

# Sampling activated mechanisms in proteins with the activation–relaxation technique

Normand Mousseau,\* Philippe Derreumaux,† G.T. Barkema,‡ and R. Malek§

\*Department of Physics and Astronomy, Condensed Matter and Surface Science Program, Ohio University, Athens, OH, USA

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

‡Theoretical Physics, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands

§Department of Physics and Astronomy, Condensed Matter and Surface Science Program, Ohio University, Athens, OH, USA

*The activated dynamics of proteins occur 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 long-time dynamics will not be observed in such molecular dynamics simulation; 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 first-order 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 review the algorithm in detail, discuss some previously published results and present simulations of activated mechanisms for a two-helix bundle protein using an all-atom energy function.*

© 2001 by Elsevier Science Inc.

**Keywords:** activation-relaxation technique (ART); molecular dynamics (MD); Optimized Potential for Efficient peptide-structure Prediction (OPEP)

## INTRODUCTION

The simulation of proteins on a biological time scale of microseconds to milliseconds is a major challenge in computational

biology. For many proteins, such simulations would allow the generation of 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  $\mu$ s and the helical  $\lambda$  repressor protein in only 20  $\mu$ s.<sup>1</sup> Such studies would also lead to an understanding of 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 modeling, fold recognition, and motif search methods.

Within the framework of molecular dynamics (MD), Duan and coworkers have actually reached the 1- $\mu$ s milestone in a simulation of a 36-residue peptide using two months of computation time on 256 parallel processors of a CRAY T3E.<sup>2</sup> There is, however, still a large gap between one 1- $\mu$ s trajectory for a 36-residue peptide and 10–100 trajectories of ms lengths for 200-residue proteins. Approaches that have been tried to go beyond the ns time scale of traditional MD simulations include simulations with multiple time steps, approximate schemes for long 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 ns of simulation time are routinely accessible.<sup>3</sup> 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.<sup>4</sup>

Most approaches to studying 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.<sup>5</sup> This list is not exhaustive and other methods have been

Corresponding author: N. Mousseau, Department of Physics and Astronomy, Condensed Matter and Surface Science Program, Ohio University, Athens, OH 45701, USA.

E-mail address: mousseau@helios.phy.ohiou.edu (N. Mousseau)

proposed recently to improve sampling while staying as close as possible to a full description of protein interactions.

The method that we discuss in this manuscript is the activation-relaxation technique (ART), introduced a few years ago by Barkema and Mousseau.<sup>6</sup> This method focuses on the slow activated dynamics of the protein, which bring it from one local minimum to another, in search for the native structure, and remove 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 real three-dimensional space. This technique can therefore generate moves involving hundreds of atoms or barriers many times higher than  $k_B T$ .

In the last few years, ART has been applied with success to a number of materials, including amorphous semiconductors,<sup>6–11</sup> Lennard-Jones clusters,<sup>12</sup> binary glasses,<sup>6,13</sup> and silica glasses,<sup>14</sup> 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.<sup>15,16</sup> 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 for the two-helix bundle.

## DESCRIPTION OF ART

It has long been recognized that the mixture of fast and slow dynamics in proteins makes their simulation difficult. In particular, thermal vibrations play very little role in the long-time dynamics of proteins, which are dominated by the kinetic energy of the solvent but nevertheless impose a very short time step. As with other techniques, ART addresses this problem by removing the fast vibrations and focusing 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 first-order saddle point. These paths are known as activated events and 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 a valley to a first-order saddle point, and from this saddle point down to the new minimum. The total energy increases in the first two stages, which represent the activation. The energy decreases in the last stage, which is therefore called the relaxation.

### 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 elsewhere.<sup>7,17</sup>

Although we know of very few formal mathematical results

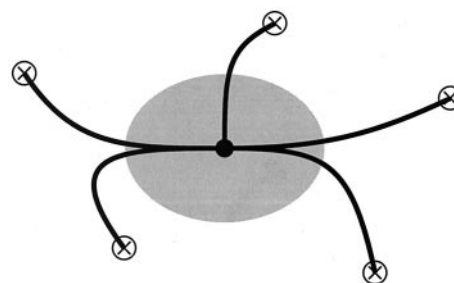
that can guide us in enumerating saddle points or searching for them from a nearby local minimum, this question has been addressed by numerous researchers in the last 20 years.<sup>18–22</sup> One point needs reemphasized, however: the curvature matrix, or Hessian, at the local minimum contains no information regarding the position or even the number of saddle points surrounding 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 Figure 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.

