Designing refoldable model molecules

I. Coluzza, H. G. Muller, and D. Frenkel
FOM Institute for Atomic and Molecular Physics, Kruislaan, 407 1098 SJ Amsterdam, The Netherlands
(Received 16 April 2003; published 13 October 2003)

We report a numerical study of the design of lattice heteropolymers that can refold when the properties of only a few monomers are changed. If we assume that the effect of an external agent on a heteropolymer is to alter the interactions between its constituent monomers, our simulations provide a description of a simple allosteric transition. We characterize the free energy surfaces of the initial and the modified chain molecule. We find that there is a region of conformation space where molecules can be made to refold with minimal free energy cost. This region is accessible by thermal fluctuations. The efficiency of a motor based on such an allosteric transition would be enhanced by ''borrowing'' heat from the environment in the initial stages of the refolding, and ''paying back'' later. In fact, the power cycle of many real molecular motors does involve such a borrowing activation step.

DOI: 10.1103/PhysRevE.68.046703 PACS number(s): 07.05.Tp

I. INTRODUCTION

Molecular motors are molecules that can convert chemical energy into mechanical energy. The effect of the chemical reaction is to induce a conformational change in the molecule. As the final conformation has a lower free energy than the initial one, the molecule has the capacity to perform an amount of work that is, at most, equal to this free energy change. The amount of work that is delivered in practice depends on many factors, such as the speed of the transformation and the mechanical coupling of the relevant ''reaction coordinate'' to the outside world.

In molecular motors, as well as many other proteins, the transition between conformations is induced by a change in the environment (e.g., a pH change), the absorption of a photon or the chemical transformation of a fuel molecule (e.g., ATP or lactose) attached to the protein. The effect of this external element is to change the structure or interactions of some ''active'' parts of the protein. These changes, in turn, lead to a rearrangement of the protein structure. The fact that molecular motors are microscopic has important consequences for their mode of operation. In fact, the second law of thermodynamics makes it impossible for a macroscopic Carnot engine to ''borrow'' significant amounts of heat from the environment. In contrast, thermal fluctuations play an important role in the behavior of molecular motors.

Changes in conformation due to altered interactions between monomers are also of interest in a different context, namely, in the design of mutations that significantly modify the native structure of a protein. In 1993, Rose and Creamer [1] formulated this problem as follows: given two distinct protein folds of similar length, what is the minimum number of amino acids that must be changed in order to transform one fold into the other? In fact, Rose and Craemer formulated the so-called ''Paracelsus challenge'': the award of a prize to anyone who could convert one protein fold into another without changing more than 50% of the original protein's sequence. A possible solution to this challenge was proposed by Dalal *et al.* [2] who designed a protein sequence that could be converted from its native β-sheet conformation into an α-helix structure by changing fewer than half of the amino acids.

At the level of the relative stability of native structures, the present study of allosteric transitions is equivalent to the problem of conformational changes due to mutations. The difference appears when we consider the actual pathway by which the molecule refolds after the change has been introduced: this pathway has little physical meaning in the case of mutation, but is of considerable interest regarding allosteric transitions. In this paper, we explore a simple model for allosteric transitions that is intermediate between a realistic, but prohibitively costly, atomistic model and a simple, but abstract, two-state model. We model the chain as a linear, polypeptidelike heteropolymer, living on a lattice. In what follows, we shall refer to this molecule as a ''protein'' and, in fact, we shall use model parameters that apply to proteins. We stress, however, that the approach is not limited to protein-based conformation switches. Our central assumption concerns the effect of a chemical reaction on the chain molecule. We assume that the chemical reaction does not directly lead to a conformational change of the molecule. Rather, we assume that it leads to a change in the effective interaction between some of the monomers in the chain.

In our model, we account for this difference in the properties of individual monomers by changing their chemical nature. This modification could be thought of, for example, as a change of the ionization state of acidic or basic residues triggered by a pH change or as resulting from binding to a metal ion. For simplicity, however, rather than introducing a new set of interactions between specific residues, we exchange them for other members of the standard set of 20 amino acids. In the language of proteins, we change the identity of one or more amino acids in the chain. Once an amino acid has been modified, the molecule may be able to lower its free energy by transforming into a different native state. In the present model, it is this thermodynamic incentive that drives the refolding to a new spatial arrangement.

Below, we first describe the techniques used to simulate our system, we then present the simulations of the refolding process, and finally we discuss some of the implications of this work.

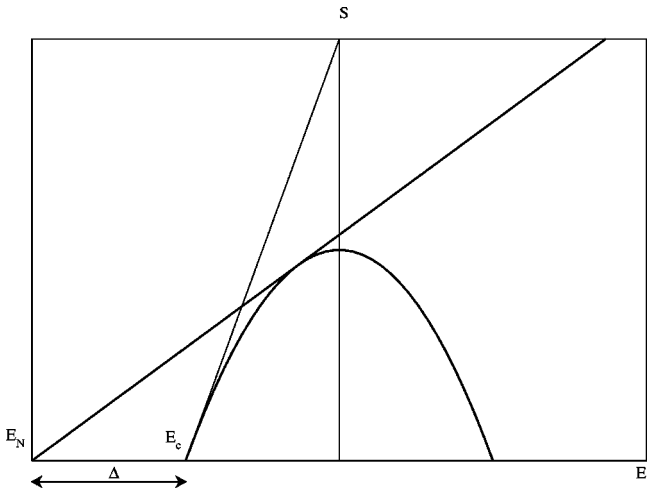

FIG. 1. (Color online) Energy spectrum of a heteropolymer on a lattice given by Eq. (3). On the horizontal axis is the configurational energy [Eq. (1)], while on the vertical axis is the entropy. E_N is the energy of the native state, while E_c is the crossing point of the parabola with the abscissa. Δ is the region of discrete states. The slope of the tangent passing through E_N defines the folding temperature, while the tangent in E_c gives the glass temperature.

