

ARTIST: An Activated Method in Internal Coordinate Space for Sampling Protein Energy Landscapes

Mi-Ran Yun,¹ R. Lavery,¹ N. Mousseau,^{1,2} K. Zakrzewska,¹ and P. Derreumaux^{1*}

¹Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico-Chimique et Université Paris, France

²Département de Physique et Centre de bioinformatique Robert-Cedergren, Université de Montréal, Québec, Canada

ABSTRACT We present the first applications of an activated method in internal coordinate space for sampling all-atom protein conformations, the activation–relaxation technique for internal coordinate space trajectories (ARTIST). This method differs from all previous internal coordinate-based studies aimed at folding or refining protein structures in that conformational changes result from identifying and crossing well-defined saddle points connecting energy minima. Our simulations of four model proteins containing between 4 and 47 amino acids indicate that this method is efficient for exploring conformational space in both sparsely and densely packed environments, and offers new perspectives for applications ranging from computer-aided drug design to supramolecular assembly. *Proteins* 2006;63:967–975. © 2006 Wiley-Liss, Inc.

Key words: ARTIST; protein energy landscapes

INTRODUCTION

Exploring energy landscapes is a major challenge for a number of problems in physics, chemistry, and biology. In the field of structural biology, these problems include protein dynamics, folding, and aggregation, often implying processes that occur on much longer time scale than thermal vibrations. For the last 30 years, conventional molecular dynamics (MD) in Cartesian coordinate space has provided a wealth of information on conformational changes occurring within the nanosecond time scale.¹ This is not sufficient, however, to provide a complete picture of the problems cited above, and two main directions have been followed to accelerate conformational sampling and increase our understanding of the structure, thermodynamics, and kinetics.

The first direction involves the development of more efficient Cartesian coordinate-based sampling techniques such as replica exchange² and ensemble dynamics.³ Replica exchange Monte Carlo (MC) or MD simulations are based on exchanges between several copies, or replicas, of the same system simulated in parallel at different temperatures. The effect of these exchanges, controlled by a Metropolis acceptance/rejection criterion, is to enable the room temperature simulations to escape from local minima. Ensemble dynamics, which uses worldwide distributed computing, generates several tens or hundreds of independent series of coupled MD trajectories at the same temperature, typically lasting tens of nanoseconds. These trajec-

tries can be started from the same point using different starting velocities, or from an ensemble of structures. All trajectories are restarted when an energy barrier is crossed in one of the simulations and a selection of the most promising trajectories is often performed at this point.

The second direction is to reduce the number of degrees of freedom by moving to internal coordinate space and, notably, restricting movements to torsional degrees of freedom. Focusing on slower modes, this procedure can significantly increase the efficiency of each move. This approach has been used in MD simulations,^{4–6} simulated annealing MC-based approaches,^{7–10} normal-mode analysis,^{11,12} and fragment-based assembly methods.¹³ It has been applied to studies of the equilibrium properties of small nucleic acid and protein models, to predicting protein structures from amino acid sequences and to refining NMR and X-ray structures. In the case of protein folding, these studies have revealed that the global optimization of an energy function is more efficient in internal coordinate space than in Cartesian coordinate space. However, for MD simulations, the use of internal coordinates has been hindered by the only moderate improvement in the time steps, which can be used (roughly 4–10 fs, compared to 2 fs using Cartesian coordinates with restrained X-H bond lengths).

It is possible to profit further from internal coordinate space representations by defining conformational moves that take into account the structure of the energy landscape. Following this idea, we propose to combine the use of internal coordinates with a first-order saddle point search method. Activated methods, inspired by transition state theory, sample the energy landscape by identifying well-defined transition states connecting two local minima. Because the activation paths are defined directly on the energy surface, activated methods that have been shown to work on a number of systems, including proteins, described in Cartesian coordinates should also be applicable to systems using internal coordinates. Several activated techniques have been developed in recent years.^{14–16} Here, we propose to test the activation–relaxation tech-

*Correspondence to: Philippe Derreumaux, Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico-Chimique et Université Paris 7, 13 rue Pierre et Marie Curie, 75005 Paris, France. E-mail: philippe.derreumaux@ibpc.fr

Received 18 October 2005; Revised 19 December 2005; Accepted 9 January 2006

Published online 7 March 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20938

nique (ART), which has already been used successfully in Cartesian coordinate studies of protein dynamics.^{17,18} This method has been shown to be the most efficient activated algorithm for exploring the energy landscapes of high-dimensional systems.¹⁹

In the following section, we give a general description of the two programs that have been merged to make internal coordinate searches of protein conformations, namely LIGAND and ART. We then discuss the sampling properties of this approach for four model proteins containing from 4 to 47 amino acids (80–740 atoms). We show that the Activation-Relaxation Technique for Internal coordinate Space Trajectories (ARTIST), using both torsional and valence angles, is an efficient approach for finding the global energy minimum of medium-size proteins. It also performs well for exploring both large and small conformational changes in densely packed environments.

MATERIALS AND METHODS

ARTIST is the result of merging two programs: LIGAND, a general purpose, internal coordinate minimization program, and ART, an algorithm generating activated events directly on the energy surface. Here, we describe these two methods separately before discussing the details of their incorporation into ARTIST.