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 uphill 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 that are spatially separated, we typically restrict 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 nonzero 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 subspace of soft directions only. Tests using this limited subspace do not show any significant improvement in the convergence rate to saddle points, however, and we are currently using the full force at every step in our simulation.

There are a number of ways to leave the harmonic basin. The best choice will depend in good part on the exact nature of the



*Figure 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.*

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.

### Convergence to the Saddle Point

Once the configuration is pushed outside the harmonic basin, a negative eigenvalue appears in the hessian, corresponding to a direction of negative curvature. The configuration is moved upward 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  $\mathcal{O}(N^3)$  operation. A number of schemes exist for extracting extremal eigenvalues and eigenvectors in  $\mathcal{O}(N)$  steps. We use here a recursion method, the Lanczós algorithm,<sup>23</sup> which provides the lowest eigenvalues with a very good accuracy in only 15–30 force evaluations. The algorithm is used extensively for the diagonalization of parts of large matrices; details can be found in an article by Sankey et al.<sup>24</sup> and a short overview can be found in Appendix A.

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

This can be seen in Figure 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.

### Relaxation to the New Minimum

Once the configuration has converged to a saddle point, defined by the 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 the configuration away slightly from the initial minimum before starting our

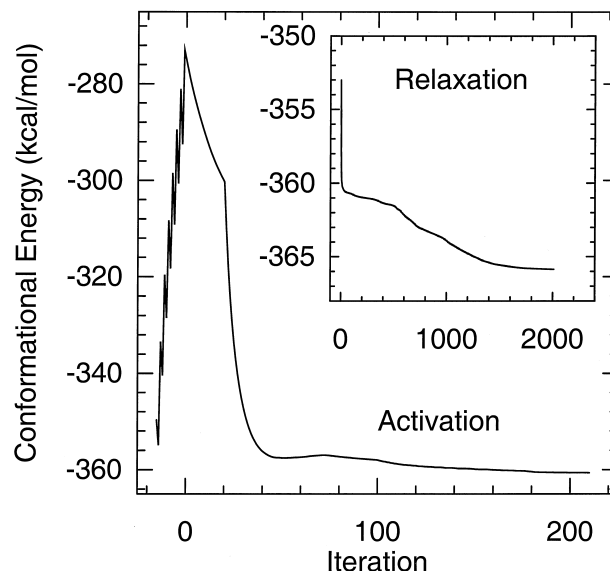


Figure 2. Conformational energy as a function of iteration number for a generic event on a two-helix bundle away from its native state. The negative iterations correspond to the first stage of the algorithm, leaving the harmonic basin. From iteration 0 and up, the configuration is pushed along a valley 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 total energy as a function of iteration number during the relaxation.

relaxation, to ensure that we move on to a new minimum. We have verified that pulling back from the saddle point brings the configuration back to the original minimum (see, for instance, Malek and Mousseau<sup>12</sup>).

Any standard minimization routine can be used for this step. We typically use a conjugate gradient minimization algorithm.<sup>25</sup>

## 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  $\Delta E$  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,<sup>10,11</sup> *a*-GaAs,<sup>8,9</sup> *g*-SiO<sub>2</sub><sup>14</sup> and Lennard-Jones glasses.<sup>6,13</sup> In all cases, as the configuration relaxes, its total energy converges to a plateau

that 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.

## 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 recently performed a detailed characterization of the properties of ART using Lennard-Jones clusters, for which considerable information already exists.<sup>26</sup> Details of the simulations and analysis can be found elsewhere<sup>12</sup>; 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 of 13, 38, and 55 atoms. 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 ensures 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 the configuration back slightly from the saddle point and minimize the total energy using a conjugate gradient scheme. This is repeated on 3,000 saddle points and we find that all configurations but 4 of the 3,000 at saddle points can be relaxed back to the initial local minimum. (The 4 saddle points 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 (the DW method).<sup>26</sup> 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 the 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 a 13-atom cluster; within 14,000 events, all 180 saddle points and 87 minima are visited at least once.

## APPLICATIONS TO PROTEINS

As a first application of ART to proteins, we study a 38-residue  $\alpha$ -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.<sup>27</sup> 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 the 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.