II. METHODS

In what follows, we use a lattice model to study active refolding in proteins. This model assumes that there are only nearest-neighbor interactions between the amino acids. The total conformational energy of a particular sequence in a given structure is given by

$$E=\sum C_{ij}S_{ij}, \tag{1}$$

where i and j are particle indices, C is the contact map defined as

$$C=\begin{cases}1 & \text{if } i \text{ neighbor of } j\\ 0 & \text{otherwise,}\end{cases} \tag{2}$$

and S is the interaction matrix. For S we use the 20×20 matrix fitted by Miyazawa and Jernigan [3] on the basis of the frequency of contacts between each pair of amino acids in nature.

A. General properties of heteropolymers on lattice

A given lattice polymer can form a large number of compact conformations. Obviously, every conformation is characterized by a different contact map. Hence, the energy of the polymer depends on its conformation. In a mean field approximation the energy spectrum of the compact structures of a random chain on a lattice has the shape shown in Fig. 1.

The mean filed expression for the entropy is [4–7]

$$S(E)=\begin{cases}N\ln\gamma-\dfrac{E^2}{2N\sigma_{\mathrm{B}}^2} & \text{if } E>E_c\\ 0 & \text{if } E\leqslant E_c,\end{cases} \tag{3}$$

where N is the number of elements in the chain, σ_B is the standard deviation of the interaction matrix, and γ is the coordination number for fully compact structures on the lattice. E_c is the (lower) crossing point of the parabola with the abscissa, $E_c=-N\sigma_{\mathrm{B}}(2\ln\gamma)^{1/2}$. The finite width of this energy spectrum reflects the fact that the system is frustrated. The native state corresponds to the least frustrated structure. If the native state is nondegenerate, this lowest-energy conformation has zero entropy. The degree of frustration of a heteropolymer is linked to the number of different monomers that it contains. This is particularly obvious in the case of a homopolymer. For such molecules, all compact states are unfrustrated and have the same energy. This picture is confirmed in our simulation, where indeed we observe a nondegenerate native state for a well designed sequence. In the following we will refer to the lowest-energy state as the native state of the heteropolymer.

In 1993 Shaknovich and Gutin [8–10] showed that it is possible to ''design'' a lattice protein in such a way that it will fold into a specific conformation. They achieved this by optimizing the sequence of amino acids, using a Monte Carlo algorithm that randomly exchanges amino acids within the chain molecule. The acceptance of such trial swaps depends on the energy change associated with the move

$$\Delta E=\sum (S'_{ij}-S_{ij})C_{ij}, \tag{4}$$

where S' (S) denotes the interaction matrix of the new (old) sequence of amino acids. During a Monte Carlo run of several hundred thousand cycles, a large number of distinct sequences are generated. The sequence S^* with the lowest energy is assumed to be the best candidate to fold into the native state.

$$E_{\mathrm{Native}}=\sum C_{ij}S^*_{ij}. \tag{5}$$

Pande *et al.* [11–13] have provided a theoretical analysis of the design of a foldable protein sequence. In particular, these authors showed that, in the context of the random energy model, the phase behavior of designed protein sequences can be predicted analytically. One of the main findings of this work is that the energy gap separating the target native state from the set of non-native compact states is inversely proportional to the design temperature (the fictitious temperature at which we perform Boltzmann sampling of different sequences for a given target conformation).

In the following section, we extend this point of view to the design of a molecule that can undergo an allosteric transition. Our model assumes that an external agent changes the effective interactions of a few monomers in the chain. Its effect on the native conformation of a chain is therefore identical to that of a small number of mutations. Our aim is then to design two sequences that differ only by a small number of ''mutations, yet fold into different target conformations.

A closely related but different problem is the design of heteropolymers that fold into a specific structure. This problem has been addressed using the so-called ''painted glob-

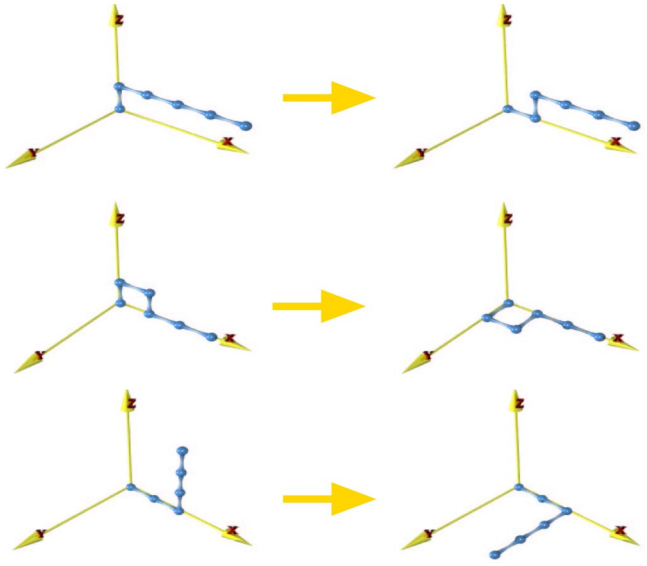

FIG. 2. (Color online) Monte Carlo moves used in the folding program. From top to bottom: corner flip, crankshaft, rotation.

