 120° . Similarly, the fourth residue can be defined with respect to the second and the third ones by $\theta(4)$ that can assume values 1 to 6. It follows that a set of four connected atoms can assume 36 conformations. Due to the sym-

Fig. 1. Three-dimensional plot illustrating the six allowed orientations for a residue with respect the two preceding ones.

metry of the system that doubles all conformations, they are actually 18 distinct conformations.³

The chain can easily grow as six new orientations are allowed for each new residue, apart from the excluded volume condition that can make not valid a new orientation for a residue if it has to occupy an already filled lattice site. The second problem is an easy conversion from the internal set of coordinates to the absolute Cartesian set. Given the first and second residues whose position is defined in the xyz space by vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively, the position of the third residue, $\mathbf{r}_3$, must be calculable through six different operations, corresponding to the six different values of $\theta(3)$, which, acting on $\mathbf{r}_1$ and $\mathbf{r}_2$, give the six possible $\mathbf{r}_3$ (or acting on $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$ give the six possible $\mathbf{r}_{23} = \mathbf{r}_3 - \mathbf{r}_2$). Due to lattice constraints, the vector $\mathbf{r}_{12}$ lies in one of the three orthogonal planes. In Figure 2A it lies, for example, in the xy plane.

Let's define as $\mathbf{t}_1$ the unit vector perpendicular to the plane in which $\mathbf{r}_{12}$ lies [in the example $\mathbf{t}_1$ is $(0,0,1)$]; $\mathbf{t}_2$ and $\mathbf{t}_3$ are the unit vectors [in the example $(1,0,0)$ and $(0,1,0)$, respectively] corresponding to the projections of $\mathbf{r}_{12}$ onto the x and y axes, respectively. The six operations corresponding to $\theta(3) = j$ ($j = 1$ to 6) are (Fig. 2B-G):

$$\theta(3) = 1 \quad \mathbf{r}_{23} = \mathbf{r}_{12} \times \mathbf{t}_1$$

$$\theta(3) = 2 \quad \mathbf{r}_{23} = \mathbf{t}_2 + \mathbf{t}_1$$

$$\theta(3) = 3 \quad \mathbf{r}_{23} = \mathbf{t}_3 + \mathbf{t}_1$$

$$\theta(3) = 4 \quad \mathbf{r}_{23} = -\mathbf{r}_{12} \times \mathbf{t}_1$$

$$\theta(3) = 5 \quad \mathbf{r}_{23} = \mathbf{t}_3 - \mathbf{t}_1$$

$$\theta(3) = 6 \quad \mathbf{r}_{23} = \mathbf{t}_2 - \mathbf{t}_1.$$

Similarly, $\mathbf{r}_{34}$ can be defined by using $\mathbf{r}_{23}$ and $\theta(4)$, and so on.

Let's now consider the side chains defined in the model by the "atoms" C^β ; their position with respect to sequential backbone

³They correspond to the 16 states described by Raghunathan & Jernigan (1997) in their Table 1 but their states 11 and 16 are counted two times in our system as they can present two orientations that appear distinct relatively to the symmetry properties of the lattice.

Fig. 2. Illustration of the three operators t_1 , t_2 , and t_3 (A) that in combination with r_{12} allow to determine the six vectors r_{23} corresponding to the six operations $\theta(3) = j$ ($j = 1$ to 6) (B-G).

segments are relatively regular as the analysis of Raghunathan and Jernigan (1997) showed that in real proteins the torsion angle $C_{i-2}^\alpha - C_{i-1}^\alpha - C_i^\alpha - C_{i+1}^\beta$ (DA2) usually differs by -120° to -190° from the backbone torsion angle $C_{i-2}^\alpha - C_{i-1}^\alpha - C_i^\alpha - C_{i+1}^\alpha$ (DA1). When represented on the cubo-octahedral lattice, the side chains occupy lattice sites and two orientations (states) are allowed for each backbone conformation (state), corresponding to a difference between the above torsion angles of about -120° and 180° . In our representation which utilizes the internal coordinates $\theta(i)$, the allowed orientations of a side chain can be described by the coordinate $\theta_{SC}(i)$ that can be easily derived from $\theta(i+1)$ as $\theta_{SC}(i) = \theta(i+1) + 2$ or $\theta_{SC}(i) = \theta(i+1) + 3$ in correspondence of DA2 – DA1 = -120° or 180° , respectively.⁴

