Sampling activated mechanisms in proteins with the activation-relaxation technique.

Normand Mousseau[(a)]
Department of Physics and Astronomy, Condensed Matter and Surface Science Program, Ohio University, Athens, OH 45701

Philippe Derreumaux
IGS, CNRS-UMR 1889, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20,France

G. T. Barkema
Theoretical Physics, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands

R. Malek
Department of Physics and Astronomy, Condensed Matter and Surface Science Program, Ohio University, Athens, OH 45701
(December 12, 2000)

The activated dynamics of proteins occurs on time scales of milliseconds and longer. Standard all-atom molecular dynamics simulations are limited to much shorter times, of the order of tens of nanoseconds. Therefore, many activated mechanisms that are crucial for the long-time dynamics will not be observed in such molecular dynamics simulations: different methods are required. Here, we describe in detail the activation-relaxation technique (ART) that generates directly activated mechanisms. The method is defined in the configurational energy landscape and defines moves in a two step fashion: (a) a configuration is first brought from a local minimum to a nearby saddle point, the activation; and (b) the configuration is relaxed to a new metastable state, the relaxation. The method has already been applied to a wide range of problems in condensed matter, including metallic glasses, amorphous semiconductors and silica glass. We will review the algorithm in detail, discuss some of the previously published results and present simulations of activated mechanisms for a two-helix bundle protein using an all-atom energy function.

I. INTRODUCTION

The simulation of proteins on a biological time scale, microseconds to milliseconds, is a major challenge in computational biology. For many proteins, such simulations would allow to generate transition state ensembles from the denatured state to the native (experimental) state. For instance, the C-terminal domain of the prion protein has been found to fold in 170 μs and the helical λ repressor protein in only 20 μs [1]. Such studies would also allow to understand how the amino-acid sequence codes for the structure and to annotate the biological function of the protein sequences that cannot be determined by comparative modelling, fold recognition and motif search methods.

Within the framework of molecular dynamics (MD), Duan and co-workers have actually reached the 1-μs milestone in a simulation of a 36-residue peptide, in two months of computation time on 256 parallel processors of a CRAY T3E [2]. There is however still a large gap between one 1-microsecond trajectory for a 36-residue peptide and 10-100 trajectories of millisecond lengths for 200-residue proteins. Approaches that have been tried to go beyond the nanosecond time scale of traditional MD simulations include simulations with multiple time steps, approximate schemes for longe range electrostatic forces and simulations for hard spheres. Multiple time step integration techniques have enlarged the time scales accessible by MD. However still only a few nanoseconds of simulation time are routinely accessible [3]. MD simulations with square-well interactions have provided a basis for understanding behavior that is observed in helical proteins, but their applicability to more realistic potential energy models remains to be determined [4].

Most approaches to study the energy landscape of proteins rely on some simplification of the problem. The solvent, for example, is most often treated as a mean-field addition to the configurational energy of models. Accelerated methods tend to further modify the energy landscape through the introduction of biases, the use of bead-descriptions or discretization [5]. This list is not exhaustive and other methods have been proposed recently to improve sampling while staying as close as possible to a full description of the protein interactions.

The method that we discuss in this manuscript is the activation-relaxation technique (ART), introduced a few years ago by some of us [6]. This method focuses on the the slow activated dynamics of the protein, which brings it from one local minimum to another, in search for the native structure, and removes the fast thermal vibrations that contribute little to the relevant dynamics of these molecules. Searching for activation paths directly in the energy landscape, ART is not sensitive to the complexity of moves as seen in the real three-dimensional space. This technique can therefore generate moves involving hundreds of atoms or barriers many times higher

than k_BT.

In the last few years, ART has been applied with success to a number of materials, including amorphous semiconductors [6–11], Lennard-Jones clusters [12] and binary glasses [6,13], and silica glasses [14], identifying a wealth of relaxation and diffusion mechanisms in these complex materials. Here, we apply ART to a two-helix bundle protein described by an all-atom potential. [15,16] The application of ART to proteins is still in its first stage and we will concentrate here mostly on the methodology.

In the next section, we give a general description of the method, then discuss parameters and difficulties encountered when applying ART to complex systems such as proteins. We then discuss recent results obtained one the two-helix bundle.

II. DESCRIPTION OF ART

As has been recognized for a long time, the mixture of fast and slow dynamics in proteins makes their simulation difficult. In particular, the thermal vibrations play very little rôle in the long time dynamics of proteins, which is dominated by the kinetic energy of the solvent, but nevertheless impose a very short time step. As other techniques, ART addresses this problem by removing the fast vibrations, and focussing on the activated events that determine the long-time structure and dynamics of proteins. This is achieved by considering moves from minimum to minimum, following paths that pass through a saddle point. Activated events, defined as such paths, are defined in the configurational energy landscape, forcing the atomic coordinates, defined in three-dimensional space, to follow. The advantage of such a method over standard real-space Monte Carlo methods is that it does not impose a predefined set of moves and can generate any amount of complexity that is naturally present in the problem.

With ART, the trajectory from a local energy minimum to another is divided into three parts: from the original minimum (where all eigenvalues of the hessian are positive) to a valley (with a negative eigenvalue), from there to a first-order saddle point, and from this saddle point down to the new minimum. The overall configurational energy increases in the first two stages, which represent the *activation*. The energy decreases in the last stage, which is therefore called the *relaxation*. Next, we will discuss the three stages one by one.