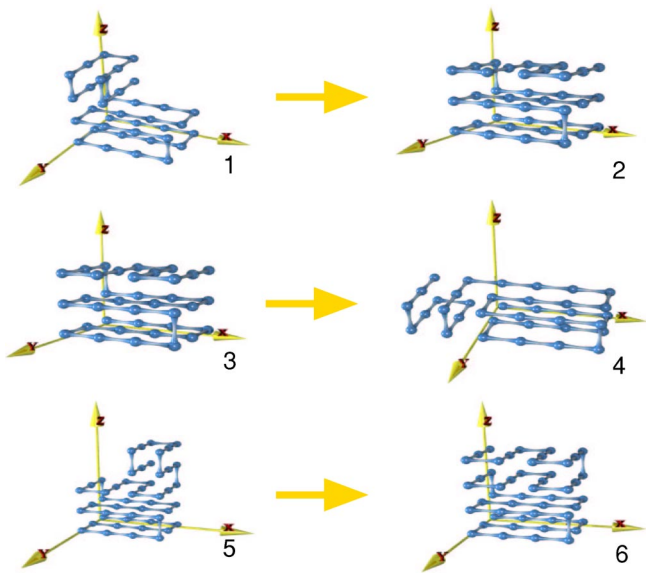

FIG. 3. (Color online) Spatial arrangement of the chain in the structures used to test the model.

ule'' model (see, e.g., Refs. [14,15]). The central idea behind this approach is to look at the target structure, and then distinguish between surface and core residues (hydrophilic and hydrophobic). The design consists of a sequence of folding and repainting steps.

B. Design of a switchable polymer

In order to design two monomer sequences that yield distinct target conformations, yet differ only in a few monomers, we used a modified version of the Shakhnovich method. Unlike the latter method, our approach does not keep the amino acid composition of a chain fixed. Rather, we allow for random changes of amino acids. In order to prevent this compositional sampling from leading to the formation of homopolymers, we introduce a (purely fictitious) compositional ''temperature.'' This has the result of increasing the compositional entropy. To perform the sampling, we combine the following acceptance criterion with the normal acceptance Metropolis rule

$$P_{\mathrm{acc}}=\min\left\{1,\left(\frac{N_P^{\mathrm{new}}}{N_P^{\mathrm{old}}}\right)^{\beta_2}\right\},$$

where β_2 is the arbitrary parameter that plays the role of a temperature and N_P is the number of permutations that are possible for a given set of amino acids. N_P is given by the multinomial expression

$$N_p=\frac{N!}{n_1!n_2!n_3!\ldots}, \tag{6}$$

where N is the total number of monomers and n_1,n_2, etc., are the number of amino acids of type 1,2, ... With this condition we can explore a large set of sequences, yet avoid the formation of homopolymers. In the absence of any *a priori* criterion to fix β_2, we used trial and error. If β_2 is too low (high temperature), the chains will tend to become homopolymers (with a degenerate native state). In contrast, when β_2 is too high (low temperature), we noticed that the lowest-energy sequences were no longer able to fold. We should therefore choose a value of β_2 that yields a compromise between these two conflicting tendencies. To explore a range of values for β_2 and at the same time limit the trapping in local minima of sequence space, we introduced a parallel-tempering algorithm for the sequence sampling at different pseudotemperatures. The definition of the acceptance criterion for the parallel-tempering trial moves is

$$P_{\mathrm{acc}}=\min\left\{1,\left(\frac{N_P^{\mathrm{new}}}{N_P^{\mathrm{old}}}\right)^{\Delta\beta_2}\right\}.$$

We found that, for the present chain lengths, the set of values $\beta_2=\{1,2,\ldots 14\}$ yielded good sequences, in the sense that the native state that was both stable and nondegenerate. We stress that the fictitious temperature parameter β_2 only plays a role in the generation of suitable sequences and does not model the subsequent chain refolding. In our sequence-generating algorithm, we impose an upper limit on the number of differences between the two sequences that we design. Of course, once we have generated two particular sequences, we still need to test whether they do indeed fold into the desired structures.

C. Folding

To study the folding of a particular model protein, we use a Monte Carlo simulation with three basic moves (Fig. 2): corner flip, crankshaft, branch rotation. The corner flip [Fig. 2(a)] involves a rotation of 180 deg of a given particle about the line joining its neighbors along the chain. The crankshaft move [Fig. 2(b)] is a rotation by 90 deg of two consecutive particles. A branch rotation is a turn, around a randomly chosen pivot particle, of the whole section starting from the

TABLE I. Sequences generated for the test structures (Fig. 3). The letters in bold are the amino acids chosen by the design program to induce the conformational change.

*WKCAVCEMNRCILCDTWKCFICEMERDGQKYPS**R**Q**K***	Sequence *A*
*WKCAVCEMNRCILCDTWKCFICEMERDGQKYPS**I**Q**M***	Sequence *B*
*WKCAVCEMNRCILCDDWKCFGCEMPRKN**PM**YTS**E**Q**H***	Sequence *C*
*WKCAVCEMNRCILCDDWKCFGCEMPRKN**EH**YTS**I**Q**P***	Sequence *D*
*HWKLHDMYVWRTKDMLPWREVDMYAQIPPITENSKAFESCRGFQC**L**N**G***	Sequence *E*
*HWKLHDMYVWRTKDMLPWREVDMYAQIPPITENSKAFESCRGFQC**N**K**G***	Sequence *F*

pivot particle and going to the end of the chain. With these moves we expect to have a good balance between cluster moves and single-particle moves.