### Interaction Potential

The 38-residue peptide was modeled with OPEP-like interactions, including an off-lattice representation and a knowledge-based potential.<sup>15,16</sup>

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, C $\alpha$ , C, O, and one bead with an appropriate van der Waals radius and position for the side chains.<sup>15</sup> This chain representation reproduces experimental structures ex-

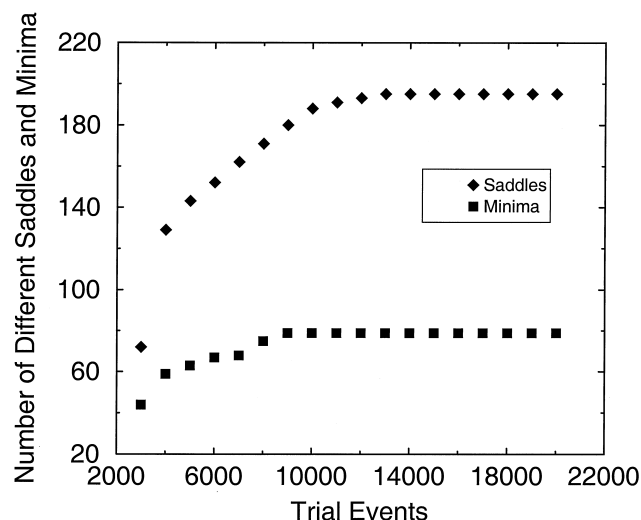


Figure 3. Number of different saddle points and minima found around a generic minimum in a 13-atom Lennard-Jones cluster as a function of ART events. Permutational isomers are removed. Each ART event starts within a randomly chosen direction.



actly. This is not always possible in other simplified or lattice-based schemes. Furthermore, this model includes many more local minima than most simplified representations that use 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 nonnative structures.<sup>15</sup> This potential was then tested on more complex topologies including two- and three-helix bundles and the  $\beta\alpha\beta$  fold.<sup>28,16</sup>

Since the main goal of this study is to determine the optimal ART strategy for proteins, we ignore the  $\alpha$ -helical,  $\beta$  strand and  $\alpha_L$  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  $2r$ . Non-bonded 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.

## 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.

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 an average 5,000 force evaluations to reach a saddle point, and about half that number to relax to a new minimum. On a fast

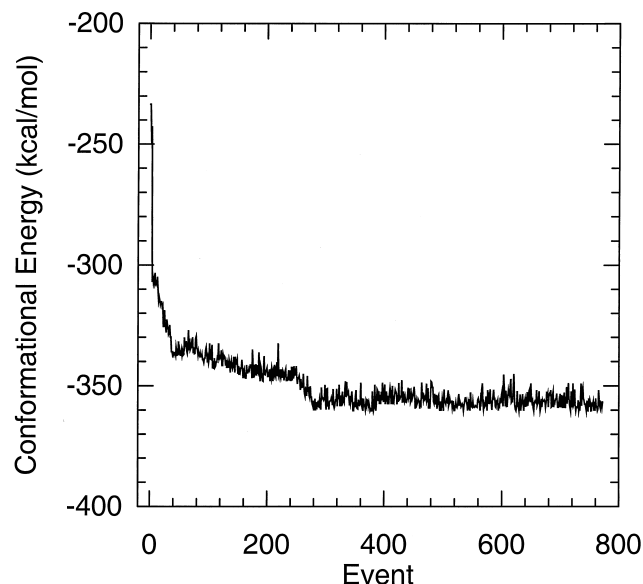


Figure 4. Sequence of energies at the final minimum as a function of event number for the two- $\alpha$  helix bundle for a temperature of 300 K. The initial configuration is the fully extended conformation. With the OPEP-like potential, the energy of the native state is  $-422$  kcal/mol. The overall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

workstation (500 Mhz DEC Alpha 21264), an event requires on average slightly less than 2 min.

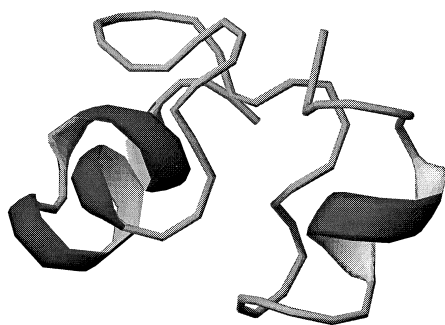
The deformation of the protein can vary significantly from event to event. The total displacement, defined by the square root of the sum of squared atomic distances between the two configurations, can be anywhere between  $0.065$  Å Root Mean Square Deviation (RMSD), and  $4.58$  Å RMSD. The space of configurations can be efficiently sampled by ART.