A. Finding a valley

Before going into the details of this first stage in ART, it is useful to discuss a few points about energy landscapes; a more complete description can be found in [7] and [17].

Although we know of very few formal mathematical results that can guide us in enumerating saddle points or searching for them from a nearby local minimum, this question has been addressed by a wide range of people over the last 20 years. [18–22] One point needs reemphasized, however: the curvature matrix, or Hessian, at the local minimum contains no information regarding the position or even number of saddle points surrouding this minimum. This can be easily understood by considering a steepest descent path, also called a minimum-energy path, connecting a saddle point to its adjacent minimum. As shown in Fig. 1, and by definition, this path reaches the local minimum along the direction parallel to that with the lowest curvature. (In systems with special symmetries, it is also possible for some minimum-energy paths to enter along directions perpendicular to the direction of lowest curvature.)

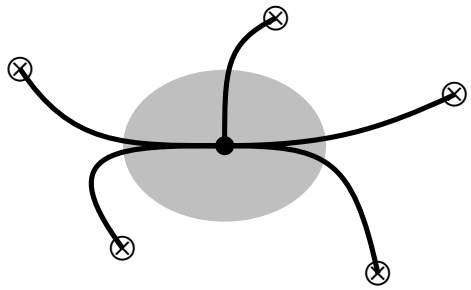

FIG. 1. Minimum energy paths, obtained by steepest descent, for example, from nearby saddle points to the local energy minimum. For generic situations, without any special symmetry, all paths enter the harmonic basin through only two points, parallel to the direction of lowest curvature.

Short of a full exploration of the basin around a minimum, the search for saddle points must therefore be preceded by another process. First, one must search for traces of a "valley", defined by a direction of negative curvature and leading to a saddle point. Once such a valley is found, the configuration can be pushed up-hill along this valley until a saddle point is encountered (the second stage).

Since the Hessian at the minimum does not contain any information regarding the location of saddle points, we simply leave the minimum in a random direction. However, to avoid the generation of multiple events which are spatially separated, we restrict typically the displacement to one atom plus its nearest neighbors. To prevent instabilities from arising, the configuration is slowly pushed along the initial random direction while the energy is minimized in the perpendicular hyperplane, until the lowest eigenvalue, which is computed at every step, falls below a given negative threshold. This threshold is non-zero to ensure a greater chance of convergence to a local saddle point.

In the case of proteins, we know that activated events do not involve an appreciable displacement along the rigid directions (bond and bond-angle stretching). It is therefore possible to select a random displacement in the sub-space of soft directions only. Tests using this limited sub-space do not show any significant improvment in the convergence rate to saddle points, however, and we are currently using the full force at every step in our simulation.

There a are number of ways to leave the harmonic basin. The best choice will depend in good part on the exact nature of the system studied. The convergence to the saddle point, to the contrary, requires few optimizable parameters and is essentially problem-independent. The success rate for finding saddle points depends therefore almost completely on the details of the algorithm chosen for leaving the harmonic basin.

B. Convergence to the saddle point

Once the configuration is pushed outside the harmonic well, a negative eigenvalue appears in the hessian, corresponding to a direction of negative curvature. The configuration is moved upwards along the eigendirection corresponding to this eigenvalue, while relaxing in the perpendicular 3N-1 dimensional hyperplane. If the negative curvature persists, the algorithm guarantees convergence to a first-order saddle point. In some cases, the negative eigenvalue becomes positive before convergence to a saddle point is reached.

Since only the lowest eigenvalue and its corresponding eigenvector is needed, it is not necessary to compute and diagonalize at every step the 3N-dimensional Hessian, an $\mathcal{O}(\mathrm{N}^3)$ operation. A number of schemes exist for extracting extremal eigenvalues and eigenvectors in $\mathcal{O}(\mathrm{N})$ steps. We use here a recursion method, the Lanczós algorithm, [23] which provides the lowest eigenvalues with a very good accuracy in only 15 to 30 forces evaluations. The algorithm is used extensively for the diagonalization of parts of large matrices; details can be found in Refs. [24] and a short overview can be found in Appendix VII.

Maximizing the energy along the direction corresponding to the negative curvature while minimizing in the 3N-1 others, the configuration can be brought in a stable manner as close as desired to the saddle point; as the force is zero at the saddle point, this becomes a stable attractor within this algorithm.

Although the final point corresponds to the transition point in transition state theory, the path followed to reach it is not the minimum energy path. As mentioned above, this would be the steepest descent path which is unstable if followed upwards. The exact trajectory followed in ART depends on a number of factors including the size of the step along the direction of negative curvature as well as on the speed at which we relax the energy in the perpendicular direction.*

This can be seen in Fig. 2 where the energy stored in the soft modes is plotted as a function of iteration number. As the energy minimization in the hyperplane perpendicular to the activated direction is performed in parallel with the displacement upward along the negative curvature, the total energy tends to decrease as the configuration converges to the saddle point. Relaxation at points along this trajectory always brings the protein back to the initial minimum, indicating that although the energy is high, the configuration has not left the basin of attraction of the initial minimum.