Proper combinations of $\theta(i)$ can characterize α -helix and β -strand regions. Figure 3 reports the three-dimensional structures obtained

from the two θ -sequences: (A) 1–3–4–6–1–3–4–6–1–3–4–6 [for θ -sequence we mean $\theta(3) = 1$, $\theta(4) = 3$, and so on] and (B) 1–5–4–2–1–5–4–2–1–5–4–2 for a 14-residues chain having all $\theta_{SC}(i) = \theta(i+1) + 2$. They are both helices characterized by a pitch of $2.83 \times 3.8/\sqrt{2} = 7.6 \text{ \AA}$, by bond angles alternatively of 90° and 120° and torsion angles of 55° and -55° , respectively. Conformation A is right and conformation B is left-handed; as α -helices in protein structures are always right-handed, conformation A can represent a satisfactory model for α -helix on the cubo-octahedral lattice. Analysis of the orientation of the side chains confirms the analogy of the model with real proteins; in fact the torsion angles that characterize this model DA1 = 55° and DA2 = -55° strictly correspond to the same angles in an ideal α -helix region of a protein where DA1 = $\sim 50^\circ$ and DA2 = $\sim -56^\circ$. A β -strand can be obtained from the two θ -sequences: (C) 2–3–2–3–2–3–2–3–2–3–2–3 or (D) 5–6–5–6–5–6–5–6–5–6–5–6 (Fig. 3), characterized by DA1 = 180° and DA2 = 71° compared to DA1 = 180° and DA2 = $\sim 56^\circ$ present in proteins.

Extension of the conformational space of a protein model

The number of conformations for a sequence of N residues defined only by the backbone C^α atoms each of which can assume six

⁴The use of the states of Raghunathan and Jernigan (1997) can give rise to some confusion: C^α states 11 and 16 in their Table 2 can originate also side-chain states 10 and 7, respectively, in addition to the two states there indicated; moreover, C^α state 7 originates states 10 and 14 (instead of 9 and 10), C^α state 10 originates states 7 and 13, C^α state 12 originates states 9 and 15, C^α state 15 originates states 8 and 12.

Fig. 3. Three-dimensional plot of right- and left-handed α -helices and β -strand modeled on the cubo-octahedral lattice. (A) Right-handed α -helix corresponding to the θ -sequence 1-3-4-6-1-3-4-6-1-3-4-6; (B) left-handed α -helix corresponding to the θ -sequence 1-5-4-2-1-5-4-2-1-5-4-2; (C, D) β -strand corresponding to the θ -sequences 2-3-2-3-2-3-2-3-2-3 or 5-6-5-6-5-6-5-6-5-6-5-6. In all the cases $\theta_{SC}(i) = \theta(i+1) + 2$.

orientations with respect to the preceding two ones is $C_T = 6^{N-2}$ but, owing to the excluded volume condition the number of valid conformations C_V is lower. In Table 1 are reported C_T and C_V for chains up to 14 residues calculated from the complete enumeration of the valid conformers. It can be seen that the growth of the conformational space is very high and the C_{Vi}/C_{Vi-1} ratio remains close to six, almost for sequences of lengths up to 14 residues.

Introduction of the side chains (C^β) makes the problem more complicate, in particular if two different orientations are allowed for each side chain. We chose to consider the presence of the side chains only for internal residues, as a side chain on the first and on the last residue cannot be oriented with the rules described above; in this way, the first and the last C^α atoms serve as N- and C-termini caps. If two orientations are allowed for each C^β , the number of total conformations for a 14-residues chain is $C_T = 6^{N-2} \times 2^{N-2} = \sim 8.9 \times 10^{12}$ with respect to $\sim 2.2 \times 10^9$ total conformations in the case of a single fixed orientation. Thus, we decided in a first approach to take the side chains fixed and chose the orientation of $DA2 - DA1 = \sim -120^\circ$ (not considering at the moment the case in which it is 180°) for two reasons: (1) ideal α -helix and β -strand regions of proteins have $DA2 - DA1 = \sim -120^\circ$; (2) this orientation allows to have chiral residues and hence chiral protein models.

Table 1 reports the results of the complete enumeration of the valid conformations, which shows that in the presence of the side chain about 1×10^9 conformers are valid for the 14-residues sequence, a number suitably tractable for further investigations.

