Chemical Physics 158 (1991) 199–219
North-Holland

Comparison of lattice Monte Carlo dynamics and Brownian dynamics folding pathways of α-helical hairpins

Antonio Rey[1] and Jeffrey Skolnick [2]
Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, CA 92037, USA

Received 21 May 1991

In the context of dynamic Monte Carlo and Brownian dynamics, trajectories of the folding pathways of α-helical hairpin proteins have been computed by two very different models and simulation schemes. The dynamic process is monitored by following the number of native contacts and, in the case of Brownian dynamics, also the torsional angles found along the computed trajectories. An examination of the resulting pathways suggests that the on-site mechanism of assembly previously found in Monte Carlo diamond lattice simulations holds in general for the initial stages of protein folding, and validates their independence with respect to the lattice geometry and the local movements employed in Monte Carlo calculations. Thus, the essential physical character of the computationally very efficient dynamic Monte Carlo simulations is confirmed.

1. Introduction

The application of computer simulation techniques to the elucidation of the properties of proteins has become rather widespread. Different methods corresponding to distinct points of view of the folding problem have been employed. It is clear that an exhaustive sampling of the configurational space is not possible due to the huge number of degrees of freedom involved in the process, and thus two main classes of algorithms have been considered.

In molecular dynamics (MD) calculations [1], a full atom description of the protein and the solvent is depicted (with the exclusion perhaps of hydrogen atoms). The movements of these atoms are computed by integration of the corresponding classical Newtonian equations. This requires the description of a very detailed many-body interaction potential and allows one to obtain the individual particle motions as a function of time. However, the enormous complexity of the system modelled in this way and the limitations in memory and CPU time in contemporary computers seriously limit the scope of this approach. Thus, MD has been mainly applied to the study of small peptides [2] and very local movements in complex systems [3], that occur on a time scale considerably shorter than that corresponding to the folding process.

On the other hand, dynamic Monte Carlo (DMC) simulations have usually taken as their starting point considerably simpler models [4]. In addition, the configurational space is often drastically reduced by considering an underlying lattice whose nodes are the only coordinates in space where the units of the model can be found. In this case, a discrete potential is defined, and a set of individual and possibly artificial movements tries to mimic the main features of the dynamics. Of course, there is not an explicit time step in these algorithms, but if the choice of the movements is carefully implemented, one can assume a direct relationship between the number of Monte Carlo steps (every one of them usually corresponds to a weighted collection of individual movements) and the real time, so that the set of successively computed conformations can be assumed equivalent to a dynamic trajectory.

In previous work, we have employed DMC simulations in a diamond lattice model to investigate the folding features of different idealized simple proteins

[1] Permanent address: Departamento de Quimica Fisica, Facultad de Ciencias Quimicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain.
[2] To whom correspondence should be addressed.

[5,6]. Both in the case of α-helical and β-sheet type structures, an on-site folding mechanism is observed, in which an all-or-none transition takes place. Folding initiates at a central turn, and from that point the different elements of regular secondary structure are assembled one after the other, in a number of simulation steps that is quite reduced in comparison with the total length of the trajectory. The tertiary interactions included in the model stabilize the resulting conformation, and when combined with local secondary structure preferences, yield the native folded state.

This technique has proved to be substantially more efficient than MD calculations in the exploration of long processes involving wide explorations across the configurational space [4]. However, at first glance, it may be questionable whether or not the resulting pathways are physical. First, since the early models were constrained to the geometry of a diamond lattice, one could question if this restricts the possible folding pathways by artificially suppressing a number of channels that could be followed in a conformational space different to that allowed by the diamond lattice. In order to check this point, while still retaining the efficiency of DMC calculations, we have performed these simulations for a completely different lattice model, in which both the geometry and coordination of the lattice are changed, and even the representation of the protein is modified with respect to our previous calculations. The first sections of this paper are devoted to the discussion of the model employed and the results obtained for the folding pathways of α-helical hairpins.

In addition, a second point still remains, and that is whether pathways based on *any* dynamic MC algorithm are physical. As indicated above, no real time is involved in these simulations, and hence the dynamics might correspond to the choice of a set of perhaps partially unphysical movements. While there are reasons to believe this is not the case, an explicit demonstration that this is so would greatly enhance our confidence in the folding pathways produced in DMC simulations.

Therefore, in order to try to check the general validity of those results, in the second part of this paper we employ an alternative methodology, namely Brownian dynamics (BD). In this approach, the solvent is only considered through the dragging effects it causes in the movement of the solute molecules and through the random displacements that it creates in the particles whose sizes are large in comparison with the solvent molecules but are still small enough to be affected by the thermal movement of the solvent [7]. A fundamental real equation of motion for the units of the model is formulated, whose solution provides a dynamic trajectory in which no artifacts are included (other than the definition of the model considered).

BD simulations are not new in the study of the folding problem, but they have been mainly applied to very detailed models of small proteins [1,8] (almost comparable in complexity with those employed in MD, although without including the solvent molecules). In order to keep the study of the whole folding mechanism computationally tractable, these previous studies have introduced a series of simplifications. Namely, certain portions of secondary structure or microdomains are kept frozen in the conformation corresponding to the final native structure. This way, the observed folding pathway a priori corresponds to a diffusion–collision mechanism, where prebuilt elements of secondary structure diffuse one against the other to yield the folded structure [8]. This result is not coincident with the conclusions obtained in DMC simulations in which, as previously mentioned, an on-site construction of both secondary and tertiary structure is observed [4]. We shall show in the last sections of this work that BD simulations, even though restricted to a very simple model, corroborate quite well the DMC simulations.

There is also one other study [9] which uses a sort of BD (a modified stochastic molecular dynamics algorithm) to fold the off-lattice version of our four member β-barrel [6], and while the folding pathway results are sketchy, they too confirm the basic validity of the lattice DMC picture of protein folding.

2. MC simulations

2.1. General description of the model

In order to further enhance the computational advantages of the MC methods, the entire configurational space is embedded onto an underlying cubic lattice. The 210 lattice protein model used here [10]

employs an α-carbon and β-carbon representation of each residue, where each α-carbon occupies a central site along with the six nearest neighbor sites on the cubic lattice. For glycines, this constitutes the entire representation of the residue. For non-glycine type residues, each side chain is described by a β-carbon representation, where the side chain is composed of four lattice sites, three of which are fcc vectors (vectors of the type $[\pm 1, \pm 1, 0]$) from the central α-carbon vertex, and the fourth site is a diamond lattice vector (a vector of the type $[\pm 1, \pm 1, \pm 1]$) which serves as the center of the hydrophobic–hydrophilic interaction. Both hard and soft core repulsive interactions are included between α-carbons. In no case can a lattice site be multiply occupied.

The local conformation of the ith residue is described by the specification of the square of the distance between the centers of α-carbons $i-1$ and $i+1$, r_θ^2, with allowed values of $r_\theta^2 = 6$, 8, 10, 12, 14, 16, and 18. The energetic preferences of these states are specified by $\epsilon_\theta(r_\theta^2)$. To fully specify the conformation of these chains, the dihedral angles associated with consecutive bonds connecting α-carbon centers must also be given. The energy associated with the dihedral angle preference is specified by ϵ_ϕ, and accounts for medium range interactions in the context of the model.

The model also contains a cooperative type interaction that has been included to mimic the effects of dipolar interactions and hydrogen bonding. The strength of this interaction is specified by a parameter ϵ_c, and it allows for secondary structure stabilization. The exact details of this interaction have been described elsewhere [10]. Here, we shall only mention that ϵ_c type interactions can couple conformations down the chain, as well as residues spatially close but far down the chain contour.

Finally, there are interactions between pairs of sidechains. Sidechains are labelled as hydrophobic and hydrophilic. Pairs of hydrophobic residues interact with an attractive potential of mean force [11] whose magnitude is $\epsilon_{\text{phob-phob}}$, typical values of which are 0.75. Pairs of hydrophilic residues interact with a weak repulsive potential of mean force, $\epsilon_{\text{phil-phil}}$, typical values of which are 0.25. Most importantly, hydrophobic–hydrophilic pairs interact with a repulsive potential of mean force $\epsilon_{\text{phil-phob}}$, whose magnitude is typically 1.0. For further details concerning the construction of the sidechain–sidechain interaction matrix, we refer the reader to reference [10].

As in diamond lattice simulations, all the energetic parameters are scaled by a reduced temperature factor, T^*. For a renaturation run (that beginning in an unfolded state), the system is started off at a large T^*, allowed to equilibrate, and then T^* is diminished. This sequence is repeated until the system passes through the transition region and a folded conformation is obtained.

The primary sequence is specified by the convention that $A_i(k)$ denotes the ith stretch in the primary sequence containing k residues that are arranged in an amphipathic aminoacid pattern and which locally prefers helical conformations. For the approximate realization of the α-helix described here, a helix consists of a consecutive sequence of $r_\theta^2 = 12$ states arranged with right-handed chirality. This produces a helix with a four residue repeat, consistent with which every first and fourth residue are hydrophobic and the remainder are hydrophilic. For this class of sequences, $\epsilon_{\theta_i}(12) = 0$ and all the other $\epsilon_{\theta_i}(r_\theta^2) = 0.25/T^*$. It should be pointed out that the above preferences are small in that the native conformation is in no way enforced, and give rise to denatured state populations with helical contents of the order of several percent.

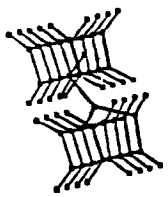
Putative turn regions are denoted by $b_i(j)$, and consist of j residues located at the interface between the putative ith and $(i+1)$th α-helical stretches. Typical turn propensities translate into native-like turn populations of about 1% in the denatured state. Thus, as indicated above, the native conformation has to be found by the dynamic Monte Carlo algorithm, which lacks a target potential associated with tertiary interactions.