Ligand

LIGAND is an internal coordinate, energy minimization program that can be applied to a single molecule or a set of interacting molecules. LIGAND grew out of earlier developments at the Laboratoire de Biochimie Théorique, notably the programs CINFLEX²⁰ and JUMNA.²¹ It is distinct from these two approaches by respectively using a simple minimizer rather than a constrained minimizer and by using pure internal coordinates, rather than a hybrid internal/helicoidal representation. LIGAND has been applied to a variety of structural studies on both proteins and nucleic acids^{11,22,23} and has also been extended to normal mode calculations,^{11,12} and to collective-variable MC²⁴ and multicopy MC simulations.²⁵ By adopting internal coordinates (limited to single bond torsions and certain valence angles as discussed below) rather than Cartesian coordinates, LIGAND reduces the number of variables necessary to model conformational freedom by roughly an order of magnitude. This directly affects the conformational landscape, reducing the number of local-energy minima and allowing larger conformational changes. The choice of internal variables also makes it easy to restrict conformational freedom to any chosen subset of variables, allowing one to study easily, for example, the flexibility of a single loop within a protein, or selected side chains interacting with a flexible ligand within a protein pocket. It is important to note, however, that although LIGAND can treat explicit molecules or ions interacting with the solute molecules, it is clear that an energy minimization approach is not adapted to explicit solvent models. Consequently, as discussed below, solvent and counterion electrostatic effects are treated with continuum models.

LIGAND uses an all-atom representation of the molecules treated. Flexibility within each molecule is limited

to dihedral rotations around single bonds and to main-chain valence angles. In the case of the polypeptides treated here, “main chain” implies that N–C α –C', C α –C'–N, and C'–N–C α valence angles are independent degrees of freedom. Any other atoms linked to these main chain atoms are moved in a coupled way by half the rotation that is applied to the moving main chain atoms.²⁰ Intracyclic valence angles are also variable within the proline ring. This ring is rendered flexible by cutting the C δ –N bond (which is reestablished using a quadratic distance restraint) and treating the C α –C δ chain as a linear system. This makes the two valence angles and the three dihedral angles involving the C δ and N atoms dependent. In addition to internal degrees of freedom, a multimolecular system includes intermolecular rotations and translations. These are defined with respect to a chosen reference frame within each molecule; the first molecule in the system being fixed in space to exclude changes in the overall position or orientation of the whole system.

Internal degrees of freedom for each molecule included in the study are analyzed using a utility program (POLYPEP in the case of proteins), which contains information on the structure and composition of all the standard amino acids. It can also be used to prepare polypeptide conformations contained in an existing PDB data file or to create standard conformations (α -helix, β -sheet, etc.). POLYPEP also defines force field data such as atomic charges, atom, bond, and valence angle types, and determines the moving atoms associated with each variable. This information is passed to LIGAND using a standard data file.

Energy calculations can be carried out with either our in-house FLEX force field²⁰ or the AMBER force field^{26,27} using the parm98 parameter set.²⁸ Solvent electrostatic damping can be treated using a dielectric constant, a sigmoidal distance-dependent dielectric function,^{21,29} or with the continuum solvent generalized Born (GB) approach using the Tsui and Case parameter set.³⁰ All the calculations carried out here use the AMBER force field and either the GB solvent representation or the sigmoidal dielectric function. No distance cutoffs are used. Minimization is performed with a quasi-Newton algorithm using analytic first derivatives of the internal and solvent energy contributions with respect to the internal degrees of freedom, and is considered to have converged when the predicted energy gain at the next step falls below 10^{-4} kcal mol⁻¹.¹¹

ART

The activation–relaxation technique (ART) is a method that defines moves directly on the energy hypersurface, bringing a system from one local minimum to another, passing via first order saddle points. Each new minimum located is accepted or rejected as a new starting point on the basis of a Metropolis criterion, allowing the procedure to sample the energy hypersurface and, given sufficient time, to locate the global energy minimum.¹⁴ ART was initially used to study activated mechanisms in a number

of materials including amorphous semiconductors, Lennard-Jones clusters and silica glasses.^{14,15} In its current implementation (ART nouveau),¹⁵ the method was also adapted to systems having high and low-frequency vibrational modes¹⁷ and was used to study the folding pathways of a β -hairpin peptide model as well as multimers of amyloid-forming peptides.^{18,31,32} In all cases, ART in Cartesian coordinate space succeeded in identifying folding mechanisms considerably more complex than those previously proposed. As an example, ART combined with a simplified generic force field (OPEP^{33–34}) revealed that the rate-limiting in the assembly of three peptide chains into a cross- β structure involves the reptation of one strand over another.³¹ Given that the passage from Cartesian to internal coordinates leads to a lower dimensional energy hypersurface, it was expected that the combination of LIGAND and ART could be a powerful tool for locating the stable energy conformations of biological polymers and their complexes.

Each event in the ART nouveau search consists of four steps:^{15,17} (1) leave the present local minimum and find a negative eigenvalue in the Hessian matrix using the Lanczos technique (because only the lowest eigenvalue and its associated eigenvector are required it is not necessary to diagonalize the full Hessian matrix); (2) push the system along the corresponding eigenvector, while minimizing the forces in the perpendicular hyperplane until the total force is close to zero; (3) push the system over the saddle point and minimize the conformation in the new local minimum; (4) apply the Metropolis criterion at the desired temperature to accept or reject the newly generated conformation. Steps (1) and (2) constitute the activation phase, and steps (3) and (4) the relaxation phase.