## RESULTS

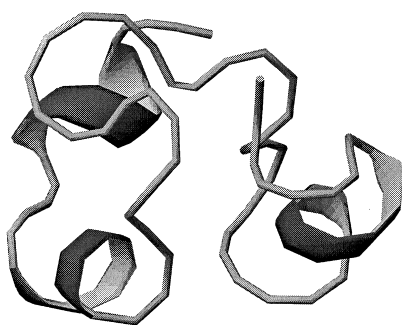
Before addressing the overall efficiency of ART for sampling the conformational energy landscape of proteins, we first discuss individual events generated by ART. The most common type of event involves almost all atoms in the protein and deforms the protein slightly without bringing it to a configuration that is qualitatively different. The typical displacement for this kind of events varies between  $0.3$  Å RMSD and  $1$  Å RMSD. 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 a large deformation of the protein, from about  $1.3$  Å RMSD, to  $6$  Å RMSD, or more. Figure 5, for example, shows 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.

Initial



Final



*Figure 5. A large ART event. The total displacement between the initial ( $-386$  kcal/mol) and the final configurations ( $-373$  kcal/mol) is  $4.58$  Å RMSD. 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.*

Taking as a threshold a displacement of  $1.3$  Å RMSD, 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.

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  $\beta$ -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 (Figure 6, left panel). Further relaxation during 800 steps at this temperature does not lead to any major structural change.

To sample the phase space sufficiently, it is necessary to simulate at higher temperatures. Figure 6 (right panel) 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 conformation 6(b) (Figure 6, right panel), residues 2–15 are essentially helical (versus 2–15 in the NMR structure), residues 24–29 are helical (versus 23–35 in the NMR structure) and the C-terminal region

is folded against residues 9–10 (versus 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 toward the  $\alpha$ -helical hairpin shape; it is possible to remain in the vicinity of the compact conformation described in conformation 6(a) (right panel) and lower the energy down to approximately  $-388$  kcal/mol. Such a variety of folding behaviors is expected in the framework of the funnel concept and has been noted for  $\alpha$ -helical proteins modeled by one particle per residue.<sup>4,29</sup> As a result, it can be very hard using single-temperature runs to find the basin of attraction of the hairpin conformation.

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

## 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 to escape 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 toward 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  $\beta$  and  $\alpha\beta$  proteins with 70 amino acids using a potential favoring contacts present in the native structure (Go-like potential).<sup>30</sup>

## APPENDIX A:

### The Lanczós Algorithm

The Lanczós algorithm<sup>23</sup> 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.<sup>24</sup>

We therefore start with a random eigenvector,  $|x_0\rangle$ , representing a small displacement vector away from the current point  $|x\rangle$ , which does not need to be a minimum. 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