2.2. Dynamic Monte Carlo algorithm

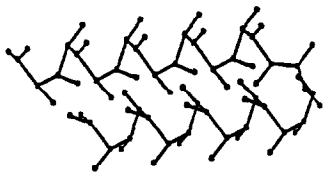
The model chain is subjected to a random sequence of elemental motions previously described [12], whose acceptance is determined by the standard asymmetric Metropolis scheme [13]. The algorithm is capable of rotating and translating assembled pieces of secondary structure; thus, the relative contribution of prefabricated versus on-site construction mechanisms to the assembly process can be

examined. This algorithm rather efficiently samples configuration space, and for the random coil state produces the correct global and local dynamics in the absence of hydrodynamic interactions [12]. Therefore, it is hoped that the observed folding pathways are physically meaningful and independent of the details of the model and choice of local moves. One such check, as stated in section 1, is to compare them with the diamond lattice results [5]. As it turns out, the pathways are qualitatively identical, therefore strongly arguing that the results are universal. Otherwise stated, they are independent of lattice, the realization of the model and the choice of local elemental moves. In the second part of this paper we shall be able to check that, in addition, they are also independent of the simulation theoretical scheme.

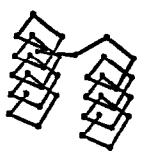
We now examine the nature of the conformational transition from the denatured state to a model α-helical hairpin containing 38 residues, depicted in fig. 1.



1



2



3

Fig. 1. Fully folded conformations of the α-helical hairpin used in DMC simulations shown in top, 1, and side, 2, views that include both α-carbons and the interacting side chain sites. A top view displaying the α-carbon backbone is shown in 3.

The desired native conformation is shown in top, 1, and side, 2, views that include both the α-carbons and the interacting side chain site. A top view displaying the α-carbon backbone alone is shown in 3. The top view, 1, in particular, shows the hydrophobic core, and the side view, 2, shows how these hydrophobic residues interdigitate to produce a close packed interface between the two helices. Observe that, as in real helices, the sidechains point towards the N-terminus. The first (second) helix contains 16 (18) residues. The native turn involves residues 17–20. Residue 17 is the last hydrophobic residue at the C-terminus of the first helix, and residue 20 is the first hydrophobic residue at the N-terminus of the second helix. Residues 18 and 19 are both hydrophilic. The native turn contains residues 17–20 in the bond angle conformations 14, 10, 8 and 14. In the fully native conformation there are 7 pairs of sidechain contacts.

A small chain comprised of 22 residues gives qualitatively identical results to those described below. Here, the longer molecule is chosen so as to increase the likelihood of long lived isolated helices, and hence the possibility of preformed construction as the dominant pathway of assembly. However, as shown below, even under these conditions on-site construction dominates.

2.3. Equilibrium results

Primary sequences of the type $A_1(16)b_1(4)A_2(18)$ were examined, in which for the A_i portions of the sequence, $\epsilon_\theta(12)=0$, and all the other $\epsilon_\theta=0.25/T^*$. For the residues included in putative helical fragments, $\epsilon_\phi=-0.6/T^*$, and the ϵ_ϕ are zero for all the other states. For the turns, $\epsilon_\theta=0$ for the native conformation, and $\epsilon_\theta=0.25/T^*$ for all the other states. $\epsilon_\phi=-0.6(\gamma)/T^*$ for native-like conformations, with $\gamma=0.6$ and 1.0 in two different cases considered, and the ϵ_ϕ are zero for all the other states. For the native conformation, the total short range free energy, $E_\theta=0$, and the total torsional energy, $E_\phi=-20.04/T^*$ and $-21.0/T^*$, for $\gamma=0.6$ and 1.0, respectively. The total side chain interaction free energy, $E_{\text{side}}=-5.25/T^*$, and the cooperative free energy, $E_c=-8.55/T^*$. Thus, the total free energy of the native conformation is $E_N=-33.84/T^*$ and $-34.8/T^*$, for the $\gamma=0.6$ and 1.0 cases. Using an analytical expression previously described [10], an estimated native state

population for the $\gamma=1.0$ case based on short range interactions in the transition region at $T^*=0.41$ of about 10^{-15} is obtained. Thus, based on intrinsic propensities alone, the native conformation is definitely not enforced.

In a total of 5 simulation runs for the $\gamma=0.6$ case and a total of 4 simulation runs for the $\gamma=1.0$ case, the in-register, folded hairpin conformation was obtained every time. However, a variety of turn conformations were observed. Some involved an additional hydrophobic contact between residues 14 and 20. For the $\gamma=0.6$ case, 3 out of 5 runs, and for the $\gamma=1.0$ case, 4 out of 4 runs produced the native turn. The conformational specificity for a unique turn population can be increased, either by further augmenting γ or by changing the intrinsic preference for turns so that the conformation permitting an additional hydrophobic contact is also the lowest energy conformation.

In fig. 2, the number of native contacts, N_C, between pairs of sidechains is plotted versus time (a) at $T^*=0.5$ under denaturing conditions, (b) at $T^*=0.43$, in the transition region, and (c) at $T^*=0.40$, under strongly folding conditions for the $\gamma=1.0$ case. In fig. 2a, as would be expected for unfolding conditions, N_C fluctuates around zero, with only an occasional contact of any sort. In the transition region, fig. 2b, N_C is seen to undergo a relatively rapid jump from 0 to 2 native contacts (this is associated with a native turn plus a single helical turn in one of the helices and a pair of helical turns in the other helix). This conformation persists for 40000 Monte Carlo time steps before the helix completely zips up. Observe that, subsequent to the formation of the helical hairpin, there are substantial fluctuations in N_C; as shown in fig. 2c, these persist at lower temperatures as well. Examination of the folding trajectory indicates that, in addition to the fully helical hairpin, there are a manifold of conformations composed of varying amounts of unfolding at the hairpin ends. Thus, we conclude that an all-or-none transition is not built into the model itself, nor into the Monte Carlo algorithm that provides the conformational sampling. If, for a given set of circumstances, a conformational transition is found to be of an all-or-none character, then this is a consequence of the physics, and it is not obligatory from the folding algorithm that we have employed.

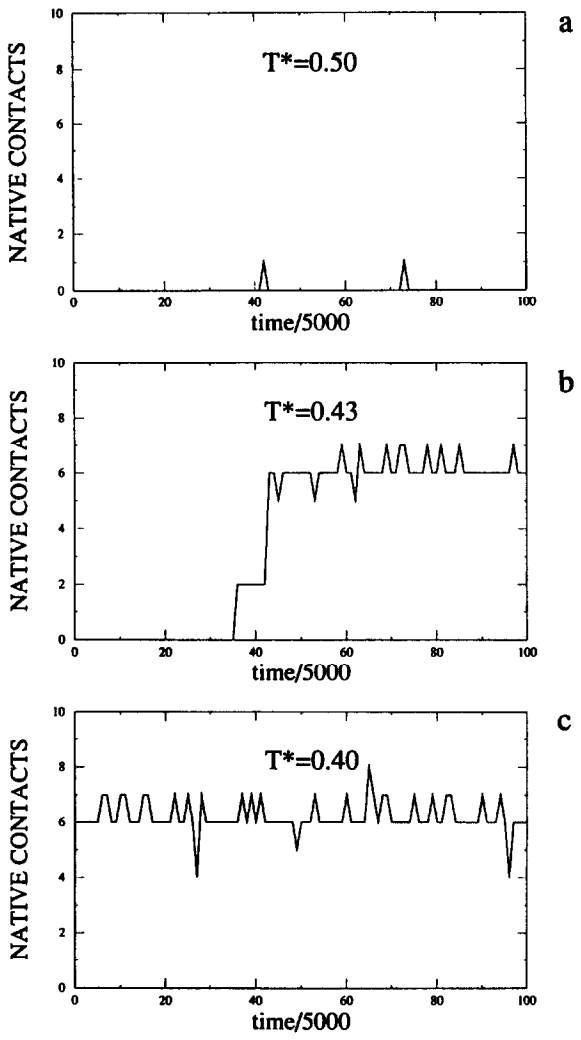


Fig. 2. Number of native contacts between pairs of side chains plotted versus number of DMC steps (a) at $T^*=0.5$, denaturing conditions; (b) at $T^*=0.43$, in the transition region; and (c) at $T^*=0.40$, under strongly folding conditions. All correspond to the $\gamma=1$ case.

2.4. Folding pathways

We next turn to a discussion of the mechanism of assembly of these model hairpins. Two representative folding trajectories will be presented in figs. 3 and 4, both dealing with the $\gamma=1.0$ case. The various snapshots are shown at perspectives that allow the important features of a particular conformation to be readily ascertained. The times indicated in the figure

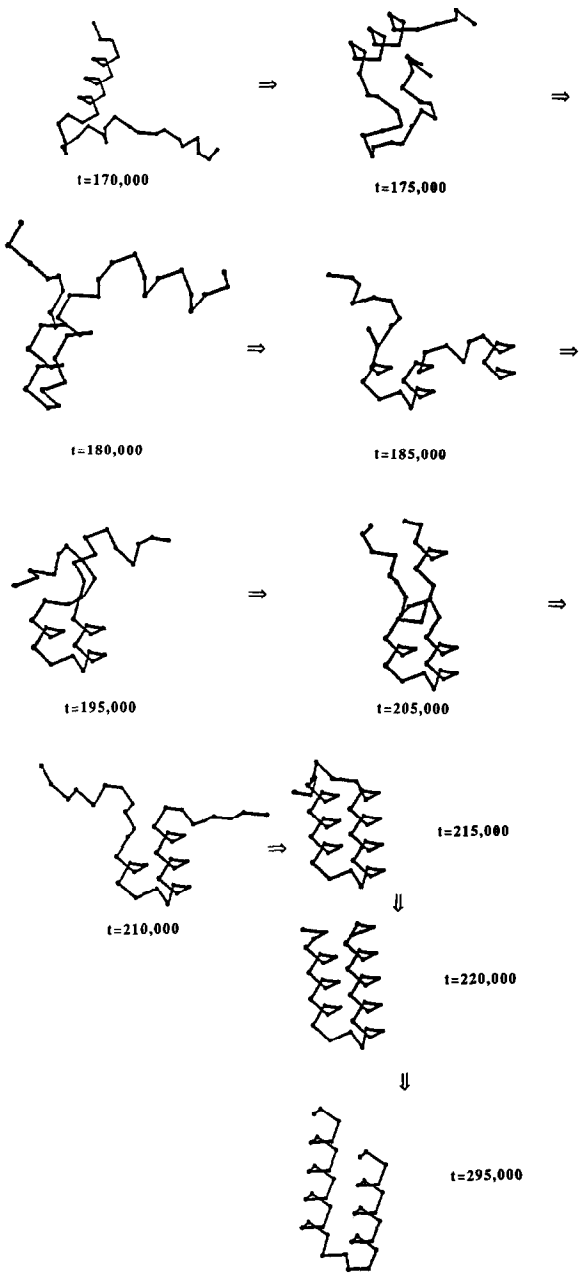


Fig. 3. DMC folding trajectory corresponding to the plot of fig. 2b, at a temperature $T^*=0.43$. See text for details.