ARTIST

The coupling of LIGAND with ART leads to ARTIST. Because all the moves are defined directly in the energy landscape, irrespective of the real-space details, the algorithm can easily generate events involving the deformation of a single angle or of all angles together. In contrast to Cartesian coordinate space where displacements of a variable involve only a local deformation, the deformation of a dihedral angle in internal coordinate space can have long-range consequences. This problem is partly controlled with ARTIST as steric hindrance impacts directly on the shape of the energy landscape. We see, however, especially in compact conformations, that a number of trial events are aborted because of atomic collisions during the activation phase. Reducing this number of collisions is possible. Improvements could come from using soft-core Lennard-Jones potentials, adaptive steps or reduced protein representations, such as that adopted by the OPEP force field.³³ All these possibilities are under study.

In what follows, activation and relaxation of all protein models are performed in internal coordinate space using all torsion angles of the backbone and the side chains and the main valence angles along the backbone (see above). Starting conformations of the model systems were gener-

ated using a random deformation to move the system away from its initial energy well. These deformations, unless otherwise indicated, consisted of perturbations, of the order of 5–10°, which were applied to a total of 20 variables randomly chosen from all φ , ψ and χ dihedral angles. The Metropolis temperature was varied between 300 and 600 K (but does not correspond to coupling to a temperature bath because the conformations are minimized and neglect all vibrational entropic contributions).

RESULTS

We test ARTIST on four model proteins: a dodecalanine, the KFFE tetrapeptide, the 47-residue N-terminal domain of RNase HI and a 28-residue α -helical hairpin. These four simple models allow us to examine the sampling properties in both sparse and dense environments.

Dodecalanine

This molecule consists of 12 alanines and 120 atoms. We ran five 500-event simulations (R1–R5) at 600 K, starting from a fully extended structure and using the AMBER force field with the GB solvent representation. As discussed above, the Metropolis acceptance rate at this temperature is 10%. This means that, on average, these runs generate trajectories crossing at most 50 different barriers. Four simulations (R1–R4) reach the same lowest energy conformation, $-191 \text{ kcal mol}^{-1}$, characterized by a full α -helix. The last run, R5, locates another helix with an energy of $-188.7 \text{ kcal mol}^{-1}$ deviating by 0.8 Å RMSD from the global minimum. The trajectories R1 and R4, shown in Figures 1 and 2, are described by the evolution of the secondary structure, energy and end-to-end C α distances (d_{1-12}) as a function of the event number. Note that d_{1-12} for a canonical, 12 residue α -helix is 16.6 Å. Secondary structure composition was computed using the DSSP program³⁵ and representative structures along the trajectories were drawn using MOLMOL program.³⁶

In Run R1, the polypeptide first collapses into a compact random coil loop [$d_{1-12} = 5 \text{ Å}$ at event 26, Fig. 1(A)]. The chain then unfolds and adopts transient helical conformations. At event 80, d_{1-12} is 23 Å and residues 6–9 are helical [Fig. 1(B)]. Subsequently, the short α -helix shifts into a 3_{10} helix and the two extremities come closer again [$d_{1-12} = 12 \text{ Å}$ at event 109 with an energy of $-178 \text{ kcal mol}^{-1}$, Fig. 1(C)]. Finally, in a single accepted event (event 112, the 21st accepted event), the system jumps to a full α -helix [Fig. 1(D)] and its energy decreases by 13 kcal mol⁻¹.

In contrast to the fast folding trajectory R1, R4 provides insights into a slower folding mechanism (the ground state being obtained at event 292, the 29th accepted event). As in run R1, the peptide first folds into a compact loop [centered on residues 5–7 at event 157, Fig. 2(A)]. It then unfolds, but from here the peptide explores a much wider distribution of helices: mixing an α -helix (residues 6–10) and a 3_{10} -helix (residues 2–4) at event 170 [Fig. 2(B)], maintaining only an α -helix (residues 2–5) at event 179 [Fig. 2(C)], adopting a