During the simulation we measured two order parameters. The first is the conformational energy [Eq. (1)] of the chain. The second is the number of native contacts in a given conformation, which is a commonly used order parameter in the study of protein folding. However, as we are considering a model with two native structures, we define the order parameter as the difference of the number of contacts that belong to two reference structures (e.g., 1 and 2) i.e.,

$$Q(C)=\sum_{i<j}^{N}[C_{ij}^{(1)}C_{ij}-C_{ij}^{(2)}C_{ij}], \tag{7}$$

where $C_{ij}^{(1)}$ and $C_{ij}^{(2)}$ are the contact maps of the reference structures and C_{ij} is the contact map of the instantaneous configuration. To be more precise, as we consider two distinct native states, we take these as the reference structures, giving a value $+1$ to every contact that belongs to structure 1 and a value -1 to every native contact of structure 2. Contacts that appear in both 1 and 2 do not contribute to this order parameter.

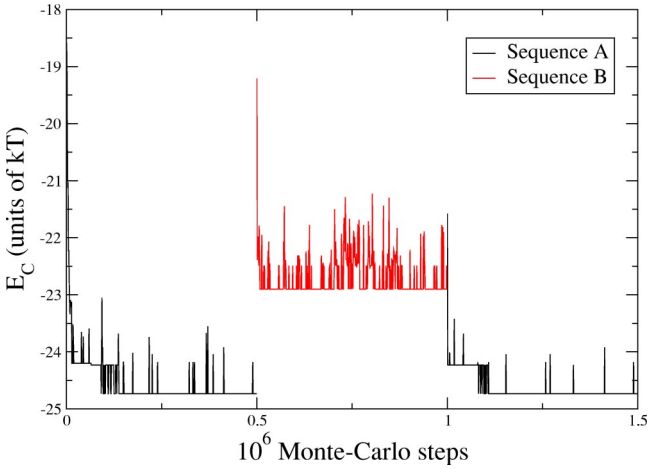

FIG. 4. (Color online) Sequence switching simulation. On the vertical axis we plot the conformational energy E_C [Eq. (1)] while the abscissa is the number of Monte Carlo steps. During the simulation we switch from sequence *A* to sequence *B*, and look at the conformation corresponding to the lower energy. For the sequence *A* the native structure is 1 [Fig. 3(a)], while for sequence *B* the structure is 2 [Fig. 3(b)]. The process is reversible; in fact, when, after 10^7 steps we switch back to sequence *A*, the lowest-energy conformation is structure 1.

The Landau free energy, as a function of the order parameter Q [Eq. (7)], is defined by

$$F(Q)=-kT\ln[P(Q)], \tag{8}$$

where $F(Q)$ is the Landau free energy of the state with order parameter Q and $P(Q)$ is the histogram that measures the frequency of occurrence of conformations with order parameter Q (see, e.g., Ref. [16]). In fact, a direct (brute-force) calculation of this histogram is not very efficient, as the system is often trapped in local minima, especially at low temperatures. To solve this problem, we combined our sampling of chain conformations with a parallel-tempering routine [17–20]. Using this approach (with 14 different temperature stages $1,\frac{1}{2},\ldots,\frac{1}{14}$) we can get efficient sampling of the accessible free energy landscape for the individual sequences. However we could not adequately sample the free energy for all the possible value of Q. We then combined the normal parallel tempering with an umbrella sampling of the polymer free energy landscape ([21,22]). In such a simulation we bias the sampling with respect to the order parameter so that all relevant conformations occur approximately equally. More specifically, we bias the sampling of a particular value of the order parameter by imposing a bias potential that is opposite and (approximately) equal to the free energy associated with that order parameter. As this free energy is not known *a priori*, the biasing potential is constructed itera-

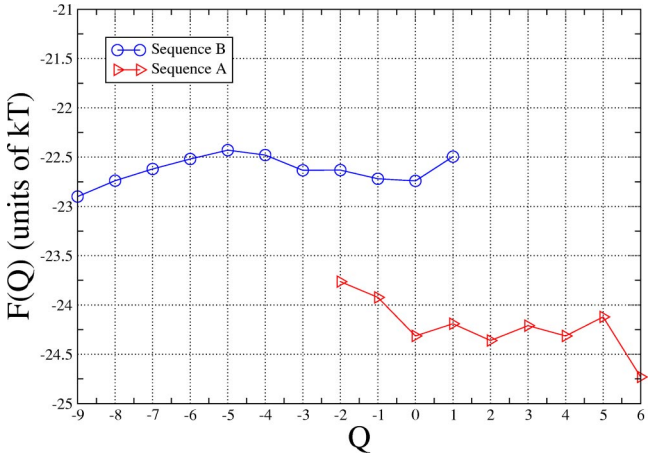

FIG. 5. (Color online) Result of the folding simulation for the sequences *A* and *B*, obtained using a parallel-tempering algorithm. On the horizontal axis is the number of native contacts Q defined in Eq. (7). On the vertical axis is the free energy $F(Q)$. This plot shows the need for improved sampling.