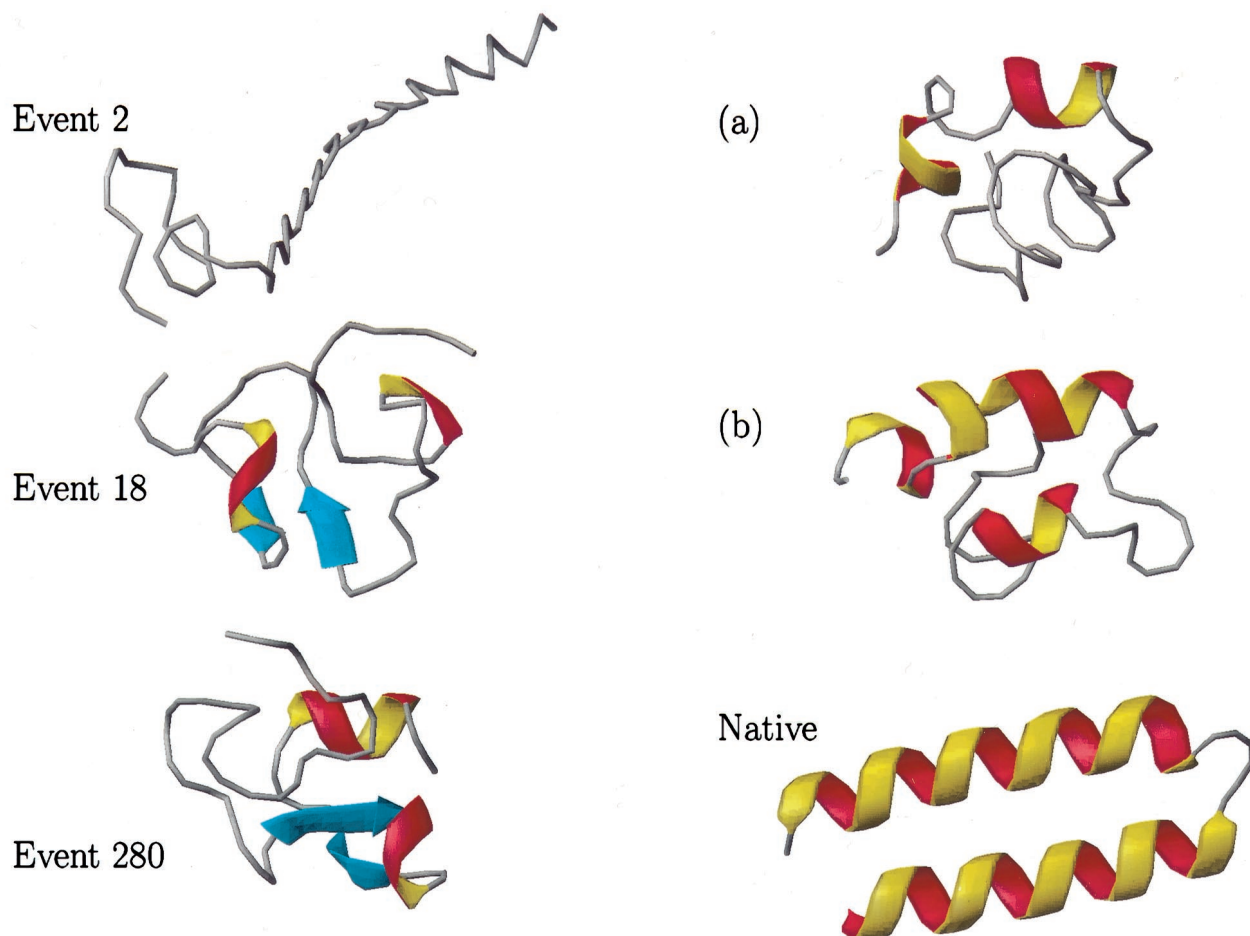


Figure 6. Left: Three conformations of the 38-residue helical hairpin model in an ART simulation at  $T = 300$  K starting from the fully extended state. The top conformation has an energy of  $-233$  kcal/mol. The middle, the 6th accepted event in the series, has an energy of  $-316$  kcal/mol, and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

vector and a second one,  $|x_1\rangle$ , perpendicular to the first. The full recursion scheme becomes:

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

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  $\mathcal{O}(N)$  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  $L$ ,

$$L = \begin{pmatrix} a_0b_1 & & & \\ b_1a_1b_2 & & & \\ & \ddots & & \\ & & b_{l-1}a_{l-1}b_l & \\ & & & b_la_l \end{pmatrix} \tag{2}$$

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 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.

## APPENDIX B:

### Parallel Tempering

The basic idea behind parallel tempering is to perform several simulations simultaneously on the same system, but at different temperatures.<sup>31</sup> The system that 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 regularly that the system that 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 at various intervals, with a certain probability 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

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

The result is that the higher-temperature simulation helps the lower-temperature one across the energy barriers in the system (see Newman and Barkema<sup>32</sup> for more details).

### ACKNOWLEDGMENT

N.M. and R.M. are grateful to the NSF for partial support under grant number DMR-9805848.