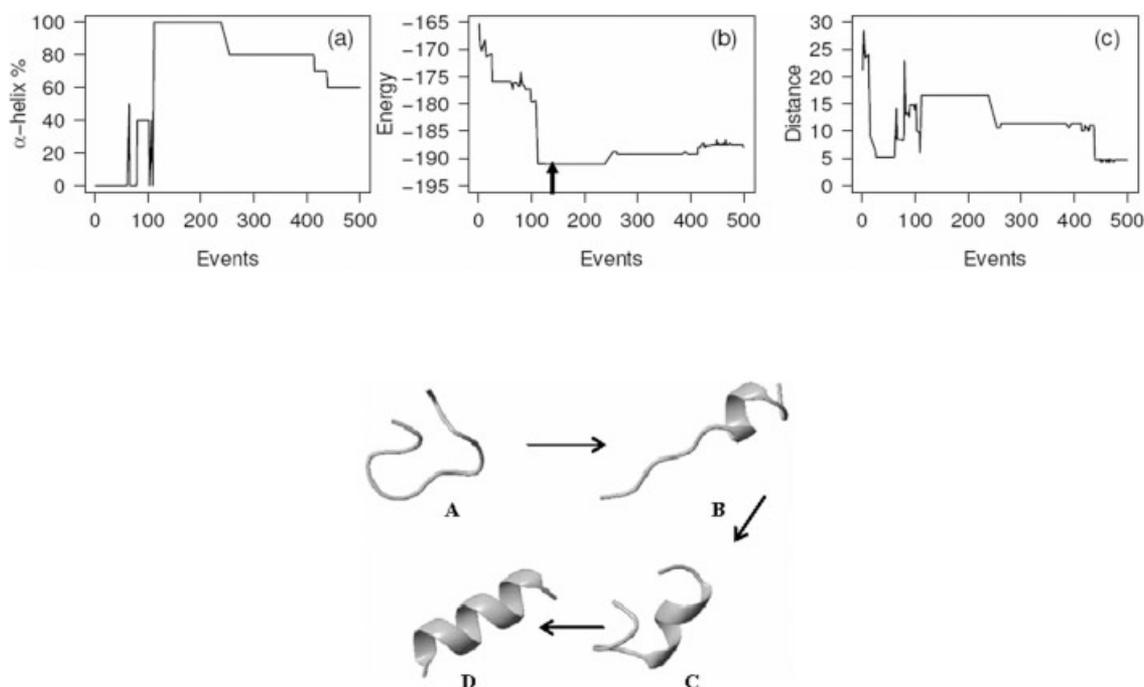


Fig. 1. Folding trajectory R1 of a dodecaalanine fragment starting from a fully extended state at 600 K. α -Helix content, energy (in kcal mol^{-1}) and end-to-end distance d_{1-12} (in \AA) as a function of event numbers (including accepted and rejected events). The ground state is first reached at event 112 (indicated by an arrow in plot **b**).

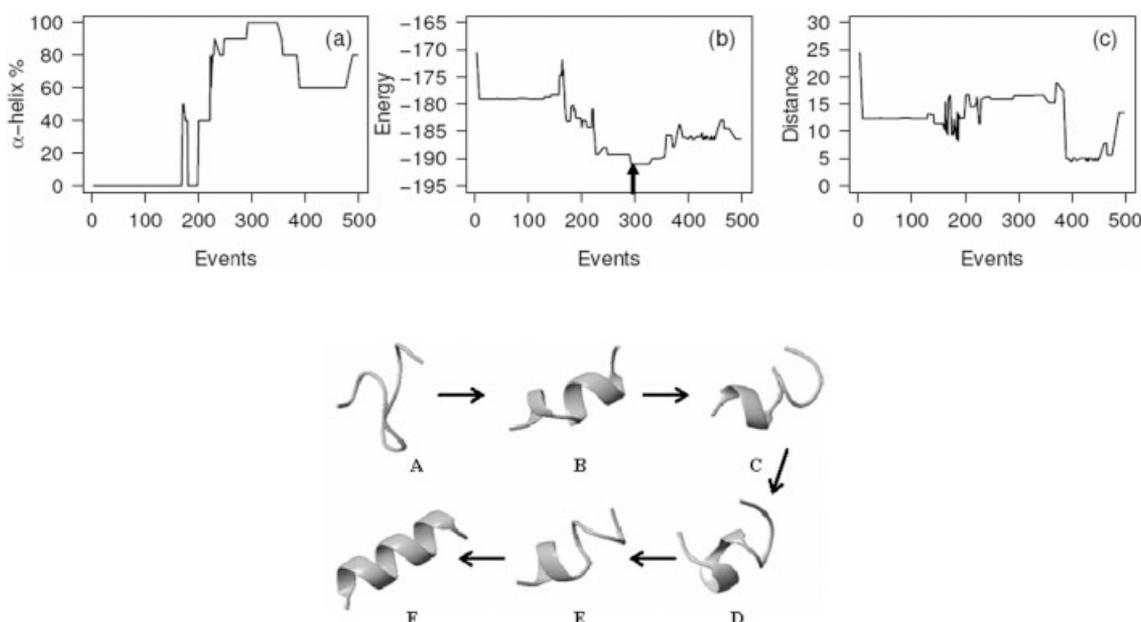


Fig. 2. Folding trajectory R4 of a dodecaalanine fragment starting from a fully extended state at 600 K. The ground state is first reached at event 292 (indicated by an arrow in plot **b**).

π -helix (residues 3–8) at event 187 [Fig. 2(D)], and again an α -helix (residues 2–5) at event 200 [Fig. 2(E)]. Finally, this short helical fragment propagates in both directions and the ground state is reached at event 292.

We compare these results with three 500-event runs at 300 K (S1–S3), starting from the same fully extended state. At this temperature, the acceptance rate is 5%. Run

S1 locates the same global minimum found by runs R1–R4, while runs S2 and S3 locate α -helices of $-189.5 \text{ kcal mol}^{-1}$, deviating by 0.8 \AA RMSD from the ground state. Because S2 and S3 correspond to trajectories over at most 25 different barriers, it is likely that longer simulations would bring them into the ground state. Figure 3 gives the geometric and energetic details of the run S1.