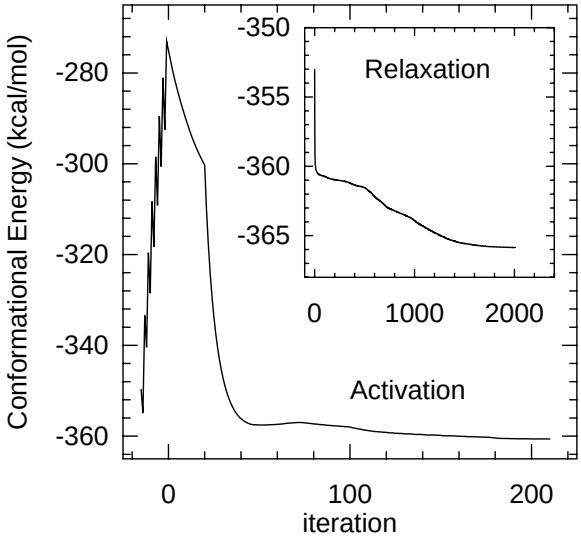

FIG. 2. Conformational energy as a function of iteration number for a generic event on a two-helix bundle away from its native state. The negative iterations correspond to the first stage of the algorithm, leaving the harmonic well. From iteration 0 and up, the configuration is pushed along a valley to a saddle point. The energy mostly goes down because the algorithm relaxes the energy in 3N-1 directions perpendicular to the bottom of the valley as it pushes the configuration up. In inset, we plot the configurational energy as a function of iteration for the minimization step.

C. Relaxation to the new minimum

Once the configuration has converged to a saddle point, defined by a total force reaching zero, an energy minimization routine can bring the configuration in either of the two minima connected by this point. We therefore push slightly the configuration away from the initial minimum before starting our relaxation, to ensure that we

*Although the saddle points found do not depend on these details, the parameters can affect strongly the success rate in finding a saddle point.

move onto a new minimum. We have tested that pulling back will bring us into the original minimum, see, for instance, Ref. [12].

Any standard minimization routine can be used for this step. We typically use a conjugate gradient minimization algorithm [25].

III. A TYPICAL SIMULATION

ART has been used for a number of applications ranging from the structural optimization of amorphous materials to the identification of relaxation and diffusion mechanisms in complex materials.

A typical optimization run starts from a configuration prepared by randomly packing atoms in a box, and locally relaxing this configuration with the result that all forces are zero. Next, an ART event is generated using the procedure described above; the new configuration is then accepted or rejected following a Metropolis test, with a probability $p = \min(1, \exp(-\Delta E/k_B T))$, where ΔE is the energy difference between the final and initial configurations. If the new configuration is accepted, the next ART event is started from the new configuration, otherwise it is started from the previous minimum. This procedure is iterated many times.

This approach was used to produce well-relaxed configurations of disordered systems such as *a*-Si [10,11], *a*-GaAs, [8,9] *g*-SiO_2, [14] and Lennard-Jones glasses. [6,13] In all cases, as the configuration relaxes, its total energy converges to a plateau which is determined by the Metropolis temperature. Variations on this scheme can be used, for example, to study typical activation mechanisms on a given energy surface.

IV. PROPERTIES OF ART

Although the basics of ART are well established, it is important to characterize in some detail the sampling biases of the method as well as the impact of the various variables that appear in the implementation of the algorithm on the events found and its completeness.

We have recently performed a detailed characterization of the properties of ART using Lennard-Jones clusters, for which considerable information already exists [26]. Details of the simulations and analysis can be found elsewhere [12]; we give here only a brief description of the procedure and the final results.

For this study, we looked at the energy landscape around a generic local minimum for three Lennard-Jones clusters with sizes 13, 38 and 55 atoms, respectively. The local minima correspond to reasonably well-relaxed structures but do not display any special symmetry that could affect our results.

The first questions addressed concern the reversibility of the trajectories as defined by three points in the configuration space: initial minimum, saddle point, and final minimum. First, we must make sure that from the saddle point, it is possible to relax back to the initial minimum as well as reach a new minimum. The former requirement ensures that the configuration has not left the original basin in the search for a saddle point, while the latter just makes sure that a real saddle point was indeed found. Because we find events, we know that new minima can be generated. To check the reversibility, we pull back slightly the configuration away from the saddle point and minimize the total energy using a conjugate gradient scheme. This is repeated on 3000 saddle points and we find that all configurations but four of the 3000 at saddle points can be relaxed back to the initial local minimum. (The four ones relaxed to a shallow minimum very close by.) These results demonstrate that if the activation is done carefully, the saddle points found really separate two neighboring minima.

Second, we verify that the whole trajectory is also reversible in the sense that a given event, going from minimum A to minimum C through saddle point B, can also be generated as a C-B-A event. This is checked by selecting at random 30 events for each cluster size, going to the new minima, and generating new events from there, to try to find a reverse path, bringing the configuration back to the initial minimum through the same saddle point. In all cases, this could be done: ART generates fully reversible paths and identifies properly the dominant points on these paths.