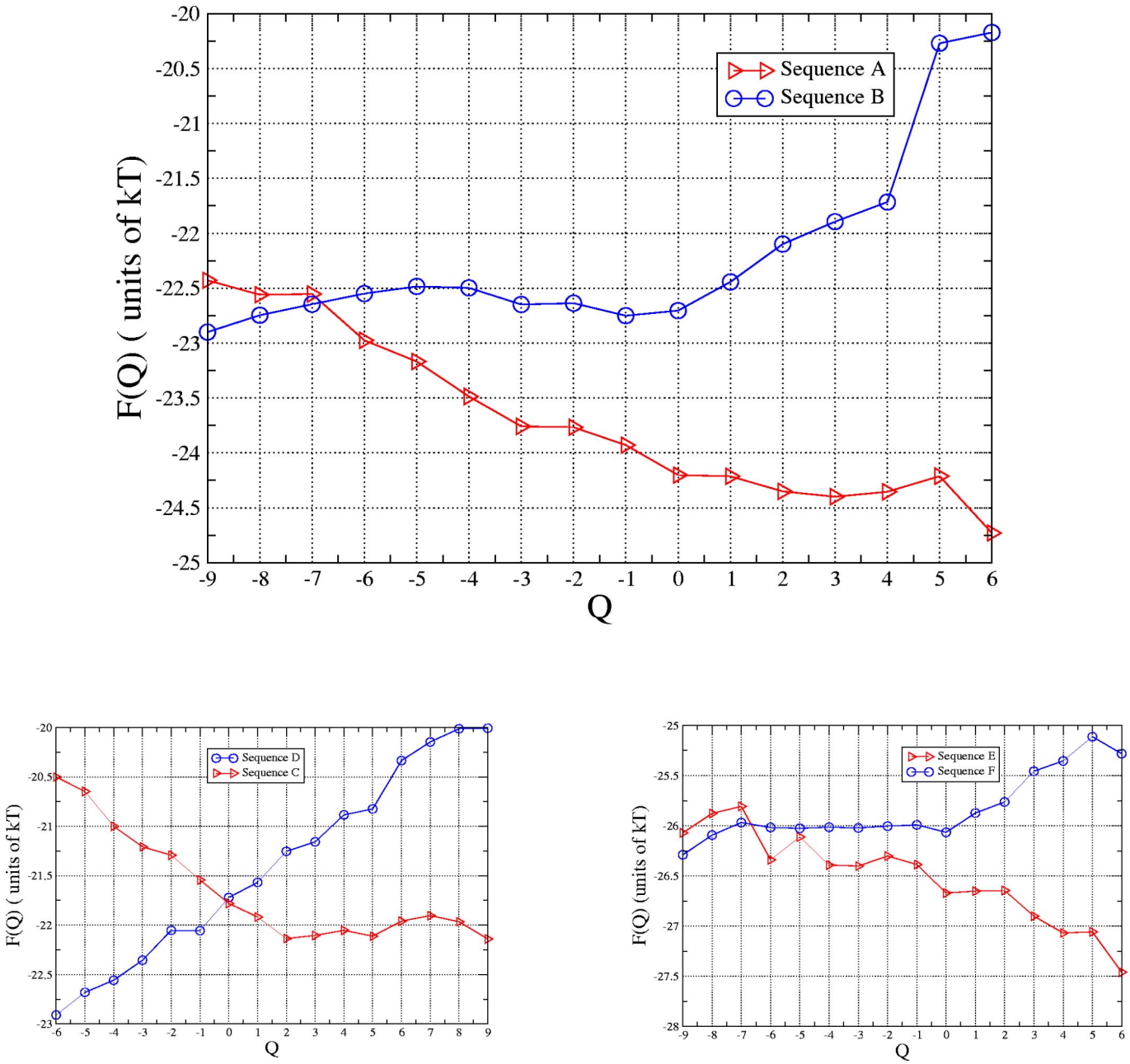

FIG. 6. (Color online) Plots of the free energy $F(Q)$ of all the sequence pairs ($A \Leftrightarrow B$, $C \Leftrightarrow D$, and $E \Leftrightarrow F$) as a function of the number of native contacts Q [Eq. (7)]. These data were obtained with a joint parallel-tempering and umbrella-sampling simulation. In this plot is visible the crossing point between the free energy curves where the energetic cost of the sequence switching is lower.

tively. In the Appendix we describe our implementation of the umbrella-sampling scheme.

III. RESULTS

To illustrate the mechanism by which allosteric transitions proceed in our model, we consider the refolding behavior of three different model molecules. In Fig. 3 we show the target structures between which the transitions occur: $1 \Leftrightarrow 2$, $3 \Leftrightarrow 4$, and $5 \Leftrightarrow 6$. Because the same procedure is applied in every case, we focus our explanation on the conformational change from structure 1 [Fig. 3(a)] to structure 2 [Fig. 3(b)]. Following the procedure explained in Sec. II we first designed a sequence that would fold into structure 1 [see Fig. 3(a)]. We explore possible amino acid sequences by using both the conventional swapping move that does not change the composition and the switch move that does. The acceptance criterion of the latter trial move depends on the parameter β_2 that has to be chosen. A typical result after 10^7 iterations is sequence A (Table I). We applied the same technique to the other initial structures 3 and 5 [Figs. 3(c) and 3(e)], and the resulting sequences are, respectively, sequence C and sequence E (Table I).

Sequences A, C, and E, listed in Table I, were used as the

starting point to design the modified sequences that would refold into structures 2, 4, and 6, respectively [see Figs. 3(b), 3(d), and 3(f)]. We limited our search to those sequences that differed by a given number of residues. For the first and the last example we constrained the sequence that formed the initial conformation to differ by, at most, two amino acids from the sequence that formed the final conformation. For the transition $3 \Leftrightarrow 4$, we imposed a threshold of four residue differences. These Paracelsus numbers are purely empirical: they are the lowest threshold for which refolding to the desired structures could be obtained in each case.

The sequences that are listed in Table I are those ones used in the simulations described below. We stress that we did not impose the positions of the ''mutations.'' But, not surprisingly, they appear to be concentrated in that part of the chain that is involved in the conformational change. Having constructed the two desired sequences, we performed a Monte Carlo (MC) simulation to study the equilibrium properties of the native state of each sequence.