(expressed as number of MC steps) are from the start of the simulation run. Fig. 3 presents the folding trajectory corresponding to the N_C versus time plot of fig. 2b, at a temperature $T^*=0.43$. This run is particularly interesting in that an intermediate of two native contact pairs lived for 40000 time steps. At $t=170000$ steps, one of the helices present in the hairpin has formed. By $t=175000$, this helix has partially dissolved. Then, at $t=180000$, the first two native contacts located near the helical turn have formed. These will persist until the successful completion of folding. Observe that there is a second broken helix that could possibly swing into place. However, the $t=185000$ snapshot indicates that this has not happened, and in fact all that is left of the first helix is a single helical turn. The unfolded tails will continue to thrash about until $t=210000$, where the second helix now contains three turns and the first helix has a single turn. Note the presence of extended chain in the unfolded tails. The latter then zips up to form 6 of the 7 native contacts at $t=220000$. It will take until $t=295000$ for the fully assembled helix containing 7 native contacts to form. The elapsed time from the first appearance of a tertiary contact until the construction of the native hairpin took 125000 time steps, with the majority of the conformation formed in 50000 time steps.

Another folding trajectory displayed at a factor of ten finer time resolution than fig. 3 is shown in fig. 4. At $t=548000$, folding of one of the two hairpin helices has initiated. By $t=548500$, three of the helical turns have formed, and by $t=550000$ the helix has zipped up to its end. This single helix persists until $t=552500$, when the first pair of native contacts has formed at the turn. The $t=553500$ snapshot shows that the first helix has partially dissolved, and a single turn of the second helix has formed. The first and second helices have four native contacts, thereby stabilizing this portion of the hairpin. The ends of the helices will continue to fluctuate in helix content. By $t=555500$ both helices contain three helical turns. Finally, by $t=556000$ both helices have four helical turns, and the molecule contains six native contacts. Thus, the folding of the majority of the molecule has occurred in 8000 time steps. It takes until $t=564000$ time steps (not shown as no new information is contained) for the full 7 native contacts to appear for the first time, and the total elapsed time from folding initiation is 16000 time steps.

For the other two runs, both at $T^*=0.41$, folding took 25000 and 35000 time steps, respectively, from the start of successful initiation. Since the statistics

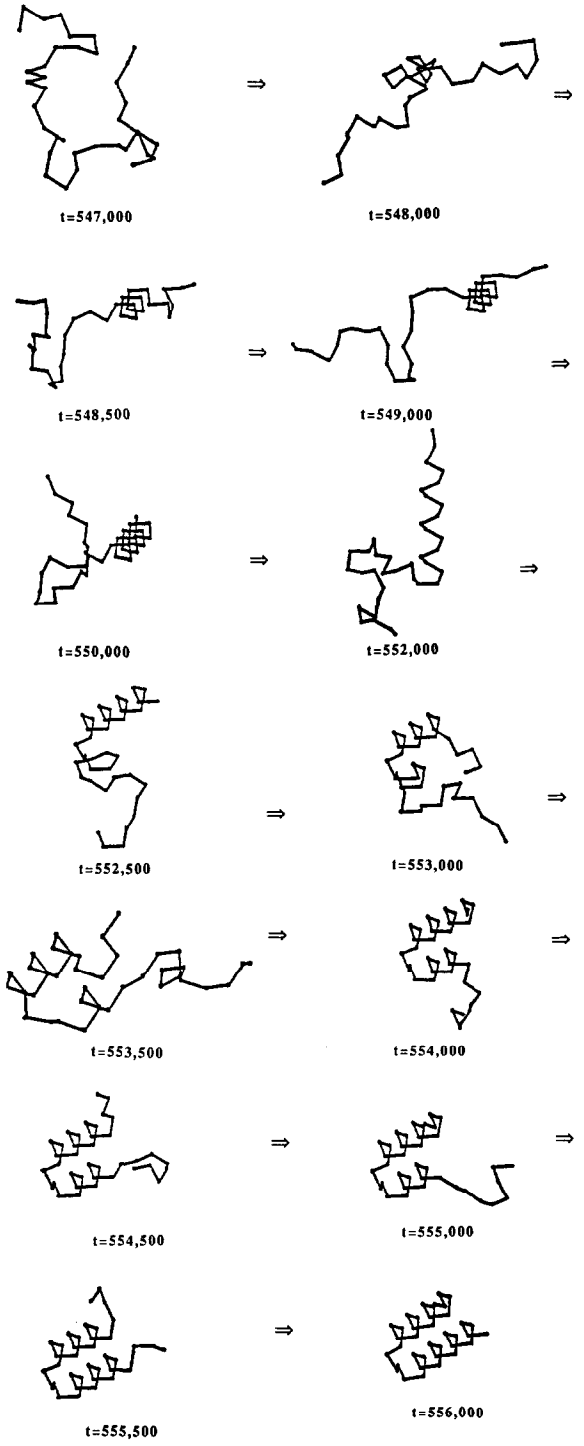


Fig. 4. DMC folding trajectory at a finer time resolution, at $T^*=0.425$.

are far too crude to obtain folding rate constants, we report these values just to give the reader a feel for the order of magnitude involved.

In all the cases observed, folding of helical hairpins has occurred by a sequential zipping up of the helices in place, starting from the turn. In many cases a single helix appears, followed by formation of the second helix beginning at the location of the native turn. It may then almost fully dissolve as in fig. 3, ($t=205000$) with side-by-side construction of both helices subsequently occurring. Other times, as in fig. 4, the majority of the first helix will persist throughout the entire assembly process, with the second helix using the first helix as scaffolding onto which it is constructed. Observed that folding is not unidirectional; helices constantly dissolve and reform during the course of assembly. In fact, many unsuccessful attempts at folding occur prior to a successful initiation. We have also observed folding to initiate from sites quite far from the native turn, in which a hydrophobic sidechain attached to the randomly coiled tail assembles onto a helix that already exists. This conformation, being unstable and non-native, relatively rapidly dissolves. An alternative method of assembly involves the direct initiation from the native turn of the pair of helices. This mechanism is not as common here, because the intrinsic probability of the native turn is very small. Using the previously mentioned analytic theory [10], we obtain an estimate of about a 0.6% turn population in the denatured state at $T^*=0.41$.

Thus, we find that the mechanism of folding is the same as in the diamond lattice model [5]. For both the diamond and the 210 lattice representation, folding of helical hairpins occurs by on-site construction, where at least one of the helices present in the hairpin zips up into place. If both a four and a twenty-four nearest neighbor lattice, using very different DMC moves, give identical results, it is fairly certain that the results are universal; i.e., they are independent of the particular lattice realization. Moreover, it is very suggestive that the results are applicable in general. Let us examine now the results of a completely different approach to the problem to see whether this is indeed the case.

3. Brownian dynamics simulations

3.1. Description of the model

On suppressing the lattice and passing to the continuous space representation necessary for the use of BD algorithms, the calculations become very computer intensive. Thus, in order to keep the system computationally tractable, we employ a very simple model, which is a continuous representation of the diamond lattice model, instead of the more detailed 210 lattice model described in the previous section.

This way, the protein is modeled as a linear chain composed of N spheres, each representing one residue of the real molecule, connected to their first neighbors by means of harmonic springs. No constraints have been introduced into our BD simulations at this stage. The chain motion is governed by the stochastic equation of Langevin, which we solve employing the algorithm originally proposed by Ermak and McCammon [14]. If the initial position of subunit i is $\boldsymbol{r}_i^0$, its position after a dynamics time step Δt is given by

$$\boldsymbol{r}_i = \boldsymbol{r}_i^0 + \left(\frac{\Delta t}{k_B T}\right) \sum_j^N \boldsymbol{D}_{ij}^0 \cdot \boldsymbol{F}_j^0 + \Delta t \sum_j^N \left(\frac{\partial \boldsymbol{D}_{ij}}{\partial \boldsymbol{r}_j}\right)^0 + \boldsymbol{R}_i(\Delta t), \tag{1}$$

where k_B is Boltzmann's constant, T is the absolute temperature, $\boldsymbol{F}_j^0$ is the force acting on unit j at the beginning of the time step, and $\boldsymbol{R}_i$ is a vector of Gaussian random numbers with zero mean and a variance–covariance matrix given by

$$\langle \boldsymbol{R}_i(\Delta t) \boldsymbol{R}_j(\Delta t) \rangle = 2\Delta t \boldsymbol{D}_{ij}^0, \tag{2}$$

$\boldsymbol{D}_{ij}^0$ being the hydrodynamic interaction tensor between units i and j. Here, we have neglected these interactions (that is, we are employing the free-draining approximation). Then, the expression of $\boldsymbol{D}_{ij}$ has the simple form

$$\boldsymbol{D}_{ij} = \left(\frac{k_B T}{6\pi\eta_0 a}\right) \boldsymbol{I} \delta_{ij}, \tag{3}$$

where δ_{ij} is the Kronecker delta, $\boldsymbol{I}$ is the 3×3 unit tensor, a is the radius of the units comprising the model, and η_0 is the solvent viscosity. Obviously, this form of the diffusion tensor is independent of the model configuration, so that the gradient term in eq. (1) vanishes (something that, on the other hand, also happens with the usual expressions for $\boldsymbol{D}_{ij}$ when hydrodynamic interactions are carefully considered [15]).

The Ermak–McCammon method is a first order algorithm; thus, it requires a smaller time step and has a greater numerical error than other more sophisticated algorithms [16]. However, it has been widely used in both macromolecule [17] and biopolymer [18] simulations; its simplicity guarantees the solution of the stochastic equations of motion in a reasonable amount of computer time, and hence it is especially suitable for the calculation of very long trajectories as are needed in our study.

The most important point for the solving of eq. (1) is the form of the forces $\boldsymbol{F}_j$ acting on the different units of the model. Since we want to check the folding pathways obtained from the MC lattice simulations, we have to keep ourselves as close as possible to the potential employed there, taking as our reference the diamond lattice model. However, no lattice is now present in BD simulations, so that instead of the discrete additive contributions defined for the different kinds of interactions, we have to construct a continuous potential whose derivatives provide the force vectors that we need to solve the equations of motion [19].