Another question is that of completeness: does ART miss any class of first-order saddle points? To answer this question, we sample the saddle points around each minima using two methods, ART and a related but differently biased eigenvector-following technique introduced by Doye and Wales [26]. Starting from the same generic minimum for both methods, the activated paths away from this minimum are systematically searched with both methods. Although the DW approach produces only a fraction of all saddle points around a given minimum, we expect that if ART missed classes, a number of events found by DW method would not be found by ART. For all cluster sizes, however, ART recovers all events generated by DW and many more. We can therefore be confident that ART can find *all* first-order saddle points around a minimum in a given system provided that the sampling is exhaustive. Figure 3 shows the number of different saddle points and minima (after removing all permutational isomers) generated by ART on 13-atom cluster; within 14000 events, all 180 saddle points and 87 minima are visited at least once.

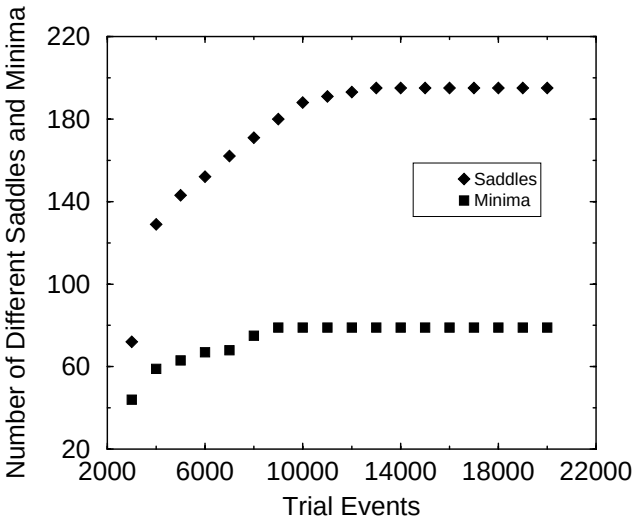

FIG. 3. Number of different saddle points and minima found around a generic minimum in a 13-atom Lennard-Jones cluster as a function of ART events. Permutational isomers are removed. Each ART event starts within a randomly chosen direction.

V. APPLICATIONS TO PROTEINS

As a first application of ART to proteins, we study a 38-residue α-helical hairpin described by an all-atom potential. This simple helix bundle was chosen because it has been characterized by NMR spectroscopy to have a helix content of 89% at pH 3.6 and 25°C. [27] Furthermore this motif is an essential constituent in protein structures. For this test, we do not try to converge to the native (experimental) conformation — it is well known that this task is currently beyond reach of all methods for unbiased potentials and unbiased conformational moves starting from the fully extended conformation, but rather we relax the protein and examine the states we can generate.

In this section, we first describe the interaction potential and then present the details of the simulation. Results are discussed in the next section.

A. Interaction potential

The 38-residue peptide was modeled by using an OPEP-like interaction including an off-lattice representation and a knowledge-based potential. [15,16]

In this work, we use a flexible-geometry peptide model where each amino-acid is represented by six particles on average, i.e. N, H, Cα, C, O and one bead with an appropriate van der Waals radius and position for the side chains [15]. This chain representation reproduces experimental structures exactly. This is not always possible in other simplified or lattice-based schemes. Furthermore, this model includes many more local minima than most simplified representations using one or two particles per amino acid.

Potential functions and parameters are commonly derived from a population of known protein structures. Unlike many simplified force fields, our potential does not contain ad-hoc biases based on foreknowledge of the target structure (e.g. *a priori* location of secondary structure elements) and was optimized on the structure of four training peptides with 10-28 residues. This was done by maximizing the stability gap between the energy of the native structure and a representative ensemble of non-native structures [15]. This potential was then tested on more complex topologies including two and three-helix bundles and the $\beta\alpha\beta$ fold [28,16].

Since the main goal of this study is to determine the optimal ART strategy for proteins, we ignore the α-helical, β strand and α_L propensity contributions for the amino acids described by the OPEP potential (Optimized Potential for Efficient peptide-structure Prediction). In summary, our OPEP-like potential includes harmonic terms for maintaining the bond lengths and bond angles near their equilibrium values and nonbonded interaction terms. Nonbonded interactions between main chain atoms were modeled by a Lennard Jones potential and a coulombic potential with all interactions included. All main chain parameters including force constants and atomic charges were taken from the literature and the dielectric function was set to $2r$. Nonbonded interactions were modeled by a 6-12 Lennard-Jones potential between side chains with hydrophobic character and by a repulsive 12-potential otherwise. The maximum contact energy is of the order of -3.0 kcal/mol for the isoleucine-isoleucine interaction.

B. Details of simulation

The simulation is performed completely with ART. Starting from the fully extended conformation, ART moves to new minima are proposed iteratively; these new minima are fully relaxed, at zero temperature. Each new event is accepted or rejected following a Metropolis criterion with a temperature of 300 K. As the configuration settles down in a deep minimum, it becomes evident that a simulated annealing, at higher temperatures, is necessary to sample the space of configurations away from this state. Results presented here include therefore runs at Metropolis temperatures of 600 K and 1200 K. An unfolding (increasing temperature) trajectory was not attempted to determine the folding temperature.

The sequence of energies at the minima is given in Figure 4. For these events, we select a local initial random displacement, involving about a third of the protein. As mentioned above, the activation is restricted to the soft

modes of the protein, excluding bond and bond-angle stretching.