Letter code and definition of energy

We chose to apply the two “color” HP model (Lau & Dill, 1989) on the cubo-octahedral lattice. Thus, only two kinds of residues are allowed: hydrophobic (H) and polar (P). The definition of energy follows from the usual definition of HP interactions on lattices

Table 1. Extension of the conformational space of N -long sequences of proteins modeled on the cubo-octahedral lattice. C_T is the number of total conformations given by the formula 6^{N-2} ; C_V is the number of valid conformations of sequences made only by the C^α “atoms”; C'_V is the number of valid conformations of sequences made of C^α and C^β “atoms” with C^β oriented in such a way that $\theta_{SC}(i) = \theta(i+1) + 2^a$

N	C_T	C_V (without C^β)	C'_V (with C^β)
3	6	6	6
4	36	36	36
5	216	214	208
6	1,296	1,270	1,194
7	7,776	7,500	6,712
8	46,456	44,186	37,368
9	279,936	259,714	206,710
10	1,679,616	1,524,334	1,139,132
11	10,077,696	8,934,126	6,254,912
12	60,466,176	2,303,266	34,261,092
13	362,797,056	305,892,894	187,284,936
14	2,176,782,336	1,787,533,448	1,022,163,386

^aThe number of distinct conformations is one-half of the values reported due to the symmetry properties of the system.

with the additional problem that there are two types of atoms: C^α and C^β . We chose to give no weight to the C^α interactions and to consider only the C^β atoms as hydrophobic or polar. When hydrophobic C^β atoms are adjacent on the lattice a favorable contact is established and a HH interaction with a contribution of -1 units to the overall energy of the molecule is considered. No contribution is given by HP or PP chain interactions.

Test cases for thermodynamic evaluations

We chose two sequences of length 10 and 14 residues (**1** and **2**, respectively) to study the energetics of the system as they are short enough to allow complete enumeration of the valid conformations, and the determination of the density of the states.

Figure 4 reports the three-dimensional plots of their native conformation, which is nondegenerate in the case of **1** and threefold degenerate in the case of **2** (apart from the symmetry of the system which doubles all conformations). Table 2 reports the density of states $n(E)$ obtained from complete enumerations. It is possible to calculate from $n(E)$ the population of states $P(E, c_k)$, i.e., the probability that the chain assumes a conformation of energy E at temperature c_k :

$$P(E, c_k) = [n(E)\exp(-E/c_k)]/Z(c_k) \quad (1)$$

where the partition function Z is the sum of the Boltzmann factors over all the conformational states:

$$Z(c_k) = \sum_E n(E)\exp(-E/c_k). \quad (2)$$

Table 3 (column a) reports the population of states for **1** and **2** at two different temperatures while in Figure 5 are plotted the

Fig. 4. Native conformations of protein models on the cubo-octahedral lattice. **1:** 10-residue chain with HP sequence-HPHHPPHH-: the native conformation has the θ -sequence 3-1-3-5-4-2-3-5 (or the symmetry related sequence 6-4-6-2-1-5-6-2); **2:** 14-residue chain with HP-sequence-HHPHPPPPHPPHH-: the native conformations have the θ -sequences 3-5-4-5-1-3-6-5-6-1-3-5 (or 6-2-1-2-4-6-3-2-3-4-6-2) (**2a**), 3-5-4-5-6-2-6-5-6-1-3-5 or (6-2-1-2-3-5-3-2-3-4-6-2) (**2b**), and 3-6-4-4-1-5-3-2-3-4-6-2 (or 6-3-1-1-4-2-6-5-6-1-3-5) (**2c**).

averaged values of energy E and of fraction of native conformations F_n vs. temperature c_k for the two chains. A transition between an unfolded state (at high c_k) and the native state (at low c_k) is observed both in the plot of E and in the plot of F_n . In both the plots the folding temperature is shifted to higher values for the longer chain.

Table 2. Density of states $n(E)$ of the protein models **1** and **2**, having HP-sequences -HPHHPPHH- and -HHPHPPPPHPPHH-, respectively^a

E	$n(E)$	
	1	2
0	775,624	684,591,464
-1	316,874	286,535,354
-2	39,126	42,663,028
-3	6,580	7,128,478
-4	844	1,061,326
-5	56	153,098
-6	26	25,642
-7	2	4,058
-8		838
-9		80
-10		14
-11		6
$\Sigma n(E)$	1,139,132	1,022,163,386

^aThe number of distinct conformations is one-half of the values reported due to the symmetry properties of the system.