As usual, the total potential can be split into a series of separate contributions: bond lengths, bond angles, torsional angles, hydrophobic–hydrophilic interactions, and cooperative interactions associated with helix formation (we consider here structures composed only of α-helices and turns). The excluded volume forces, although not explicitly mentioned, have been included in the model through the repulsive part of the hydrophobic–hydrophilic pair potential.

For both bond lengths and bond angles, we consider harmonic potentials of the form

$$U_{bl} = \tfrac{1}{2} k_{bl} \sum_{j=2}^{N} (b_j - b_j^0)^2, \tag{4}$$

$$U_{ba} = \tfrac{1}{2} k_{ba} \sum_{j=3}^{N} (\theta_j - \theta_j^0)^2, \tag{5}$$

where b_j^0 is the equilibrium bond length between units $j-1$ and j, corresponding to the lattice unit, θ_j^0 is the equilibrium bond angle centered in unit $j-1$, that is

made equal to the tetrahedral angle, and b_j and θ_j are the time dependent values of these quantities. The force constants k_{bl} and k_{ba} have the same value for all the units of the model.

The form of the torsional potential is somewhat more complicated. A diamond lattice allows for the three values of the torsional angle ϕ, corresponding to the *trans*(t), *gauche*$^+$ (g^+) and *gauche*$^-$ (g^-) states, which are usually considered in the rotational isomeric state model [20]. We want to consider situations, both in helical fragments and in turns, where only one of the states (e.g., g^- in right handed helices) is energetically favoured with respect to the other two states. In addition, the shape of our potential function (one deeper minimum and two other minima roughly equivalent to each other, as shown in fig. 5) is not well reproduced neither by the usual cosine expansions [19], nor by a Fourier sine and cosine series [21], for a wide set of parameters. Hence, we have employed a more flexible procedure, consisting of the numerical construction of our potential from a combination of harmonic functions, and the subsequent fit of the result to a Redlich–Kistner polynomial

$$U(\phi) = U_{cis} + x_1 x_2 \sum_{i=1}^{n_{terms}} c_i (x_2 - x_1)^{i-1}, \qquad (6)$$

where U_{cis} is the energy corresponding to the cis barrier, $n_{terms} - 1$ is the order of the polynomial, c_i being adjustable coefficients, and x_1 and x_2 are simple functions of the torsional angle (expressed in radians),

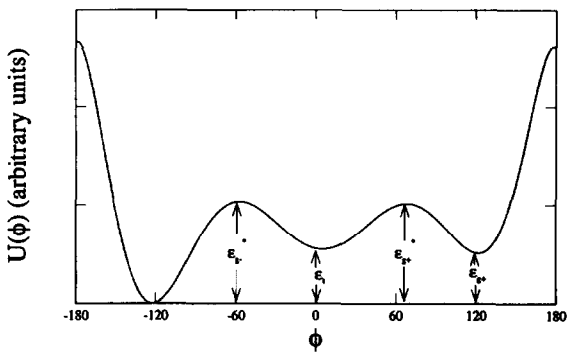


Fig. 5. Shape of the torsional potential employed in BD simulations with favoured helical secondary structure.

$$x_1 = \frac{1}{2} + \frac{\phi}{2\pi}; \quad x_2 = 1 - x_1 = \frac{1}{2} - \frac{\phi}{2\pi}. \qquad (7)$$

According to this definition, and now expressing the torsional angles in degrees, the cis barrier is situated at $\phi = \pm 180°$, and the trans conformation corresponds to $\phi = 0$. The gauche minima are situated at $\phi = \pm 120°$.

For the different parameters employed in the initial numerical construction of the potential (mainly the energetic depth of the three minima and the barriers between them), we have found that a sixth order polynomial ($n_{terms} = 7$) fits the potential function rather well. The fitting is not perfect, since the energetic differences are made smaller and the positions of the minima are slightly shifted from their previously defined values. This is not a real problem, and as a matter of fact is doubly advantageous. For the gauche$^-$ states, it yields right handed helical conformations more similar to actual ones than those obtainable on a diamond lattice. For the trans states, it precludes possible divergences in the computation of the Cartesian components of the forces [19].

For the hydrophobic–hydrophilic interactions, two different options are considered, as was done in the MC diamond lattice simulations [5]. When the interacting residues are a pair of philic–philic or philic–phobic residues, a repulsive potential is defined. On the other hand, for a phobic–phobic pair, a Lennard-Jones potential is considered, where the attractive part appears with a minimum at a distance corresponding to one lattice unit, and a repulsive core acts at short distances. The form of this potential is an 8–6 Lennard-Jones equation, where the exponent of the repulsive part has been lowered to eight from twelve. This is done to avoid a huge increase in the magnitude of the force caused by a close approach of the interacting beads at high temperatures, where the Brownian displacements that mimic the jostling of the protein units due to the thermal agitation of the solvent are more important. Thus, the mathematical forms of these potentials are

$$U_{pho\text{-}pho} = 4\epsilon_1 \left[\left(\frac{\sigma_1}{r} \right)^8 - \left(\frac{\sigma_1}{r} \right)^6 \right], \qquad (8)$$

$$U_{pho\text{-}phi} = U_{phi\text{-}phi} = 4\epsilon_2 \left(\frac{\sigma_2}{r} \right)^8. \qquad (9)$$

In both cases, a cut-off distance is introduced to save lengthy computations. This distance is chosen so that the magnitude of the potential is quite small at it. This way, no corrections are required to account for this small discontinuity in the calculation of the forces.

As previously mentioned, both potentials are repulsive at small distances, so the excluded volume effect does not have to be explicitly included in the model as a separate contribution.

In eqs. (8) and (9), r is the distance between any pair of residues in the model, that is, no target potential associated with tertiary interactions is assumed. On the other hand, the final contribution to the potential, the helical cooperativity interaction, acts only between residues j and $j+4$ when both are included in a helical portion of the protein. These interactions try to mimic the cooperative stabilization of helical structures due to hydrogen bonds (also a rather specific process) and tend to favour the growth of helical fragments without the necessity of forcing the secondary structure of the model through the torsional potential. In this case, the same form of an 8–6 Lennard-Jones potential given in eq. (8) is assumed, although the parameters and the cut-off distance are different (the value of σ, σ_3, is now chosen so that the potential minimum corresponds to the distance between units j and $j+4$ in a helical portion constructed on a diamond lattice).

The combination of this last contribution and the torsional potential favouring the secondary structure corresponding to the native conformation gives as a result a high population of helical states not only in the folded but also in the unfolded state. Although this was the case considered in the initial diamond lattice MC simulations [5], later MC simulations have included weaker native propensities for the secondary structure [22], consistent with the marginal population that is found experimentally in the unfolded situation [23]. This is also the case in the DMC simulations described in the first part of this paper. In order to account for this, we have also run simulations in which the conformational preferences associated with the torsional potential are inverted in the putative helical regions; that is, the gauge$^-$ conformation corresponding to the native state has a *higher* energy than the other two states. In this case, in order to avoid the formation of bulges due to the spherical definition of the helical cooperativity interactions, a restriction is introduced so that it only appears when the two internal torsional angles that define the conformation of residues $j-(j+4)$ are in the correct gauche$^-$ state. Employing the terminology of the helical-coil transition theory [24], the effect of these modifications is to increase the tendency to propagate a helical fragment while simultaneously increasing the cost of helix initiation.

In addition to the parameters defining the different contributions to the potential, an important quantity is the length of the time step Δt appearing in eq. (1). As mentioned previously, the simplicity of the algorithm obliges us to use a small value of Δt. Yet, practicality recommends using as large time step as possible, in order to save computation time. A compromise value was chosen that is still consistent with the average values of bond lengths and bond angles obtained along the trajectory. It could be claimed that constraining these quantities, especially the bond lengths, to constant fixed values would allow for an increase in Δt, without significantly perturbing the physical behavior of the system [25]. While that is true when one is only concerned with the statistical averages of the model geometry, we checked that the use of a larger time step causes problems for those conformations where the hydrophobic interactions attract two phobic residues into compact structures corresponding to the native structure. In this case, even with the reduction in the exponent of the repulsive part included in the potentials of eqs. (8) and (9), a bigger time step sooner or later results in a too close approach of the residues; the corresponding increase in the repulsive force blows the molecule apart in the following simulation step. Hence, a smaller value for Δt is preferable, even at the cost of larger number of steps and computation time, and therefore the introduction of rigid constraints was not useful.

The value of the time step and of the different parameters characterizing the simulation are included in tables 1 and 2. The following reduction of units has been employed in the simulation algorithm: lengths are scaled according to b^0, the equilibrium length between two units of the model (that is assumed identical to the diamond lattice unit, although this last consideration has no effect on these calculations); the friction coefficients appearing in the free-draining diffusion coefficient are reduced by

Table 1
Parameters of the system and the simulation. Torsional potential favouring secondary structure (g^- states favoured in the putative helical regions)

Bond length potential
$b^0=1$ (fundamental length unit)
$k_{bl}\{(b^0)^2/K_B\}\simeq 1.2\times 10^3$ K

Bond angle potential
$\theta^0=109.5°$
$k_{ba}\{1/K_B\}=1.2\times 10^2$ K

Torsional potential (values after fitting to eq. (6))
$\phi_{g\pm}=\pm 105°$
$\phi_t=1.3°$
$\phi^*_{g\pm}=\pm 58°$ (location of the barriers between different minima)
gauche$^-$ state
$\epsilon_{g^-}/K_B=0$ K
$\epsilon_{g^+}/K_B=\epsilon_t/K_B=4.0$ K $\quad \epsilon^*_{g^+}/K_B=\epsilon^*_{g^-}/K_B=13.5$ K
gauche$^+$ state
$\epsilon_{g^+}/K_B=0$ K
$\epsilon_{g^-}/K_B=\epsilon_t/K_B=4.3$ K
$\epsilon^*_{g^+}/K_B=\epsilon^*_{g^-}/K_B=13.6$ K
trans state
$\epsilon_t/K_B=0$ K
$\epsilon_{g^-}/K_B=\epsilon_{g^+}/K_B=3.4$ K
$\epsilon^*_{g^+}/K_B=\epsilon^*_{g^-}/K_B=12.0$ K