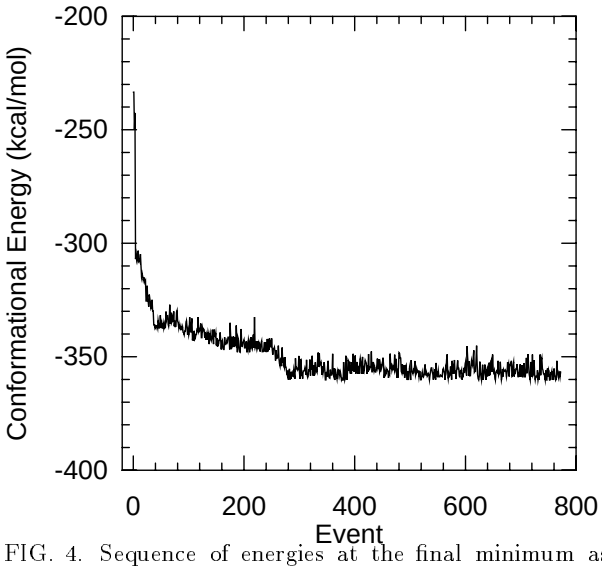

FIG. 4. Sequence of energies at the final minimum as a function of event number for the two-α helix bundle for a temperature of 300 K. The initial configuration is the fully extended conformation. With the OPEP-like potential, the energy of the native state is -422 kcal/mol. The overall acceptance rate for this run is about a third.

At each step, the lowest eigenvalue and its corresponding eigenvector are obtained using a 15 step Lanczós recursion, starting from the eigenvector identified at the previous iteration. To ensure that there is no bias introduced with this scheme, a 25 step lanczós is computed every 10 steps starting from a random vector; no significant change in eigenvector or eigenvalue is found between these two approaches. With the current parameters, it takes on average 5000 force evaluations to reach a saddle point, and about half that to relax to a new minimum. On a fast workstation (500 Mhz Dec Alpha 21264), an event requires on average slightly less than two minutes.

The deformation of the protein can vary significantly from event to event. The total displacement, defined by the square root of the sum of squared atomic distances between the two configurations, can be anywhere between 1 Å and 70 Å. The space of configurations can be efficiently sampled by ART.

C. Results

Before addressing the overal efficiency of ART for sampling the conformational energy landscape of proteins, we first discuss individual events generated by ART. The most common type of events involves almost all atoms in the protein, and deforms it slightly without bringing it to a configuration which is qualitatively different. The typical total displacement for this kind of events varies between 5 and 15 Å. Although these events reflect the structure of the conformational energy landscape, we find that a series of them rarely leads to a qualitatively different configuration; these moves tend to average out the total displacement to zero.

The challenge for ART (and any other simulation technique) lies in the generation of more significant moves (often called basin-hop moves). These are generally associated with large deformation of the protein, from about 20 Å to 70 Å or more. Figure 5, for example, show two configurations separated by a single barrier. The proportion of these events among all those generated depends on a number of factors including parameters and the conformation itself; the more stable the protein, the more difficult it is to generate hops. Taking as a threshold a displacement of 20 Å, we find that for our best simulations around 5 % of the events are significant. This is still slightly low because most of these moves are rejected; we are actively working on improving this ratio.

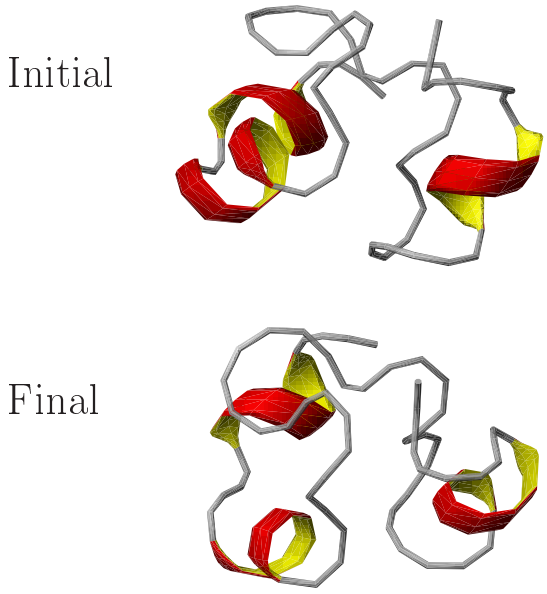

FIG. 5. A large ART event. The total displacement between the initial (-386 kcal/mol) and the final configurations (-373 kcal/mol) is 70 Å. Although the energy difference is too big for this move to be physically relevant, this move was chosen because the differences between the two conformations can be easily seen directly.

Now that we have discussed individual events, we can look at trajectories. Starting from the unfolded conformation, the protein rapidly folds in a compact state; in five to eight accepted events, the protein reaches a metastable energy minimum, dropping its energy from about -190 to -320 kcal/mol. The conformation reached is almost spherical and, as expected, bears little resemblance to the native state. This new conformation shows

a number of hydrogen bonds involving helical and β-strand segments, and turns out to be rather stable. With a Metropolis temperature of 300 K, the protein relaxes down to -360 kcal/mol in less than 300 to 500 events (less than 70 to 100 accepted events). At this point, the structure still shows a compact, spherical shape (Fig. 6). Further relaxation during 800 steps at this temperature does not lead to any major structural change.