Table 3. Comparison of the relative population of states $P(E, c_k)$ of the protein models **1** and **2** at the two temperatures $c_k = 0.4$ and 0.6 : (a) calculated from the density of states, and (b) determined through simulations 1×10^7 time steps long with the CI algorithm as averages in the second half of the simulations for 1,000 independent molecules

E	1				2			
	$c_k = 0.4$		$c_k = 0.6$		$c_k = 0.4$		$c_k = 0.6$	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
0	0.4	0.4	12.5	12.5	0.0	0.0	8.0	8.1
-1	1.7	1.8	26.9	26.9	0.0	0.0	17.8	18.0
-2	2.6	2.7	17.6	17.6	0.1	0.1	14.0	14.1
-3	5.4	5.3	15.7	15.7	0.2	0.2	12.4	12.6
-4	8.4	8.4	10.6	10.6	0.3	0.3	9.8	9.8
-5	6.8	6.7	3.7	3.8	0.5	0.5	7.5	7.5
-6	38.5	38.8	9.2	9.2	1.1	1.1	6.6	6.6
-7	36.1	36.0	3.7	3.7	2.2	2.1	5.5	5.4
-8					5.4	5.4	6.1	6.1
-9					6.3	5.9	3.1	2.9
-10					13.5	14.7	2.8	2.7
-11					70.3	69.6	6.4	6.2

Application of the CI approach to the cubo-octahedral lattice

The CI algorithm relies on the consideration that, when a contact between two hydrophobic residues is established, e.g., between residues i and j , this interaction creates a heterogeneity in the conformational freedom of the different residues of the molecule; all the residues in the loop from i to j (which we call HH-loop) have less mobility than the residues outside them. We defined (Toma & Toma, 1996) cooling factors, $f(i)$, to describe the mobility of each residue and run dynamic simulations with a Monte Carlo approach based on the values of $f(i)$ rather than the overall energy of the molecule.

Here we apply the CI algorithm to protein sequences modeled on the cubo-octahedral lattice. The conformations of a protein model of N residues are described by $N-2$ internal coordinates $\theta(i)$ defined as above. To the conformation is associated a value of energy $E = -N_{HH}$ where N_{HH} is the number of contacts between hydrophobic side chains. To each residue in the sequence is associated a cooling factor $f(i) = -n_{HH}$, where n_{HH} is the number of HH-loops in which the residue is involved.

Evolution of the system during the simulations can be represented by the following steps: (1) generation of a starting conformation (either extended or randomly generated); (2) random choice of a residue to be moved, for example, the i th; (3) use a criterion of mobility to decide if it has to be moved or not: move it if $Rnd < \exp[f(i)/c_k]$, where Rnd is a random number between 0 and 1; if not, count a time step and go to 2; (4) random choice of the movement, i.e., random choice of the value of $\theta(i)$ while taking invariant all the other θ coordinates (this corresponds to a pivot move); (5) control of the validity (excluded volume condition) of the structure deriving from the movement; if not valid, count a time step and go to 2; (6) if the structure is valid, the new conformation is accepted and its energy evaluated; a time step is

Fig. 5. (A) Averaged energy E and (B) fraction of native conformation F_n calculated from the density of states as function of temperature c_k for the two protein models **1** and **2**.

counted and the cooling factors $f(i)$ assume the values deriving from the contacts present in the new conformation; go to 2.

The procedure for counting time steps is slightly different from that used in our first paper (Toma & Toma, 1996) to have a correct simulation of the thermodynamic parameters (see below). In fact, from the above six steps it results that a time step is counted in all the cases following an attempt to change a $\theta(i)$, also in the cases in which the old conformation is retained because the Monte Carlo condition is not satisfied or the new conformation is not valid. In the previous paper, time steps were counted only in the cases in which new conformations were accepted.

Simulations

We performed several simulations with the CI algorithm on the cubo-octahedral lattice on **1** and **2**, which started from an extended conformation; after a pre-equilibration period at a very high temperature (this corresponds to the generation of a random conformation) they were suddenly cooled to a selected temperature in the range $c_k = 0.3$ – 1.0 . The simulations were long enough to ensure that equilibrium was reached in the first half of the simulation, whereas in the second half data on the population of the various states were collected; in Table 3 (column b) are reported, as examples, the population of states thus obtained at two temperatures as averages from 1,000 independent molecules of **1** and **2**. The agreement between these values and those obtained by complete enumeration (column a) is very good thus indicating that our algorithm obeys the correct thermodynamics derived from the definition of energy.⁵

The efficiency of the approach is demonstrated by the fact that in the proper range of temperatures the process of folding to the native conformations is very fast: for example, at $c_k = 0.4$ half of the 1,000 molecules of **1** reach for the first time the native conformation after about 8,400 time steps (the fastest molecule after

only 35 time steps). Similarly for **2** half of the molecules reach the native conformation in 560,000 time steps (the fastest needing only 500 time steps). Figure 6 shows how fast is the approach to equilibrium at $c_k = 0.4$ by reporting the averaged value of energy vs. time steps; only about 1.5×10^4 time steps are necessary to the ensemble of 1,000 molecules of **1** to reach equilibrium (and about 1.5×10^6 to the ensemble of molecules of **2**).