Hydrophobic–hydrophilic potential
$r_{cut\text{-}off}/b^0\in(2.5\text{–}3.5)$
$\sigma_1/b^0=\sigma_2/b^0=\sqrt{3/4}\simeq 0.87$
Attractive–repulsive
$\epsilon_1/K_B\in(24\text{–}48)$ K
Purely repulsive
$\epsilon_2/K_B\in(9\text{–}18)$ K

Helical cooperativity potential
$\epsilon_3/K_B\in(12\text{–}30)$ K
$\sigma_3/b^0=2.0$
$r_{cut\text{-}off}/b^0\in(3.5\text{–}4.0)$

Transition temperature: $T_{tr}\in(1.5\text{–}1.7)$ K

Range of the annealing processes: $T\in(3.0\text{–}1.0)$ K

Reduced time step: $\Delta t=2.0\times 10^{-5}$

Table 2
Parameters of the system and the simulation. Inverted torsional potential in the putative helical fragments (the parameters not included in this table are identical to those of table 1)

Torsional potential (values after fitting to eq. (6))
$\phi_{g\pm}=\pm 109°$
$\phi_t=1.0°$
$\phi^*_{g\pm}=\pm 63°$ (location of the barriers between different minima).
gauge$^-$ state
$\epsilon_{g^-}/K_B=10.7$ K
$\epsilon_{g^+}/K_B=\epsilon_t/K_B=0$ K
$\epsilon^*_{g^+}/K_B=\epsilon^*_{g^-}/K_B=19.7$ K
gauge$^+$ state
$\epsilon_{g^+}/K_B=0$ K
$\epsilon_{g^-}/K_B=\epsilon_t/K_B=11.5$ K
$\epsilon^*_{g^+}/K_B=\epsilon^*_{g^-}/K_B=21.5$ K
trans state
$\epsilon_t/K_B=0$ K
$\epsilon_{g^-}/K_B=\epsilon_{g^+}/K_B=9.6$ K
$\epsilon^*_{g^+}/K_B=\epsilon^*_{g^-}/K_B=18.7$ K

Hydrophobic–hydrophilic potential
$r_{cut\text{-}off}/b^0\in(3.0\text{–}3.5)$
$\sigma_1/b^0=\sigma_2/b^0=\sqrt{3/4}\simeq 0.87$
Attractive–repulsive
$\epsilon_1/K_B\in(21\text{–}27)$ K
Purely repulsive
$\epsilon_2/K_B\in(9\text{–}12)$ K

Helical cooperativity potential
$\epsilon_3/K_B\in(78\text{–}96)$ K
$\sigma_3/b^0=2.0$
$r_{cut\text{-}off}/b^0\in(4.0\text{–}4.5)$

Transition temperature: $T_{tr}\in(1.1\text{–}1.2)$ K

Range of the annealing processes: $T\in(1.6\text{–}0.8)$ K

$\xi=6\pi\eta_0 a$, so that it is not necessary to specify the radius a of the beads (although of course at least an estimation is required in order to guess the equivalence with real units). For the potentials, the usual factor k_BT, with k_B being the Boltzmann constant, is employed in the reduction. In these conditions, the time unit is given by $\xi(b^0)^2/k_BT$. Supposing that the solvent is water, a value of $a/b^0=0.3$, $b^0\sim 4$ Å and the transition temperature found in our simulations, the reduced time step $\Delta t=2\times 10^{-5}$ used here roughly corresponds to a real time step of the order of some picoseconds (about three or four orders of magnitude larger than the time steps used in MD calculations). If we consider that real proteins fold in a time scale of the order of seconds or minutes, even taking into account the possible reduction in some order of magnitude due to the small size and simplicity of our model, one can have an idea of the difficulty inherent in this kind of simulations.

Here, it is important to clarify that the apparently small values of the temperature employed in the simulations and listed at the end of tables 1 and 2 result from purely numerical considerations, and are not related with a physical scale. The justification for

these small values is that the potentials, after being transformed to real units, are about two or three orders of magnitude lower than those employed in detailed simulations [26]. The first reason that explains this fact is that we are dealing with virtual bond lengths and bond angles, since our model does not consider all the chemical details of the structure included into the units that comprise it. The second reason is purely practical. All the values have been scaled according to those employed in the construction of the torsional potential, chosen so that the analytical form fitted according to eq. (6) could properly reproduce the desired shape of the potential function. Thus, the temperature scale has to be considered in a relative sense, and does not mean that the folding process is taking place at a regime close to the absolute zero. The units of the temperature have been kept for consistency with the other values in the tables.

In the next section, we shall describe the different folding pathways obtained through both procedures (favoured and unfavoured helical secondary structure) for the case of an α-helical hairpin with $N=22$. The representation of the model is equivalent to 3 in fig. 1, although the number of units is less.

3.2. Analysis of BD trajectories

The starting point for every BD trajectory is an initial conformation randomly constructed on a diamond lattice, the only restriction being the prohibition of multiple occupancy of the lattice sites. It is important to remark once more that this is the only stage during the whole simulation in which the lattice appears, since all the movements constituting the dynamic pathway are computed in a continuous space. In some cases, an annealing process is employed, beginning at a high temperature and progressively reducing it in discrete steps. The number of steps for every temperature is large enough to guarantee thermal equilibrium. The trajectory is stored in all the steps in order to keep track of the time evolution of the system. This annealing sequence was the only way to proceed in the initial simulations, in which the different factors appearing in the global potential of the model made an a priori estimation of the transition temperature very difficult. Once this temperature was determined, several simulations were also run at constant temperature, the results obtained being comparable to the corresponding annealing simulations.

The quantitative analysis of the simulations poses some practical problems. Since the size of the molecule is smaller than in the cases considered in the diamond [5] or the 210 lattice, and the number of native contacts in the folded conformation is also reduced, the monitoring of these quantities as a function of time or simulation steps is not usually enough to create a good picture of the dynamics. Similarly, for small molecules the radius of gyration $\langle S^2 \rangle$ does not substantially differ between the folded and the unfolded states, and thus it is not shown. On the other hand, a simultaneous enumeration of the native contacts, together with the torsional angles that have the value corresponding to the folded structure, allows one to visualize the folding pathway. For further clarity, in one case we also include snapshots obtained along the trajectory, that help to understand the folding process. It is important to remember that the chain is composed of 22 units, arranged into two right handed α-helices joined together via a central turn. The sequence of states for the 19 internal torsional angles corresponding to the native conformation is $(g^-\text{-}g^-\text{-}g^-\text{-}g^-\text{-}g^-\text{-}g^-\text{-}g^-)\text{-}(t\text{-}g^-\text{-}g^+)\text{-}(g^-\text{-}g^-\text{-}g^-\text{-}g^-\text{-}g^-\text{-}g^-\text{-}g^-\text{-}g^-\text{-}g^-)$, where the brackets are included to mark the different regions (helix I)-(turn)-(helix II). In addition, there are 5 native contacts between pairs of hydrophobic residues in the folded state. Proceeding out from the central turn, the pairs of residues involved in these contacts are (14-9), (15-6), (18-5), (19-2), and (22-1). The other 12 units are considered as hydrophilic residues in the calculation of the tertiary interactions.

We begin by considering the trajectories computed in the context of the model that favours the native secondary structure; that is, one whose potential for the torsional angles has a deeper minimum in the state corresponding to the native conformation (g^- in the helical fragments, and the corresponding states in the central turn as described above). A representative example of the folding pathway obtained is shown in fig. 6, where the existing native contacts and the correctness of the torsional state are monitored as a function of the number of simulation steps. A spot in these plots indicates the existence of the correct contact or torsional angle.

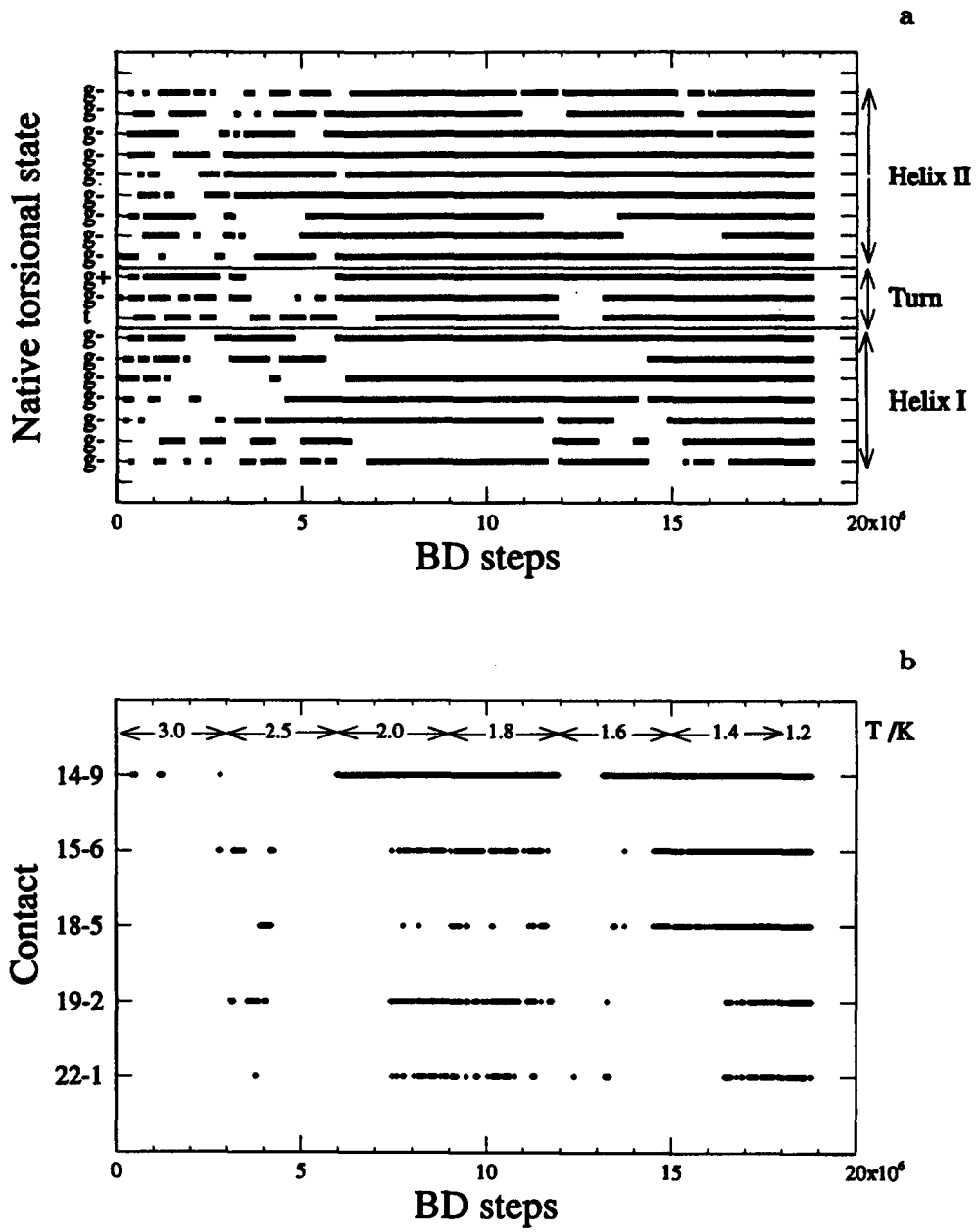


Fig. 6. Results of an annealing BD folding trajectory with g^- favouring torsional potential in the helical regions. (a) Time evolution of the torsional states with the number of simulation steps. (b) Dynamic mapping of the native contacts. See text for details.