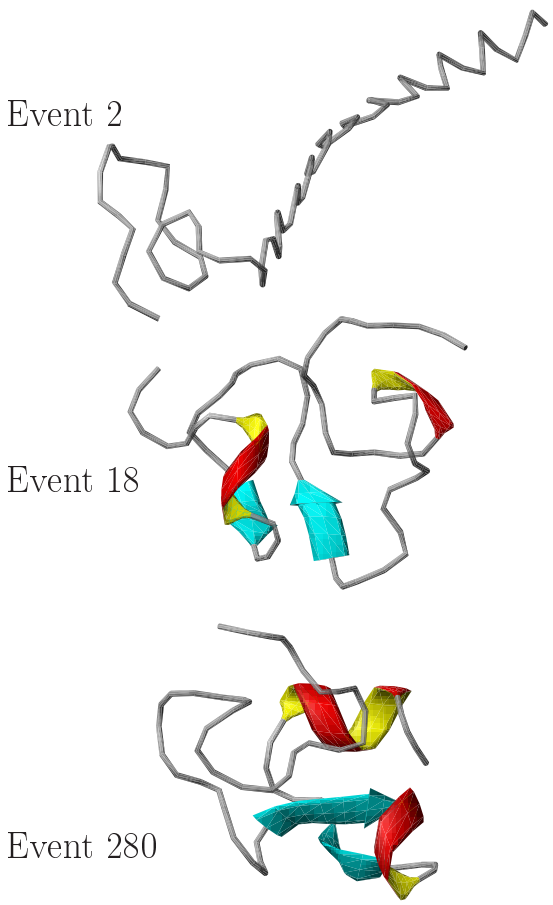

FIG. 6. Three conformations of the 38-residue helical hairpin model in an ART-simulation at T=300K starting from the fully extended state. The top conformation has an energy of -233 kcal/mol. The middle, the 6th accepted event in the series, -316 kcal/mol and the last one (61st accepted event), -360 kcal/mol.

In order to sample the phase space sufficiently, it is necessary to simulate at higher temperatures. Figure 7 shows a series of snapshots from one run at 1200 K. At this temperature, it is possible to open up the spherical conformation and refold the protein in a native-like shape. In the conformation 7(b), residues 2-15 are essentially helical (vs. 2-15 in the NMR structure), residues 24-29 are helical (vs. 23-35 in the NMR structure) and the C-terminal region is folded against residues 9-10 (vs. 2-3 in the NMR structure). We emphasize that we did not try to escape from this partially unfolded state. However, not all runs move easily towards the α-helical hairpin shape; it is possible to remain in the vicinity of the compact conformation described in 7(a) and lower the energy down to -388 kcal/mol or so. Such a variety of folding behaviors is expected in the framework of the funnel concept and has been noted for α-helical proteins modeled by one particle per residue [4,29]. As a result, it can be very hard, using single-temperature runs, to find the hairpin basin of attraction.

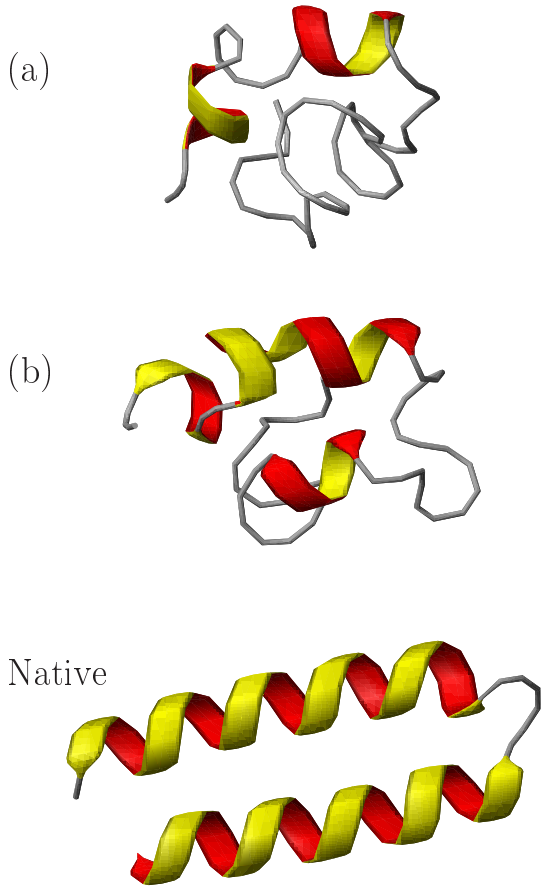

FIG. 7. Starting from a spherically compact conformation (a), of energy -360 kcal/mol, a 1000-event ART simulation at T=1200 K leads to conformation (b), at -377 kcal/mol, which starts to show the two-helices present in the native state (c).

To go beyond these limitations, it is necessary to use more advanced sampling techniques such as parallel tempering (See Appendix VIII). We are currently in the process of implementing this technique.

VI. DISCUSSION AND CONCLUSION

We have shown through a series of snapshots that the ART algorithm is able to generate in a reasonable amount of computer time a wide range of conformational substates and states for the 38-residue helical hairpin model.