Free energy calculations

First, we checked if changing from sequence *A* to *B* (see Table I) did, indeed, induce the desired conformational change. To this end, we started with a random coil of a molecule with sequence *B*. We used a standard MC simulation to let this structure fold. After the chain had reached its native structure (1), we changed the sequence from *A* to *B* and continued the simulation. After sequence *B* reached its native state (2), we switched back to sequence *A*, to verify that the refolding works both ways. In Fig. 4 we plot the conformational energy of the chain [Eqs. (1)] as a function of the number of MC steps, highlighting the time windows corresponding to each sequence. In each window, we see that immediately after the sequence switch, the system is in a state of very high potential energy, but then the chain quickly relaxes into its new native state. This shows that it is indeed possible to induce a conformational change with a relatively small modification of the chemical nature of some amino acids along the chain. The same procedure was also applied to the other sets of sequences. The results thus obtained were qualitatively similar to that we obtained for the *A*-*B* pair.

Subsequently we studied the free energy profile using the parallel tempering described in Sec. II. The results of these simulations are shown in Fig. 5 and exhibit the characteristic behavior expected for a molecule that can undergo a folding transition. At low temperatures, the native state has a free energy that is lower than that of the molten globule (characterized by an order parameter close to zero). However, to study how one structure refolds into the other, we need to know the free energy landscape of, for example, sequence *A* in the vicinity of the native structure of *B*, and vice versa.

In order to improve the sampling of conformations that would hardly be sampled in a brute-force simulation, we proceed according to the method explained in Sec. II and we use the biasing technique described in the Appendix. The result of the simulation is the complete spectrum of the free energy for each of the two sequences. In Fig. 6 we plot the

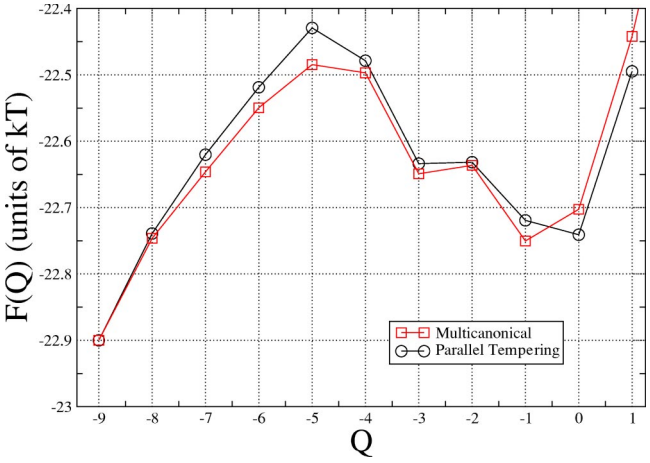

FIG. 7. (Color online) Comparison between the free energy $F(Q)$ calculated with an umbrella-sampling simulation and a parallel-tempering one, in a window of the order parameter Q [Eq. (7)] were the two methods should give the same results. The agreement between the two methods provides confidence in the validity of our results.

free energy of sequences *A* and *B*. As a consistency test, we compare in Fig. 7 the easily accessible part of the free energy of sequence *B* calculated with and without the umbrella-sampling scheme. Now that we know the free energy curves for sequences *A* and *B*, we can study the effect of a sequence change. The crossing point of the curves (Fig. 6) corresponds to a value of the order parameter for which we can change the sequence from *A* to *B* without changing the free energy.

While the order parameter that we have used thus far allows us to discriminate between states that are close to one native state or the other, it is less convenient to probe the intermediate region. The reason is that there are many different conformations with an order parameter close to zero. Not all of these conformations are equally important for the refolding process. We therefore need a second-order parameter that allows us to get more detailed information about the free energy landscape in the intermediate state. We found that the conformational energy of the chain was suitable as a second-order parameter.

Figure 8 shows the free energy landscapes for the folding of sequences *A* and *B*. Interestingly, the two surfaces show an overlap close to the crossing point of the curves in Fig. 6. This means that in the region of overlap, it is possible to change the amino acid sequence without changing the conformational energy of the chain. This suggests that in those conformations of the chain the mutated amino acids are not in contact with the rest of the chain. The possibility of changing the sequence without affecting the potential energy of the chain facilitates the action of an external agent. Similar behavior has been postulated for real protein motors that undergo a progressive change in the conformation. For instance, in the hand-over-hand model of kinesin by Schief and Howard [23], the external agent only acts if the protein is ready to accept it. Such a behavior could easily be described by an extension of the present model where we only allow sequence switching when thermal fluctuations bring the chain into a favorable conformation. As can be seen from