When the temperature is well above the transition value, quite chaotic movement is observed, mainly driven by the amplitude of the random displacements of the units. The transition rate between different torsional states is quite high, and although some native contacts can be sporadically observed, the magnitude of the tertiary interactions in these conditions is not sufficient to keep them stable. Of course,

since the phobic–phobic interaction is not restricted to the native contacts, many other contacts occur from time to time without being recorded in fig. 6. In summary, the behavior of the chain at high temperature is in complete agreement with that expected for a random coil.

When the temperature is decreased, the influence of the Brownian component of the movements is reduced, while the effect of the forces corresponding to the model potential begins to be apparent. As a consequence, it is possible to find helical fragments that live for several thousand simulation steps, although sooner or later the torsional angles change state and the helix dissolves. Something similar simultaneously happens in the central turn. This turn, however, is of crucial importance in the folding pathway, as can be noticed on observing the region corresponding to 5–13 million steps in figs. 6a and 6b. At the beginning of this time period, there is a high helical population in the residues corresponding to the helical fragments, especially those in helix II (in other simulations beginning with a different conformation or with slightly modified potential parameters, helix I appears first). Since the turn is not formed, the chain adopts a rather extended conformation, without any possibility for the tertiary interactions to stabilize the structure. Nevertheless, once the central turn is in position, the number of tertiary interactions quickly increases, as is observed in the map of native contacts. This effect, together with the temperature decrease, is enough to stabilize the preformed helix II for times considerably longer than those previously seen. The discontinuous lines appearing in the torsional states of helix I indicate that this helix is still not formed. However, the native contacts clearly show that the hairpin conformation is more or less achieved between 7 and 11 million time steps. At about 12 million steps, a fluctuation dissolves the central turn. Immediately, the tertiary interactions disappear, and the helical fragments also partially dissolve. When the turn is again recovered, the tertiary contacts begin to stabilize the structure, starting obviously from those positions closer to the central turn. The different torsional angles acquire their native values, in a process that can be clearly catalogued as on-site construction, at least for one of the helices (the stability of the first helix that is formed varies slightly for the different trajectories computed; in three of them it is partially built simultaneously with the second one from the central turn, while in other six cases it has been previously formed and remains stable while the second helix is being built on it).

In the last steps of the simulation, the chain is in its folded conformation and stays in it, stabilized both by the influence of the secondary and the tertiary interactions. The temperature is low enough so that the native conformation is stable, and the rotational and translational diffusion of the whole folded molecule can be observed. In some cases, "breathing" movements in which one of the helices move slightly apart from another due to small distortions (without changing state) in the torsional angles are observed; this results in the disappearance of the native contact joining the ends of the chain. However, these movements are very fast and do not affect the global structure of the folded state.

An important result of these simulations is that they require about 20–40 million steps to obtain a folded conformation from a pure denatured state. Taking into account the equivalence between the simulation time step Δt and the real time, the total length of the trajectories corresponds to the order of ten microseconds. Even if one considers the purely qualitative character of this estimation, it is remarkable its proximity to the physical time one could expect for a small protein with a very simple structure (since this is the value observed for the folding of helical domains in real proteins), and is consistent with the time scales of isolated helix assembly seen in polypeptides [27].

In nine out of ten of the folding trajectories computed this way, the main features of the dynamic pathways are equivalent to those described in the preceding paragraphs. Local details as to which helix forms first (if any) or the exact sequence of steps appearing in the last stages of the simulation differ slightly from run to run, but the essential scheme of on-site construction of one or even both helices from the central turn (with the first formed helix, when it exists, acting as some kind of scaffold for the growth of the second helix), is always observed.

However, one of the simulations showed a slightly different pattern. The torsional states and native contacts for this trajectory are shown in fig. 7, corresponding again to an annealing process. The most important part of these plots is that seen between 14 and 20 million steps. At the beginning of this period,

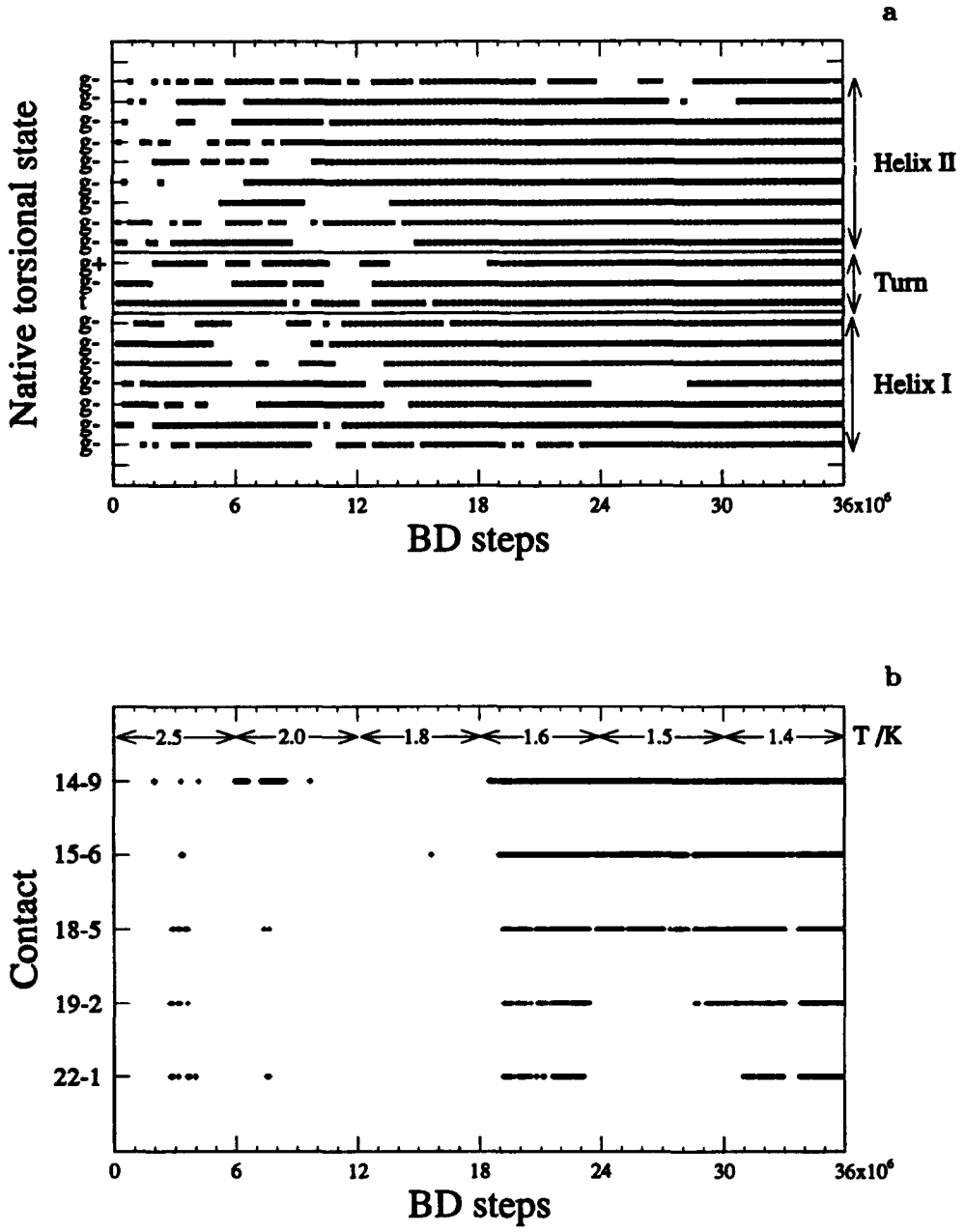


Fig. 7. Results of an annealing BD folding trajectory with g^- favouring torsional potential in the helical regions. (a) Time evolution of the torsional states with the number of simulation steps. (b) Dynamic mapping of the native contacts. See text for details.

both helices are formed, even though the turn is not yet in the correct state, with the logical consequence that tertiary contacts are absent. Then, in a quite fast process, the diffusive approach of the two helices corrects the wrong position of the central turn and the native conformation arises. This can be also observed in the almost simultaneous appearance of the five native contacts in fig. 7b, in comparison with the

gradual growing of the number of native contacts in fig. 6b.

While the formation of this hairpin could be considered as a support for the diffusion–collision model of assembly, the following steps in the dynamics of this very trajectory reveal the existence of a competitive mechanism, i.e., helix dissolution/creation on-site. At about 24 million steps, one can observe that the structure is not yet completely frozen in the folded state, but it is still able to dissolve partially. And the helices do not diffuse apart, but rather they create local defects in certain torsional states that give rise to the existence of a bulge in helix I. At the same time, the end of helix II loses its native conformation, which is simultaneously recovered by both helices, that remain now in the correct folded structure for the rest of the simulation.