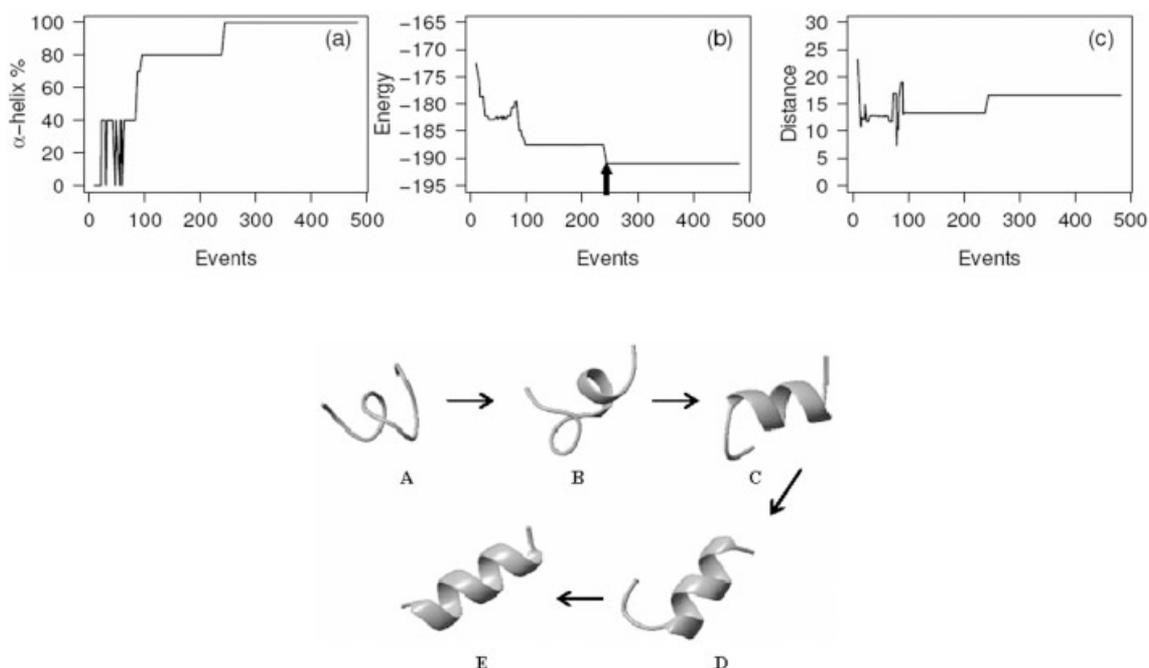


Fig. 3. Folding to a full α -helix starting from a fully extended state at 300 K. The ground state is first reached at event 244 (indicated by an arrow in plot b).

The formation of an α -helix in run S1 starts with the formation of random coil states with a d_{1-12} of 11 Å [Fig. 3(A)]. A helical fragment spanning residues 7–10 then forms at event 63 [Fig. 3(B)], and progressively propagates toward the N-terminus. At event 91, the α -helix involves residues 4–10 [Fig. 3(C)], while, at event 175, the helix spans residues 4–11 [Fig. 3(D)]. The lowest energy state is reached at event 244 (accepted event 23). Analysis of runs S2 and S3 indicates that folding can involve propagation of a helical nucleus either from the N- to the C-terminal or from the center of the chain to its extremities.

Taken together, these simulations underline the diversity of possible folding routes leading to a full α -helix. It is interesting to note that these mechanisms found using ARTIST resemble those found with earlier Cartesian-based simulations on related models. Intermediates with 3_{10} -helices and π -helices were found during all-atom MD simulations of (AAQAA)₃ in implicit solvent,³⁷ and direct transitions between conformations with small helical content and full α -helices, as shown in Figure 1(C), were also found by ensemble dynamics simulations of a 21-alanine peptide in explicit solvent.³ These results also underline the improved efficiency of moving in internal coordinate space, the trajectories converging generally to the ground state of dodecalanine *in less* than 30 accepted ARTIST steps.

KFFE

It is important to determine the impact of long-side chains on the global optimization of the energy function. To address this issue, we examine the KFFE tetrapeptide (80 atoms), which is expected to be disordered in solution³²

and has been shown to form amyloid fibrils *in vitro*.³⁸ We perform 36 runs from a fully extended conformation and 8 runs starting from randomly chosen conformations, for 2000 events at 600 K using the AMBER force field. For these more extensive simulations we use the less costly sigmoidal dielectric function representation of solvent electrostatic damping. Among 44 runs, we find that six runs locate the same lowest energy random coil state with an energy of -97.4 kcal mol⁻¹, 13 runs deviate from the global minimum by an all-atom RMSD of 1.3 Å and an energy difference of 1.7 kcal mol⁻¹, while the remaining 25 runs locate conformations in the -92 to -94 kcal mol⁻¹ energy range. Figure 4 shows the superposition of one higher energy structure (-95.7 kcal mol⁻¹) on the global minimum. As seen, all backbone and side chain atoms of the phenylalanine and glutamic acid residues superpose very well. The structural deviation results from different states of the χ_3 and χ_4 torsional angles of the lysine residue. This simple model illustrates the difficulty of escaping from local minima in the case of polypeptides with long side chains, even when using internal coordinate space and high temperature sampling.

N-Terminal Domain of RNase HI

While the previous simulations focus on convergence to the lowest energy conformation and transition paths from extended conformations to the ground state, we now examine the sampling properties of the algorithm in a densely packed environment. Such conformational flexibility around the native state is of interest for *in silico* drug design experiments because deformations of the receptor protein and the drug are often observed during binding. It

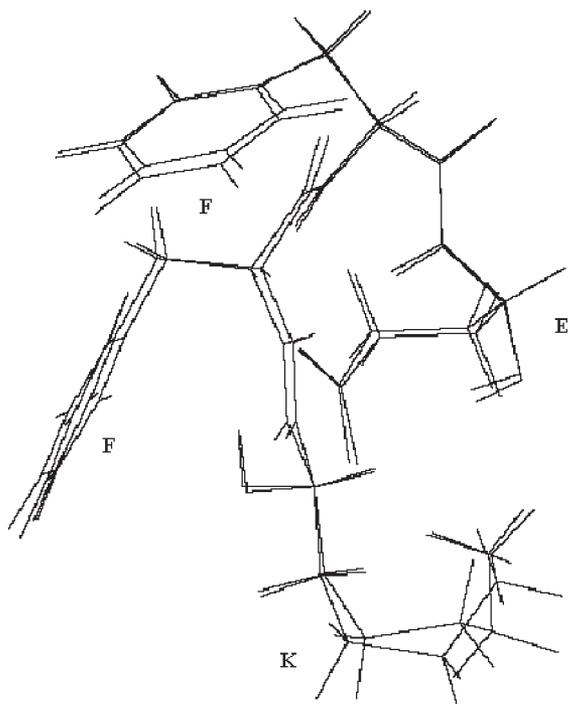


Fig. 4. Superposition of a higher energy structure on the ground state of KKFE. Structural deviation results from the χ_3 and χ_4 torsional angles of the lysine residue: (t,g⁻) vs. (g⁺,g⁺).