Discussion

Following the suggestion of Raghunathan and Jernigan (1997) that the cubo-octahedral lattice can allow a more realistic representation of proteins for lattice simulations, we have devised the use of a set of internal coordinates θ , which need for their definition only three contiguous “atoms,” as C_i^α is defined by $\theta(i)$ with respect to C_{i-1}^α and C_{i-2}^α only. In this respect the model, though being a three-dimensional model, presents a complexity typical of a two-dimensional model. In fact, orientation of a residue on a three-dimensional model usually needs two internal coordinates, bond angles (with respect to the two preceding residues) and torsion angles (with respect to the three preceding residues).

Moreover, conversion of the internal coordinates θ to xyz coordinates is very easy as it needs only the vectorial product of unit vectors or the sum of unit vectors, thus having very low time requirements in a computer simulation. Side chains (C^β) can easily be introduced without any need of a new set of coordinates as θ_{SC} are strictly linked to θ . Moreover, if there is the need to change orientation to C^β , changing $DA2 - DA1$ from -120° to 180° or vice versa, this can be made simply by changing the value of θ_{SC} by ± 1 .

Secondary structure elements occasionally formed during a simulation can be easily identified by the θ -sequence as a turn of right-handed α -helix is defined by sequence 1–3–4–6 and a β -strand by sequences 2–3–2–3 or 5–6–5–6.

When the CI approach was applied to this system, very satisfactory results were obtained. In fact, simulations under the cooling factors control allow to reproduce the thermodynamics theoretically estimated from the complete enumeration of a 10 and a 14-residue chain. Assuming that this holds also for longer chains, the CI approach on the cubo-octahedral lattice makes

⁵It is worthy of note that simulations with the CI algorithm on the square lattice with the procedure described in Toma and Toma (1996) follow correct thermodynamics if time steps are counted according to the present paper (L. Toma & S. Toma, unpubl. results).

Fig. 6. Plot of the averaged energy E as function of time steps obtained from simulations on 1,000 independent molecules of (A) 1 and (B) 2.

possible to run simulations for chains for which the complete enumeration of the valid conformations is not possible.

Work is in progress to determine the influence of the sequence on the results, in particular on the factors that can determine the preference for a particular secondary structure element; the CI approach on the cubo-octahedral lattice could allow to test the hypothesis that the only hydrophobic effect is able to determine the native conformation of a protein or, alternatively, the hypothesis that other factors play significant roles.

Acknowledgments

This work has been supported by funds from the Ministero dell'Università e della Ricerca Scientifica e Tecnologica (Roma) and from the University of Pavia.

References

- Agarwala R, Batzoglou S, Dancík V, Decatur SE, Hannenhalli S, Farach M, Muthukrishnan S, Skiena S. 1997. Local rules for protein folding on a triangular lattice and hydrophobicity in the HP model. *J Comput Biol* 4:275–296.
- Chan HS, Dill KA. 1998. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins* 30:2–33.
- Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding—A perspective from simple exact models. *Protein Sci* 4:561–602.
- Govindarajan S, Goldstein RA. 1997. Evolution of model proteins on a foldability landscape. *Proteins* 29:461–466.
- Irbäck A, Sandelin E. 1998. Local interactions and protein folding: A model study on the square and triangular lattices. *J Chem Phys* 108:2245–2250.
- Kolinski A, Galazka W, Skolnick J. 1998. Monte Carlo studies of the thermodynamics and kinetics of reduced protein models: Application to small helical, β , and α/β proteins. *J Chem Phys* 108:2608–2617.
- Lau KF, Dill K. 1989. A lattice statistical mechanics model of the conformational and sequence space of proteins. *Macromolecules* 22:3986–3997.
- Lazaridis T, Karplus M. 1997. “New view” of protein folding reconciled with the old through multiple unfolding simulations. *Science* 278:1928–1931.
- Onuchic JN, Lutheyschulten Z, Wolynes PG. 1997. Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem* 48:545–600.
- Park BH, Levitt M. 1995. The complexity and accuracy of discrete state models of protein structure. *J Mol Biol* 249:493–507.
- Raghunathan G, Jernigan RL. 1997. Ideal architecture of residue packing and its observation in protein structures. *Protein Sci* 6:2072–2083.
- Šali A, Shakhnovich E, Karplus M. 1994. How does a protein fold? *Nature* 369:248–251.
- Toma L, Toma S. 1996. Contact interactions method: A new algorithm for protein folding simulations. *Protein Sci* 5:147–153.
- Toma L, Toma S, Subirana JA. 1998. Simulation of polymer crystallization through a dynamic Monte Carlo lattice model. *Macromolecules* 31:2328–2334.