The final conclusion we can extract from this set of calculations is that, even though certain processes involve the diffusion of preformed microdomains of secondary structure along the dynamic pathway, on-site construction seems to be the mechanism mainly responsible for the folding process. This result is particularly meaningful when one takes into account that, due to the torsional potential employed, the percentage of native population for the torsional states in the unfolded conformations is quite high (reaching even 40–50% depending on the exact trajectory) for temperatures only slightly larger than the transition temperature. Since the diffusion–collision mechanism of preformed helices appears only marginally under these conditions and the native populations existing in real proteins in the unfolded state are still quite smaller than those employed so far here, we can conclude that the on-site construction appears to be a more probable mechanism for the early stages of protein folding; an identical conclusion was obtained in the lattice MC simulations [4–6,10].

In order to clarify this point even further, let us discuss now the results obtained with inverted preferences associated with the torsional potential in the putative helical regions. The practical way of getting this is to define a potential with two equivalent minima corresponding to the t and g^+ states, while the minimum for the g^- state is energetically higher. Thus, the g^- conformation is still stable, but its stability is considerably reduced with respect to the previous BD simulations, so that the helical contents in the unfolded state are rather lower (about 10–15%), and are a closer approximation to physical reality. The parameters that differ in this series of simulations with respect to the previous ones are included in table 2.

In fig. 8, we show the time evolution of the torsional states and the native contacts for one of the trajectories obtained under the present conditions, in which the temperature was kept constant at the midpoint transition value ($T = 1.2$ K, in this case). The snapshots extracted from the trajectory are included in fig. 9. As a general comment before describing in detail the different steps observed in the trajectory, it is important to compare the differences between figs. 6a and 8a. It is quite evident through this comparison the effect of inverting the propensities for the torsional potential in the putative helical fragments. Now the transitions among the states are substantially faster, and the stability of the helices is much more coupled to the existence of tertiary interactions than it was before, even in spite of the increase of the helical cooperative interactions employed in the simulations (compare the values of ϵ_3 in tables 1 and 2).

In fig. 9, we can see in **1** that the initial conformation of the trajectory is purely random, as it was in the previous BD simulations. At **2** and **3**, it can be seen that some kind of turn has been formed, although from fig. 8a we can appreciate that it is not the native one. As a consequence of this turn, some tertiary interactions occur, the chain acquires a hairpin-like structure, and the number of native contacts increases. However, since the turn is not in the correct conformation, the number of hydrophobic contacts is not enough to stabilize the structure. Thus, the system wanders around this hairpin conformation for several million steps (snapshots **3** to **5**), but finally it dissolves, with the tertiary interactions disappearing (see fig. 8b about 9 million steps). Hence, we can consider that at 11000000 steps, snapshot **6** in fig. 9, the process begins again. Nevertheless, the previous steps clearly indicate that the folding process is not enforced at all in the simulation scheme, but the system has to find its own way across the configurational space driven by a quite physical potential (for the model considered).

The folding now seems to take a different path. In fig. 8b, on coming to 15 million steps one can observe that there is a high helical population in the two

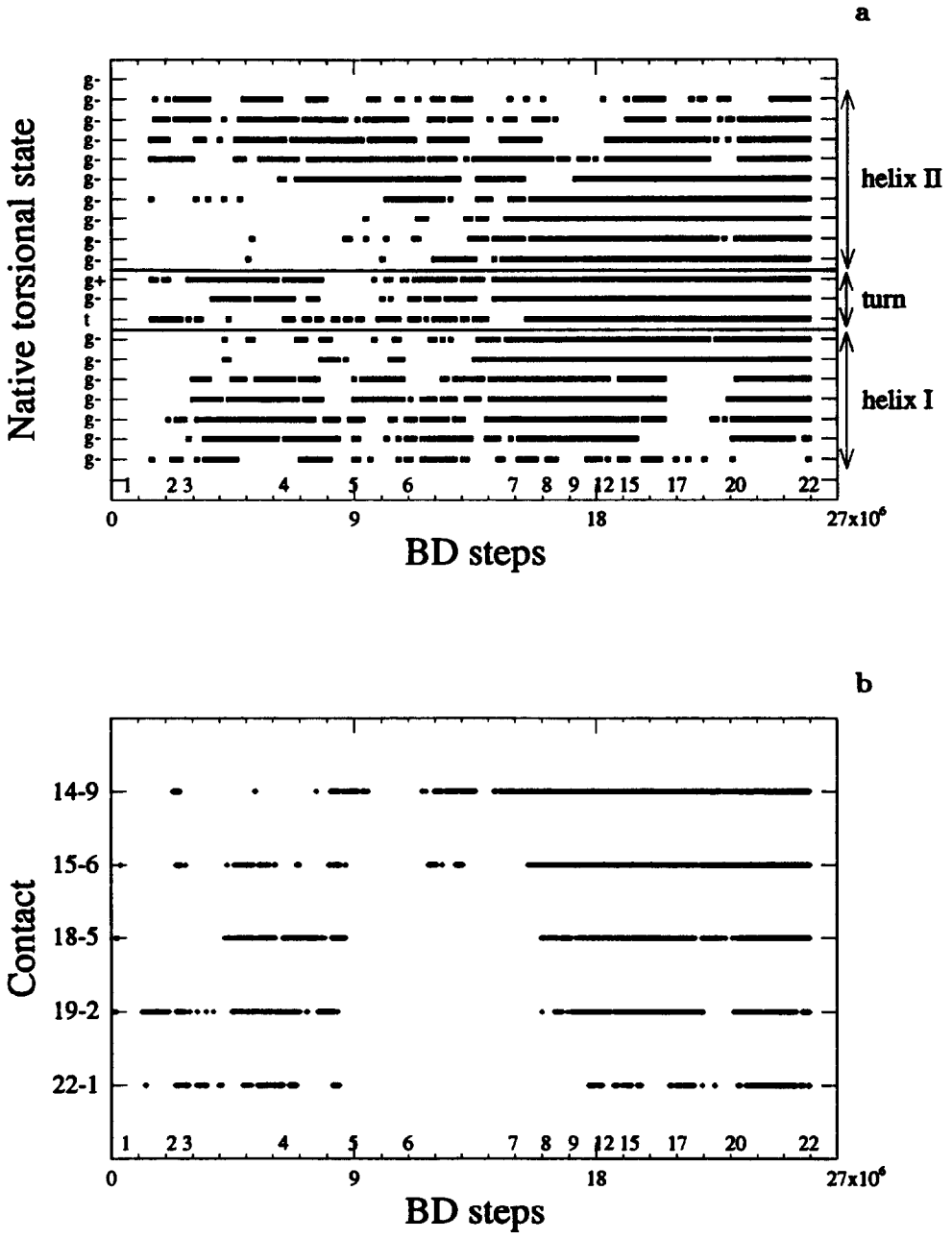


Fig. 8. Results of a constant temperature BD folding trajectory with an inverted torsional potential in the helical regions. (a) Time evolution of the torsional states with the number of simulation steps. (b) Dynamic mapping of the native contacts. The numbers (1–22) over the BD time step axis correspond to snapshots shown in fig. 9.

branches of the hairpin, even though it is opened and lacks native contacts. This is confirmed by 7 in fig. 9. However, the two helices do not diffuse to yield the native conformation. On the contrary, helix II dissolves and begins again to form from the central turn attached to helix I, which remains stable during the