### REFERENCES

- 1 Wildegger, G., Liemann, S., and Glockshuber, R. Extremely rapid folding of the C-terminal domain of the prion protein without kinetic intermediates. *Nat. Struct. Biol.* 1999, **6**, 550–553
- 2 Duan, Y. and Kollman, P. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science* 1999, **282**, 740
- 3 Schlick, T., Barth, E., and Mandziuk, M. Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation. *Ann. Rev. Biophys. Biomol. Struct.* 1997, **26**, 179–220
- 4 Zhou, Y. and Karplus, M. Interpreting the folding kinetics of helical proteins. *Nature* 1999, **401**, 400–403
- 5 Shakhnovich, E.I. Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opinion Struct. Biol.* 1999, **7**, 29–40
- 6 Shea, J-E., Onuchic, J.N. and Brooks, C.L. Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A. *Proc. Natl. Acad. Sci. U.S.A.* 2000, **96**, 12512–12517
- 7 Barkema, G.T., and Mousseau, N. Event-based relaxation of continuous disordered systems. *Phys. Rev. Lett.* 1996, **77**, 4358–4361
- 8 Mousseau, N., and Barkema, G.T. Traveling through potential energy landscapes of disordered materials: The activation-relaxation technique. *Phys. Rev. E* 1998, **57**, 2419
- 9 Mousseau, N., and Lewis, L.J. Topology of amorphous tetrahedral semiconductors on intermediate lengthscales. *Phys. Rev. Lett.* 1997, **78**, 1484–1487
- 10 Mousseau, N., and Lewis, L.J. Structural, electronic and dynamical properties of amorphous gallium arsenide: a comparison between two topological models. *Phys. Rev. B* 1997, **56**, 9461–9468
- 11 Barkema, G.T., and Mousseau, N. Identification of relaxation and diffusion mechanisms in amorphous silicon. *Phys. Rev. Lett.* 1998, **81**, 1865–1868
- 12 Mousseau, N. and Barkema, G.T. Activated mechanisms in amorphous silicon: an activation-relaxation-technique study. *Phys. Rev. B* 2000, **61**, 1898–1906
- 13 Malek, R., and Mousseau, N. Dynamics of Lennard-Jones clusters: A characterization of the activation-relaxation technique. 2000, Preprint cond-mat/0006042
- 14 Mousseau, N. Cooperative motion in Lennard-Jones binary mixtures below the glass transition. 2000, Preprint cond-mat/0004356
- 15 Mousseau, N., Barkema, G.T., and de Leeuw, S.W. Elementary mechanisms governing the dynamics of silica. *J. Chem. Phys.* 2000, **112**, 960–964
- 16 Derreumaux, P. From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential. *J. Chem. Phys.* 1999, **111**, 2301–2310
- 17 Derreumaux, P. Generating ensemble averages for small proteins from extended conformations by Monte Carlo simulations. *Phys. Rev. Lett.* 2000, **85**, 206–209
- 18 Mousseau, N., and Barkema, G.T. Exploring high-dimensional energy landscapes. *Comp. Sci. Eng.* 1999, **1**, 74–82
- 19 Rothman, M.J. and Lohr Jr., L.L. Analysis of an energy minimization method for locating transition-states on potential-energy hypersurfaces. *Chem. Phys. Lett.* 1980, **70**, 405–409
- 20 Cerjan, C.J., and Miller, W.H. On finding transition-states. *J. Chem. Phys.* 1981, **75**, 2800–2806
- 21 Simons, J., Jorgensen, P., Taylor, H., and Ozment, J. Walking on potential-energy surfaces. *J. Phys. Chem.* 1983, **87**, 2745–2753
- 22 Berry, R.S., Davis, H.L., and Beck, T.L. Finding saddles on multidimensional potential surfaces. *Chem. Phys. Lett.* 1988, **147**, 13–17
- 23 Ionova, I.V. and Carter, E.A. Ridge method for finding saddle-points on potential-energy surfaces. *J. Chem. Phys.* 1993, **98**, 6377–6386
- 24 Chekmarev, S.F. A simple gradient-method for locating saddles. *Chem. Phys. Lett.* 1994, **227**, 354–360
- 25 Lanczos, C. *Applied Analysis*. Dover, New York, 1988
- 26 Sankey, Otto F., Drabold, David A., and Gibson, Andrew. Projected random vectors and the recursion method in the electronic-structure problem. *Phys. Rev. B* 1994, **50**, 1376–1381
- 27 Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. *Numerical Recipes*. Cambridge University Press, Cambridge, 1988
- 28 Doye, J.P.K., and Wales, D.J. Surveying a potential energy surface by eigenvector-following. *Z. Phys. D* 1997, **40**, 194–197
- 29 Fezoui, Y., Connolly, P.J., and Osterhout, J.J. Solution Structure of alpha-t-alpha, a Helical Hairpin Peptide of De Novo Design. *Prot. Sci.* 1997, **6**, 1869–1877
- 30 Derreumaux, P. Predicting helical hairpins from se-



- quences by Monte Carlo simulations, *J. Comput. Chem.* 2000, **21**, 582–589
- 31 Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. Funnels, pathways, and the energy landscape of protein-folding - a synthesis. *Proteins* 1995, **21**, 167–195
  - 32 Taketomi, H., Ueda, Y., and Go, N. Studies on protein folding, unfolding and fluctuations by computer simulation. *Int. J. Pept. Protein Res.* 1975, **7**, 445–459
  - 33 Marinari, E., and Parisi, G. Simulated tempering - a new Monte-Carlo scheme. *Europhys. Lett.* 1992, **19**, 451–458
  - 34 Newman, M.E.J., and Barkema, G.T. Monte Carlo methods in statistical physics. Oxford University Press, Oxford, 1999.