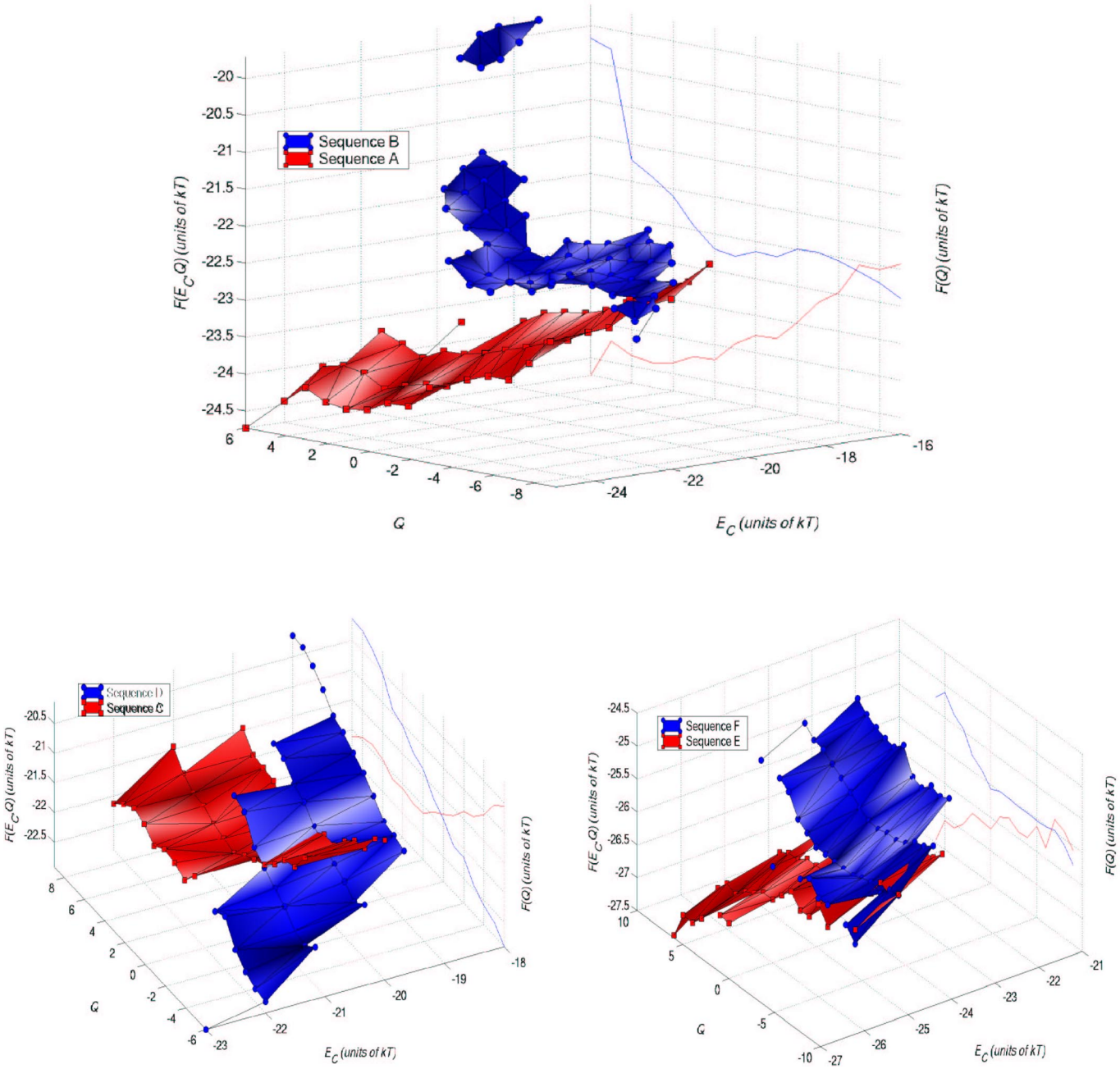

FIG. 8. (Color online) Plot of the free energy landscape of the folding of the sequences pairs $A \Leftrightarrow B$, $C \Leftrightarrow D$, and $E \Leftrightarrow F$. The free energies $F(E_C,Q)$ are functions of the conformational energy E_C [Eq. (1)] and of the number of native contacts Q [Eq. (7)]. The interesting feature of the plot is the overlapping area. The configurations in this area are common to both sequences. Comparing these plots with the corresponding ones in Fig. 6 (for convenience, we have replot it on the z-y plane) we see that the states have the same conformational energy. From the parallel tempering simulation we know that those configurations can be reached just by thermal fluctuations.

Fig. 5, both proteins can reach conformations with an order parameter between -2 and 1 by spontaneous thermal fluctuations of the order kT.

Clearly, the present model only deals with a single, albeit essential, aspect of a molecular motor, namely, the property of a motor head to undergo an allosteric transition. For one thing, we do not consider explicitly the reaction between the external agent (e.g., ATP) and the chain molecule: we simply assume that the effect is to expose some other amino acids in the molecule. It should be possible to construct a model where these changes follow in a natural way from the chemical natures of both the chain molecule and the external agent. Then it would be interesting to see how the work that can be performed by the molecule (i.e., the difference in free energy between the initial and final states in the refolding process) depends on the free energy change associated with the chemical reaction with the external agent.

IV. DISCUSSION

Recently, Borovinskiy and Grosberg [26] reported a numerical study that focused on another aspect of refolding in

design not just the initial and final states of the model protein, but also to design the sequence such that every single step of the conformational change would release approximately equal amounts of free energy. This imposed property was based on the idea that a protein stores free energy like a spring. Our model differs from this approach because we do not constrain the path by which the conformational changes proceed. In particular, by not imposing how the free energy is stored in the molecule, we find a barrier between the two states, the height of which depends strongly on the starting configuration of the chain. The refolding process is assisted by ''uphill'' thermal fluctuations that put the system in a favorable initial condition and effectively reduce the amount of chemical work that would be wasted in the refolding process. Evidence for the relevance of thermal fluctuations in initiating refolding comes from experimental studies on motor proteins [23] and is captured at a phenomenological level by thermal ratchet models [24]. Surprisingly, we find that, even without designing the pathway for refolding, our model spontaneously reproduces the ''springlike'' gradual release of free energy during refolding.