has indeed been recognized that a rigid treatment of the protein backbone is a severe limitation in current docking algorithms.^{39,40} To investigate this problem, we study the deformation of the N-terminal domain of RNase HI (PDB entry 1QHK⁴¹) consisting of 47 residues (numbers 6–52) and 740 atoms starting from the native structure. The conformation of this fragment has an α/β type topology with three β -strands (residues 7–11, 18–22, and 39–41) and two α -helices (residues 23–29 and 44–51). The fragment is first minimized and then subjected to five short ARTIST simulations of 100 events at 600 K, using the AMBER force field and the sigmoidal dielectric function. At this temperature, the Metropolis acceptance rate is 8%.

Figure 5 shows the evolution of the N-terminal domain structure as a function of the accepted event numbers for run R2. The results are further analyzed in Table I, which gives the C α RMSD between all accepted structures, and in Figure 6, which presents the changes in the φ and ψ dihedral angles and the C α positions with respect to the initial minimized structure. Very similar structures are generated in the remaining four runs.

The first accepted structure (ST1) deviates by 1.6 Å RMSD from the minimized experimental structure (ST0). The second and third accepted structures explore distinct conformations within the native basin: the RMSD between ST2 and ST3 being 1.8 Å, while ST3 and ST2 are respectively 1.6 and 1.0 Å from ST0. From here, ARTIST is able to move the protein to a new basin of attraction. ST5 deviates by 3.1–3.9 Å RMSD from all previous structures and is characterized by the disruption of one β -strand allowing a significant displacement of the C-terminal

helix. This conformational change is further propagated in ST8, which deviates by 2.6 Å RMSD from ST5 and 4.6–5.5 Å RMSD from ST0–ST3, and is characterized by the reorientation of the first helix with respect to the N-terminal region.

The changes in the backbone dihedral angles between ST0 and ST1, and ST0 and ST8 are shown in Figure 6. ST1 results essentially from 90° changes for a few backbone torsions of the residues 32–42 [Fig. 6(A)], leading to a displacement of 4 Å of the C-terminal residues [Fig. 6(B)]. However, torsional changes in fact occur collectively throughout the fragment with the φ/ψ values of most amino acids being modified by at least 5–10° [Fig. 6(A)]. The final ST8 structure has many residues that have been subjected to large-amplitude torsional angle changes [Fig. 6(C)] and have moved by 6–8 Å from their positions in ST0 [Fig. 6(D)].

We can compare these internal coordinate-based results with the structures generated using ART-OPEP in Cartesian coordinate space, although the latter have been obtained using the simplified OPEP force field.³³ Several 50-event runs were attempted at 600 K and all generated conformations which differed by only 1.0–1.3 Å RMSD from the ST0 structure. This comparison, although biased by the use of two different force fields (ART and ARTIST cannot use the same force field yet), nevertheless demonstrates the improved efficiency of locating transition states in internal coordinate space.

Two-Domain Protein Model

Finally, we investigate the behavior of the algorithm with respect to the conformational sampling of two protein domains connected by a linker. This is an important problem because there is both experimental and theoretical evidence that multidomain proteins are not rigidly packed, but are rather dynamic structures.^{42,43} For these initial tests, we use a helical-hairpin model derived from the PDB entry 1ABZ⁴⁴ and consisting of two α -helices (residues 6–16 and 22–33) connected by a loop (residues 17–21). This system contains 399 atoms and has 242 internal coordinates. We will look only at the diversity of structures that can be generated by a *single accepted* activation-relaxation event.

Because we are mainly interested in the relative motion of the two helices, only the φ and ψ torsional angles of the loop residues and the two adjacent residues at either end (residues 15–23) are subjected to random displacements in step (1) of the activation procedure. Figure 7 shows the first accepted structures from six single event runs at 600 K, starting from the minimized experimental structure (R0) and using different random seeds. ARTIST is seen to explore a wide range of α -hairpins differing in helical content and relative helical orientation. In most structures both helices are shortened by two to three residues (e.g., helix 1 in runs R1 and R2 spans residues 6–13, while in R5 it spans residues 8–15). Our procedure for leaving a local minimum does not, however, preclude the destabilization of a helix and, in run R4, helix 1 is converted into a folded substructure with adjacent 3_{10} - (residues 7–9) and α -heli-

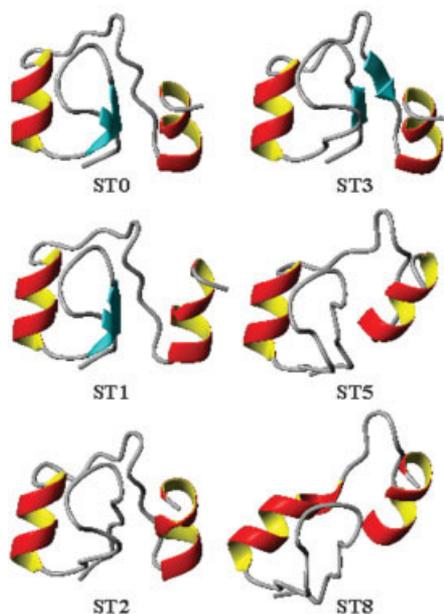


Fig. 5. Evolution of the N-terminal domain of RNase HI as a function of the accepted event numbers.