Transitions from compact random structures to partially helical-hairpin structures are detected. This preliminary result is significant since ART does not use a set of conformational moves, and escapes from compact structures is a very difficult task by all-atom MD simulations. In addition, the potential of mean force used to describe the polypeptide energy surface is general (i.e. not biased towards a particular conformation) and includes energy components resulting from both high and low-frequency motions. Although further study is needed to optimize the ART strategy, ART can be a useful tool for exploring two specific areas of the protein folding problem. The first area involves searches for iso-energetic conformations of enzymes with hinge-bending motion capabilities. The second direction involves determination of folding mechanisms for all β and $\alpha\beta$ proteins with 70 amino-acids using a potential favoring contacts present in the native structure (Go-like potential). [30]

ACKNOWLEDGEMENT

NM is grateful to the NSF for support under grant number DMR-9805848

VII. APPENDIX A: THE LANCZÓS ALGORITHM

The Lanczós algorithm [23] provides an efficient way to extract a limited spectrum of eigenvalues and eigenvectors for a high-dimensional problem without requiring the evaluation and diagonalization of the full $3N \times 3N$ matrix. This algorithm is a recursion method based on the repeated application of the local Hessian on a random displacement vector. [24]

We therefore start with a random eigenvector, $|\vec{x}_0\rangle$, representing a small displacement vector away from the current point (which does not need to be a minimum), $\vec{x}$. Applying the Hessian matrix H onto this vector is equivalent to computing the difference between the force at $|x\rangle + |x_0\rangle$ and at $|x\rangle$. We do not need to compute, therefore, the Hessian at any point.

The result of the application of the Hessian on $|x_0\rangle$ can be decomposed as a linear combination of this initial random vector and a second one, $|x_1\rangle$, perpendicular to the first. The full recursion scheme becomes:

$$\begin{aligned}
H|x_0\rangle &= a_0|x_0\rangle + b_1|x_1\rangle \\
H|x_1\rangle &= a_1|x_1\rangle + b_1|x_0\rangle + b_2|x_2\rangle \\
H|x_2\rangle &= a_2|x_2\rangle + b_2|x_1\rangle + b_3|x_3\rangle \\
&\vdots \\
H|x_{l-1}\rangle &= a_{l-1}|x_{l-1}\rangle + b_{l-1}|x_{l-2}\rangle + b_l|x_l\rangle \\
H|x_l\rangle &= a_l|x_l\rangle + b_l|x_{k-1}\rangle
\end{aligned} \tag{1}$$

where we force a closure at $l < 3N$, with $3N$ the dimension of the problem. The application of the Hessian to the position requires a single force evaluation and is an $\mathcal{O}(\mathcal{N})$ operation. The complete construction of the $l \times l$ matrix requires therefore $3lN$ operations. If $l \ll 3N$, this step dominates over the diagonalization of the tridiagonal matrix L,

$$L = \begin{pmatrix}
a_0 & b_1 & & & & \\
b_1 & a_1 & b_2 & & & \\
 & \ddots & \ddots & \ddots & & \\
 & & b_{l-1} & a_{l-1} & b_l & \\
 & & & b_l & a_l &
\end{pmatrix} \tag{2}$$

Because of the artificial (and rather brutal) closure in the recursion, only the lowest eigenvalues and eigenvectors are accurate. This does not cause any problem here since we only require the very lowest eigenvalue and its corresponding eigenvector. Typically, $10 < l < 30$ is sufficient to extract the lowest eigenvalue with sufficient accuracy. This is especially so because we can make use of the eigendirection used in the previous iteration as a seed for the new recursion.

VIII. APPENDIX B: PARALLEL TEMPERING

The basic idea behind parallel tempering is to perform several simulations simultaneously on the same system, but at different temperatures. [31] The system which runs at the lower temperature will have a stronger tendency to be lower in energy. This tendency can however suppress activated processes heavily, and consequently slow down the dynamics. During relaxation from an out-of-equilibrium situation, it happens therefore regularly that the system which runs at a low temperature gets stuck in a region of phase space with relatively high energy, while the system running at a higher temperature relaxes faster and actually reaches lower energies.

In parallel tempering, one swaps the states of the system in two of the simulations every so often, with a certain probability that is chosen so that the states of each system still follow the Boltzmann distribution at the appropriate temperature. More precisely, if the simulations at T_1 and T_2 have energies E_1 and E_2, respectively, the usual acceptance probability for a swap is given by

$$A_{swap} = \mathrm{Min}\left[1, \exp\left(-\left(\frac{1}{k_b T_1} - \frac{1}{k_b T_2}\right)(E_2 - E_1)\right)\right]. \tag{3}$$

The result is that the higher-temperature simulation helps the lower-temperature one across the energy barriers in the system. (See Ref. [32] for more details.)

IX. REFERENCES

[a] E-mail: mousseau@helios.phy.ohiou.edu

[1] G. Wildegger, S. Liemann, and R. Glockshuber, *Extremely rapid folding of the C-terminal domain of the prion protein without kinetic intermediates*, Nat. Struct. Biol. **6**, 550-553 (1999).

[2] Y. Duan and P. Kollman, *Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution*, *Science* **282**, 740 (1999).

[3] T. Schlick, E. Barth and M. Mandziuk, *Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation*, Ann. Rev. Biophys. Biomol. Struc. **26**, 179-220 (1997).

[4] Y. Zhou and M. Karplus Interpreting the folding kinetics of helical proteins, *Nature* **401**, 400-403 (1999).