To summarize, we have introduced a simple model to describe the behavior of a protein undergoing an allosteric transition. The protein is approximated by a linear heteropolymer on a lattice. The role of the external signal is played by an effective change of the amino acids along the chain. With this model we want to demonstrate that by destabilizing some essential elements of a conformation we can induce the chain to refold into a different structure. We can control this process by using a sequence design algorithm. With a folding program we characterized the equilibrium properties of the chain before and after the signal. Using the order parameters derived from the number of native contacts and the conformational energy, we compared the free energy landscape of the two sequences. The most important feature of the free energy plots is the overlapping region. The structures in this window are those where the contacts between the amino acids and the ''mutated'' residues are broken. In fact, they have the same conformational energy. In these particular states the energy cost for the transition is very low. We also emphasize that these states are accessible by thermal fluctuations. We believe that our model is able to reproduce the general behavior of allosteric transitions in proteins, where the external agent uses thermal fluctuations to lower the free energy cost of its action. This is also the basis of thermal ratchet models for molecular motors, where the thermal fluctuations are essential to drive the system.

ACKNOWLEDGMENTS

We would like to thank Michele Vedruscolo for an inspiring discussion about the Paracelsus challenge. I.C. thanks Mark Miller and Chinmay Das for advice on computational issues. We gratefully acknowledge detailed comments and suggestions by A. Y. Grosberg. The work of the FOM Institute was part of the research program of FOM and was made possible by financial support from the Netherlands organization for Scientific Research (NWO). An NCF grant of computer time on the TERAS supercomputer is gratefully acknowledged.

APPENDIX

Umbrella sampling and parallel tempering. Umbrella sampling is a method that speeds up the sampling of a rugged free energy landscape by effectively flattening it. A simple way to flatten the landscape is to add a biasing potential to the normal Hamiltonian. To estimate this biasing potential we use an iterative method. During the simulation we sample the probability $P(Q)$ of finding a conformation with order parameter Q [Eq. (7)]. After a specified number of steps we calculate the new biasing potential W with the following recursive equation

$$W_i(Q,T) = W_{i-1}(Q,T) - K \ln P(Q,T) W_0(Q,T) = 0, \tag{A1}$$

where the index i indicates the iteration, and K is a constant which we set to 0.5. Once we have the new biasing potential we add it to the energy in the acceptance criterion of every move. The potential W depends on the instantaneous structure of the system via the order parameter Q, but it also depends on the temperature. This temperature dependence is important when we combine umbrella sampling with parallel tempering. Each temperature has its own biasing potential. The acceptance rule for a temperature swapping move in the parallel-tempering algorithm is then

$$P_{\mathrm{acc}} = e^{\Delta\beta\Delta E + \Delta W}, \tag{A2}$$

$$\Delta W = W(Q_i,T_j) - W(Q_j,T_j) + W(Q_j,T_i) - W(Q_i,T_i),$$

where i and j are replica indices. A similar procedure has recently been used in a paper by Faller *et al.* [25].

[1] G.D. Rose and T.P. Creamer, Proteins: Struct., Funct., Genet. **19**, 1 (1994).
[2] S. Dalal, S. Balasubramanian, and L. Regan, Nat. Struct. Biol. **4**, 548 (1997).
[3] S. Miyazawa and R. Jernigan, Macromolecules **18**, 534 (1985), Table VI.
[4] G. Tiana, Ph.D. thesis, The Neils Bohr Institute, Copenhagen, 2000.
[5] E.I. Shakhnovich and A.M. Gutin, Proc. Natl. Acad. Sci. U.S.A. **90**, 7195 (1993).
[6] B. Derrida, Phys. Rev. B **24**, 2613 (1981).
[7] In the definition of the entropy the constant $\sqrt{\pi\sigma_B^2}$ is ignored, as explained by Derrida [6].
[8] E.I. Shakhnovich and A.M. Gutin, Protein Eng. **6**, 793 (1993).
[9] E.I. Shakhnovich, Phys. Rev. Lett. **72**, 3907 (1993).
[10] E.I. Shakhnovich and A.M. Gutin, Nature (London) **346**, 773 (1990).
[11] V.S. Pande, A.Y. Grosberg, and T. Tanaka, J. Chem. Phys. **103**, 9482 (1995).

[12] V.S. Pande, A.Y. Grosberg, and T. Tanaka, Biophys. J. **73**, 3192 (1997).
[13] V.S. Pande, A.Y. Grosberg, and T. Tanaka, Rev. Mod. Phys. **72**, 259 (2000).
[14] E.N. Govorun, V.A. Ivanov, A.R. Khokhlov, P.G. Khalatur, A.L. Borovinsky, and A.Y. Grosberg, Phys. Rev. E **64**, 040903 (2001).
[15] M-T. Kechadi, R.G. Reilly, K.A. Dawson, Y.A. Kuznetsov, and E.G. Timoshenko, Nuovo Cimento Soc. Ital. Fis., D **20**, 2383 (1998).
[16] D. Frenkel and B. Smit, *Understanding Molecular Simulations* (Accademic Press, New York, 2002), p. 184.
[17] D. Frenkel and B. Smit, *Understanding Molecular Simulations* (Ref. [16]), p. 389.
[18] A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkunov, and P.N. Vorontsov-Vel'yaminov, J. Chem. Phys. **96**, 1776 (1992).
[19] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).
[20] C.J. Geyer and E.A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).
[21] G.M. Torrie and J.P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[22] B.A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
[23] W.R. Schief and Jonathon Howard, Curr. Opin. Cell Biol. **13**, 19 (2001).
[24] F. Jülicher, A. Ajdari, and J. Prost, Rev. Mod. Phys. **69**, 1269 (1997).
[25] R. Faller, Q. Yan, and J.J. de Pablo, J. Chem. Phys. **13**, 5419 (2002).
[26] L. Borovinskiy and A.Yu. Grosberg, J. Chem. Phys. **118**, 5201 (2003).