TABLE I. The C α RMSD (in Å) between All Generated Structures of the N-Terminal Domain of RNase HI

RMSD (C α)	ST0	ST1	ST2	ST3	ST5	ST8
ST0	0.00					
ST1	1.63	0.00				
ST2	1.66	1.64	0.00			
ST3	1.02	2.05	1.76	0.00		
ST5	3.72	3.47	3.14	3.92	0.00	
ST8	5.36	5.04	4.61	5.50	2.61	0.00

cal (residues 10–13) fragments. Overall, the angle between the optimal axes of two the helices varies between 180° (antiparallel, in R0 and R7), 90° (perpendicular, in R2) and 0° (parallel, in R6).

As seen in Figure 8, in a single activation–relaxation event ARTIST can generate moves of varying complexity, associated with very different distributions of backbone dihedral angle changes. In addition, flexibility is not only limited to the backbone, but also takes place within the side chains. Analysis of all generated structures shows that, on average, 15% of all the side chain torsional angles are modified by more than 20° in a single ARTIST event and the maximum angular change reaches 90°.

The diversity of the structures that can be generated by a single activation–relaxation event is independent of the details of the force field used. Using AMBER with GB solvent representation, or AMBER with the sigmoidal dielectric function, ARTIST captures very similar minima with interhelical angles varying between 180° and 0°.

CONCLUSIONS

This is the first report, to the best of our knowledge, of using an activated search method in internal coordinate

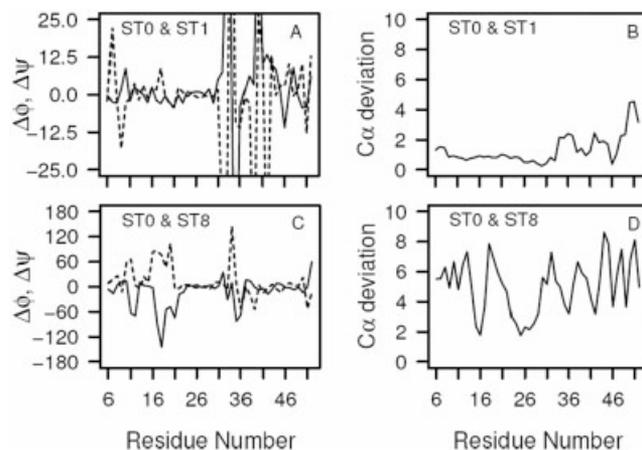


Fig. 6. Structural analysis of the ST1 and ST8 structures of the N-terminal domain of RNase HI. Changes in the φ and ψ dihedral angles (in degrees) of ST1 (A) and ST8 (C) as a function of the residue numbers with respect to the minimized ST0 structure. $\Delta\psi$ angles in dotted lines. Changes in the C α positions (in Å) between the ST1 (ST8) and ST0 structures using a best-fit superposition procedure (B, D).

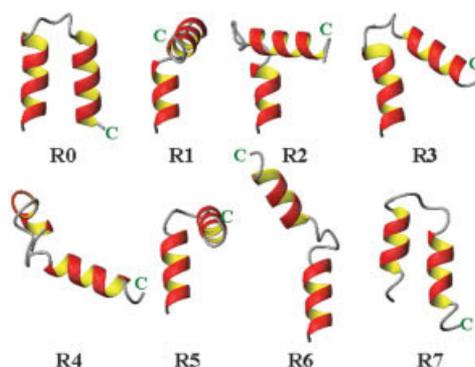


Fig. 7. Ensemble of generated models after a single activation–relaxation event starting from the minimized structure of a two-domain protein model. In each model, the C-terminal end is indicated.

space for exploring protein energy surfaces. This work differs from previous studies using internal coordinates in that our moves are not encoded in a predefined list of perturbations, but are generated automatically in conformational space via the activation–relaxation procedure, leading to moves of any complexity and size. This algorithm also differs from double-ended saddle point techniques^{45,46} in that the next local minimum is unknown during each attempted event. ARTIST is also relatively fast, the averaged CPU time on a 2.1 GHz microprocessor is (1.5, 8) min for generating a single activation–relaxation event on the protein models studied here with (399, 740) atoms using AMBER and the GB solvent representation and no distance cutoffs of the nonbonded interactions.

We have demonstrated that this algorithm, the activation–relaxation technique for internal coordinate space trajectories (ARTIST), is able to deal with a wide range of sampling problems in both sparsely and densely packed environments. These include (1) finding folding paths from an unfolded state to the ground state; (2) generating