[5] E.I. Shakhnovich, *Theoretical studies of protein-folding thermodynamics and kinetics*, Curr. Opinion in Struct Biol. **7**, 29-40 (1997); J-E. Shea, J.N. Onuchic and C.L. Brooks, *Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A*, Proc. Natl. Acad. Sci. USA **96**, 12512-12517 (2000).

[6] G. T. Barkema and N. Mousseau, *Event–based relaxation of continuous disordered systems*, Phys. Rev. Lett. **77**, 4358-4361 (1996).

[7] N. Mousseau et G. T. Barkema, *Traveling through potential energy landscapes of disordered materials: The activation-relaxation technique*, Phys. Rev. E **57**, 2419 (1998).

[8] N. Mousseau et L. J. Lewis, *Topology of amorphous tetrahedral semiconductors on intermediate lengthscales*, Phys. Rev. Lett. **78**, 1484-1487 (1997).

[9] N. Mousseau et L. J. Lewis, *Structural, electronic and dynamical properties of amorphous gallium arsenide: a comparison between two topological models*, Phys. Rev. B **56**, 9461-9468 (1997).

[10] G. T. Barkema et N. Mousseau, *Identification of relaxation and diffusion mechanisms in amorphous silicon*, Phys. Rev. Lett. **81**, 1865-1868 (1998).

[11] N. Mousseau et G. T. Barkema, *Activated mechanisms in amorphous silicon: an activation-relaxation-technique study*, Phys. Rev. B **61**, 1898-1906 (2000).

[12] R. Malek and N. Mousseau, *Dynamics of Lennard-Jones clusters: A characterization of the activation-relaxation technique*, Preprint cond-mat/0006042 (2000).

[13] N. Mousseau, *Cooperative motion in Lennard-Jones binary mixtures below the glass transition*, Preprint cond-mat/0004356 (2000).

[14] N. Mousseau, G.T. Barkema et S.W. de Leeuw, *Elementary mechanisms governing the dynamics of silica*, J. Chem. Phys. **112**, 960-964 (2000).

[15] P. Derreumaux, *From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential*, J. Chem. Phys. **111**, 2301-2310 (1999).

[16] P. Derreumaux, *Generating ensemble averages for small proteins from extended conformations by Monte Carlo simulations*, Phys. Rev. Lett. **85**, 206-209 (2000).

[17] N. Mousseau and G.T. Barkema, *Exploring high-dimensional energy landscapes*, Comp. Sci. Eng. **1**, 74-82 (1999).

[18] M. J. Rothman and L. L. Lohr Jr., *Analysis of an energy minimization method for locating transition-states on potential-energy hypersurfaces*, Chem. Phys. Lett. **70**, 405-409 (1980).

[19] C. J. Cerjan and W. H. Miller, *On finding transition-state*, J. Chem. Phys. **75**, 2800-2806 (1981).

[20] J. Simons, P. Jorgensen, H. Taylor and J. Ozment, *Walking on potential-energy surfaces*, J. Phys. Chem. **87**, 2745-2753 (1983).

[21] R. S. Berry, H. L. Davis and T. L. Beck, *Finding saddles on multidimensional potential surfaces*, Chem. Phys. Lett. **147**, 13-17 (1988); I. V. Ionova and E. A. Carter, *Ridge method for finding saddle-points on potential-energy surfaces*, J. Chem. Phys. **98**, 6377-6386 (1993).

[22] S. F. Chekmarev, *A simple gradient-method for locating saddles*, Chem. Phys. Lett. **227**, 354-360 (1994).

[23] C. Lanczos, Applied Analysis (Dover, New York, 1988).

[24] Otto F. Sankey, David A. Drabold and Andrew Gibson, *Projected random vectors and the recusion method in the electronic-structure problem*, Phys. Rev. B **50**, 1376-1381 (1994).

[25] W. H. Press et al., *Numerical Recipes*, Cambridge University Press, Cambridge, 1988.

[26] J. P. K. Doye and D. J. Wales, *Surveying a potential energy surface by eigenvector-following*, Z. Phys. D **40**, 194-197 (1997).

[27] Y. Fezoui, P. J. Connolly, J. J. Osterhout, *Solution Structure of alpha-t-alpha, a Helical Hairpin Peptide of De Novo Design*, Prot. Sci. **6**, 1869-1877 (1997).

[28] P. Derreumaux, *Predicting helical hairpins from sequences by Monte Carlo simulations*, J. Comput. Chem. **21**, 582-589 (2000).

[29] J.D. Bryngelson, J.N. Onuchic, N.D. Socci and P.G. Wolynes, *Funnels, pathways, and the energy landscape of protein-folding - a synthesis*, *Proteins* **21**, 167-195 (1995).

[30] H. Taketomi, Y. Ueda and N. Go *Studies on protein folding, unfolding and fluctuations by computer simulation*, Int. J. Pept. Protein Res. **7**, 445-459 (1975).

[31] E. Marinari and G. Parisi, *Simulated tempering – a new Monte-Carlo scheme*, Europhys. Lett. **19**, 451-458 (1992)

[32] M. E. J. Newman and G. T. Barkema , *Monte Carlo Methods in statistical physics*, Oxford University Press (Oxford), 480 pp., 1999.