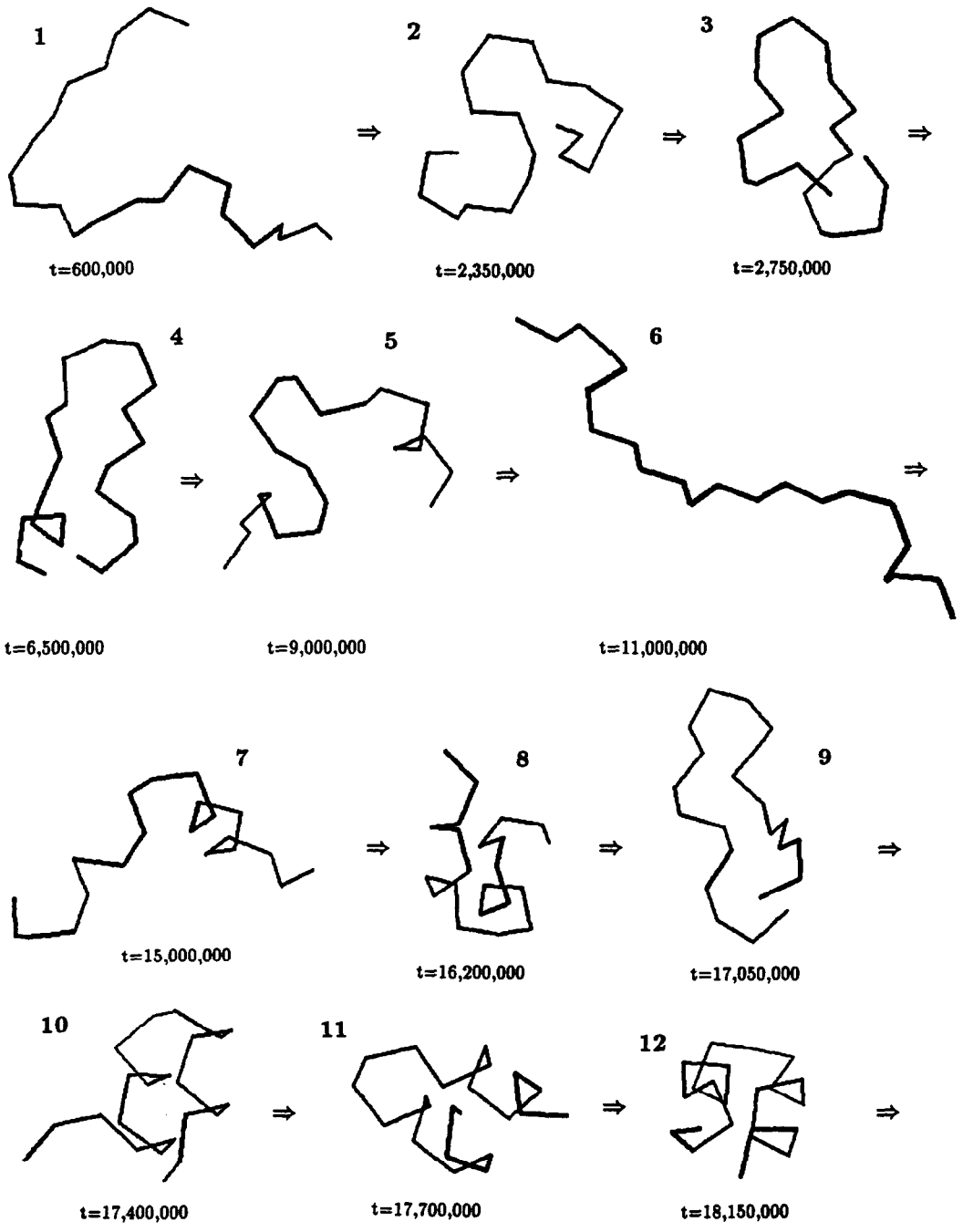
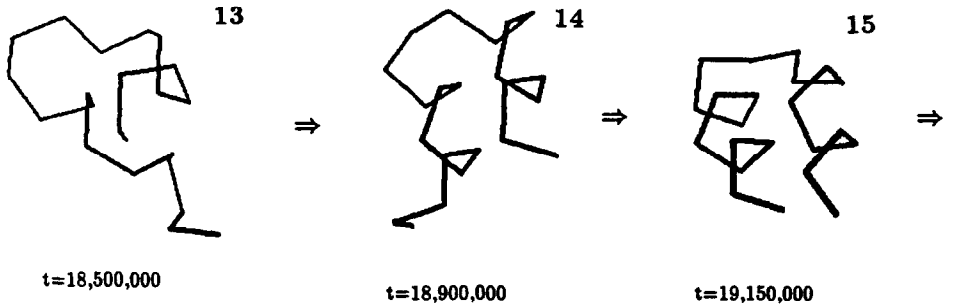
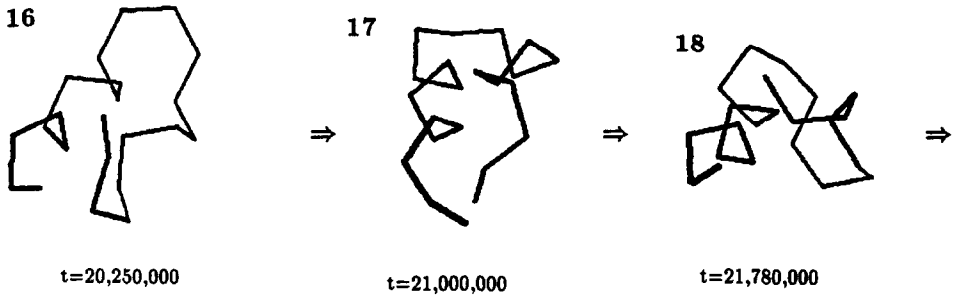
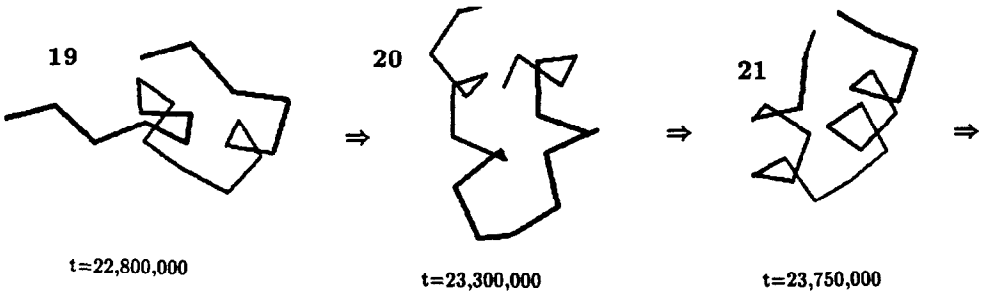


Fig. 9. Snapshots of the BD trajectory corresponding to fig. 8. See text for details.







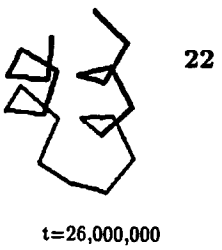


Fig. 9. (continued).

whole process (with some minor modification in the torsional states of the end bonds). Snapshots **8** to **15** show detailed images of the process. Also in figs. 8a and 8b we see how the torsional states of helix II successively take on their native state values, while simultaneously the five native contacts appear one after the other, beginning from the central turn. Hence, once more, the formation of one helix and the central turn seems to be the limiting step in the folding pathway, and once that is accomplished the whole folding happens in a relatively short time (about 2000000 million steps in this trajectory, less than 10% of the total simulation time at constant temperature). It is important to note how similar this pathway is to the on-lattice trajectories shown in figs. 3 and 4.

Of course, since we are at the transition temperature, the folded structure is not completely stable, and in snapshots **16** to **19** of fig. 9 we can see how it can partially dissolve. In this case, it is helix I which changes the torsional states of some of its bonds, and the weakening of the tertiary interactions brings as a consequence a partial dissolving of helix II. Since the turn retains its native conformation (although a substantially longer simulation would show the global unfolding of the model), the refolding process is again quite fast.

In these conditions, five trajectories run at constant temperature showed exactly the same pattern. Only one out of three trajectories corresponding to annealing processes showed something similar to the preformed diffusion–collision mechanism. However, since now the helices are not completely stable in the absence of tertiary interactions (something that we have checked running a trajectory for a single helix at the transition temperature), the time required to fold at the transition temperature is longer than in the simulations which favour native torsional states in the helical fragments, so that the observed pathway in these annealing conditions could be affected by a too fast cooling rate.

4. Summary and conclusions

In this work we have studied the folded pathway of α-helical hairpins through two rather different methods. DMC simulations of a detailed model in the 210 lattice show that the pathways previously obtained in diamond lattice models are independent of the geometry and coordination of the lattice, and of the set of movements used to simulate the dynamic procedure. Furthermore, BD off-lattice simulations show quite conclusively that those results are also independent of the physical scheme underlying the simulation technique. Since the results of the BD calculations, obtained through the solution of a fundamental real equation of motion, are fully consistent with the DMC folding pathways, the latter are clearly validated. This is a very important result if we take into account the different computational efficiency of the two types of algorithms. The calculation of a BD folding trajectory takes between 10–25 CPU hours (depending on whether only a temperature or a whole annealing process is considered) on a Cray Y-MP (about 15 times more on a Sun Sparc 2 workstation). Nevertheless, the DMC simulations of these simple models take a time of the order of minutes in the Sparc 2 workstations. If we consider that any real protein is much more complicated than the model employed in the diamond lattice (whose continuous version has also been employed in the BD calculations), it is clear that BD is not a suitable method for studying the whole folding mechanism of a real protein, unless one introduces a series of simplifications (e.g., prebuilt helices) that, as we have shown, seriously affect the resulting dynamics. On the other hand, we have proven with this work that the introduction of the lattice and the partially artificial movements included in DMC simulations do not have any significant effect on the pathways. Moreover, if one relates the approximate folding times on the 210 lattice and BD models, an elemental time step on the 210 lattice roughly corresponds to 1–10 nanoseconds, values consistent with previous conjectures [10]. Hence, we are now much more confident in the possibility of using DMC algorithms to reproduce the folding pathways of real proteins [22], with a certain security that the observed results can appropriately reproduce the physical features of the folding process.

Acknowledgement

This work was supported in part by NIH grant GM37408 and a grant from the Polymer Program of the National Science Foundation. AR also acknowl-

edges a M.E.C./Fulbright Scholarship from the US-Spanish Joint Committee for Cultural and Educational Cooperation. The Cray CPU time was kindly provided by Cray Research, Inc.

References

[1] J.A. McCammon and S.C. Harvey, Dynamics of Proteins and Nucleic Acids (Cambridge Univ. Press, Cambridge, 1987).
[2] K.D. Gibson and H.A. Scheraga, J. Comput. Chem. 11 (1990) 468.
[3] M. Karplus and G.A. Petsko, Nature 347 (1990) 631.
[4] J. Skolnick and A. Kolinski, Ann. Rev. Phys. Chem. 40 (1989) 207.
[5] A. Sikorski and J. Skolnick, J. Mol. Biol. 212 (1990) 819.
[6] J. Skolnick and A. Kolinski, J. Mol. Biol. 212 (1990) 787.
[7] N.G. van Kampen, Stochastic Processes in Physics and Chemistry (North-Holland, Amsterdam, 1981).
[8] S. Lee and M. Karplus, Biopolymers 26 (1987) 481.
[9] J.D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. USA 87 (1990) 3526.
[10] J. Skolnick and A. Kolinski, J. Mol. Biol. (1991), in press.
[11] T.L. Hill, Statistical Mechanics (McGraw-Hill, New York, 1956).
[12] A. Kolinski, M. Milik and J. Skolnick, J. Chem. Phys. 94 (1991) 3978.
[13] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. Chem. Phys. 21 (1953) 1087.
[14] D.L. Ermak and J.A. McCammon, J. Chem. Phys. 69 (1978) 1352.
[15] H. Yamakawa, Modern Theory of Polymer Solutions (Harper & Row, New York, 1971).
[16] M. Fixman, Macromolecules 19 (1986) 1195.
[17] S.A. Allison and J.A. McCammon, Biopolymers 23 (1984) 363;
A. Rey, J.J. Freire and J. Garcia de la Torre, J. Chem. Phys. 90 (1989) 2035.
[18] S. Lee and M. Karplus, J. Chem. Phys. 81 (1984) 6106;
R.J. Lewis, S.A. Allison, D. Eden and R. Pecora, J. Chem. Phys. 89 (1988) 2490.
[19] M.P. Allen and D.J. Tildesley, Computer Simulation in Liquids (Oxford Univ. Press, New York, 1987).
[20] P.J. Flory, Statistical Mechanics of Chain Molecules (Wiley, New York, 1969).
[21] A. Chung-Phillips, J. Chem. Phys. 88 (1988) 1764.
[22] J. Skolnick and A. Kolinski, Science 250 (1990) 1121.
[23] H.J. Dyson, M. Rance, R.A. Houghten, P.E. Wright and R.A. Lerner, J. Mol. Biol. 201 (1988) 201.
[24] D. Poland and H.A. Scheraga, Theory of Helix-Coil Transitions in Biopolymers (Academic Press, New York, 1970).
[25] W.F. van Gunsteren and M. Karplus, Macromolecules 15 (1982) 1528.
[26] J. Skolnick and E. Helfand, J. Chem. Phys. 72 (1980) 5489.
[27] R. Zana, Biopolymers 14 (1975) 2425.