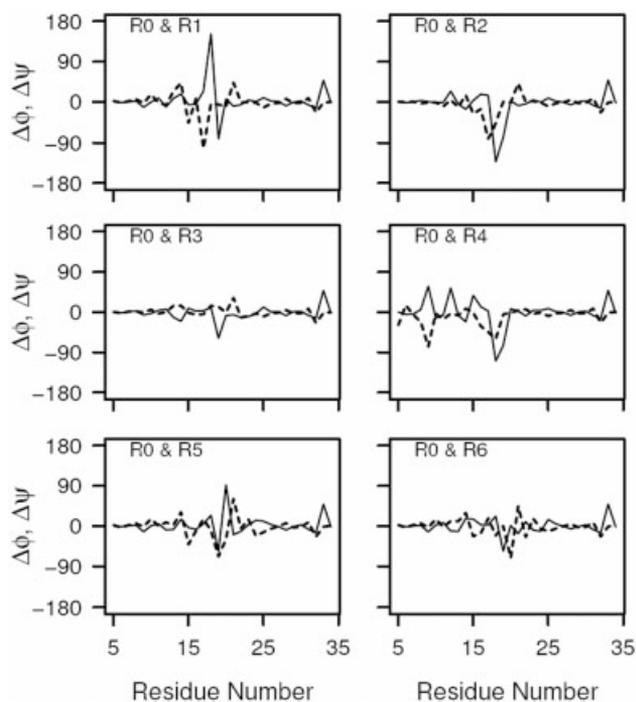


Fig. 8. Changes in the ϕ and ψ dihedral angles (in degrees) of the generated models from the initial R0 structure for a two-domain protein model. $\Delta\psi$ angles in dotted lines.

native-like conformations characterized by the simultaneous motions of the loops, secondary structures, and side chains starting from a compact native protein conformation; and (3) identifying the transition state ensemble representing the motion of one domain with respect to another. Whether all these motions indeed occur at room temperature *in vitro* cannot be determined because of uncertainties in the force field used, but, within the limits of our representation, we demonstrate that they correspond to physically and mechanistically feasible transitions between local minima over well-defined first-order saddle points.

In line with earlier works, these simulations confirm that energy landscape sampling is more efficient in internal coordinate space than in Cartesian coordinate space despite a rather low Metropolis acceptance rate. Between 300 and 600 K, only 5 to 10% of ARTIST events are accepted. This is considerably lower than the rate found in Cartesian coordinate space for the same temperature range (25–40%). This difference is largely due to the fact that ARTIST's moves are less local in nature, involving larger displacements, and is a direct consequence of the higher sampling efficiency of ARTIST. Although there is still room for improvement in the activation phase, the present results clearly demonstrate that this method holds considerable promise for a variety of protein flexibility, folding, docking and assembly problems.

Contributions

Richard Lavery and Krystina Zakrzewska are the principal authors of LIGAND. Normand Mousseau is one of the

authors of ART in Cartesian space. Mi-Ran Yun and Philippe Derreumaux merged both programs and adapted ART for internal coordinate space sampling.

ACKNOWLEDGMENTS

R.L. and K.Z. would like to thank Karine Bastard, Fabien Cailliez, Catherine Etchebest, Isabelle Navizet, Chantal Prévost, and Cyril Rebol for their contributions to the development of the LIGAND and its associated utility programs. N.M. is supported in part by grants from NSERC (Canada) and the Canada Research Chair program. P.D. thanks the IDRIS and CINES centers in France for their generous computational support.

REFERENCES

- Karplus M, Kuriyan J. Molecular dynamics and protein function. *Proc Natl Acad Sci USA* 2005;102:6679–6685.
- Lin CY, Hu CK, Hansmann, UH. Parallel tempering simulations of HP-36. *Proteins* 2003;52:436–445.
- Sorin EJ, Pande VS. Exploring the helix-coil transition via all-atom equilibrium ensemble simulation. *Biophys J* 2005;88:2472–2493.
- Schwieters CD, Clore GM. Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J Magn Reson* 2001;152:288–302.
- Mazur AK. Titration in silico of reversible B-A transition in DNA. *J Am Chem Soc* 2003;125:7849–7859.
- Chen J, Im W, Brooks CL III. Application of torsion angle molecular dynamics for efficient sampling of protein conformations. *J Comput Chem* 2005;26:1565–1578.
- Ripoll DR, Vila JA, Scheraga HA. Folding of the villin headpiece subdomain from random structures. Analysis of the charge distribution as a function of pH. *J Mol Biol* 2004;339:915–925.
- Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol* 1994;235:983–1002.
- Srinivasan R, Rose GD. Ab initio prediction of protein structure using LINUS. *Proteins* 2002;47:489–495.
- Liu Y, Beveridge DL. Exploratory studies of ab initio protein structure prediction: copy simulated annealing, AMBER energy functions, and a generalized Born/solvent accessibility solvation model. *Proteins* 2002;46:128–146.
- Duong TH, Zakrzewska K. Influence of drug binding on DNA flexibility: a normal mode analysis. *J Biomol Struct Dyn* 1997;14:691–701.
- Duong TH, Zakrzewska K. Sequence specificity of bacteriophage 434 repressor-operator complexation. *J Mol Biol* 1998;280:31–39.
- Simons KT, Strauss C, Baker D. Prospects for ab initio protein structural genomics. *J Mol Biol* 2001;306:1191–1199.
- Barkema GT, Mousseau N. Event-based relaxation of continuous disordered systems. *Phys Rev Lett* 1996;77:4358–4361.
- Malek R, Mousseau N. Dynamics of lennard-jones clusters: a characterization of the activation-relaxation technique. *Phys Rev E* 2000;62:7723–7728.
- Doyle JPK, Wales DJ. Surveying a potential energy surface by eigenvector-following. *Z Phys D* 1997;40:194–197.
- Wei G, Mousseau N, Derreumaux P. Exploring the energy landscape of proteins: a characterization of the activation-relaxation technique. *J Chem Phys* 2002;117:11379–11387.
- Santini S, Wei G, Mousseau N, Derreumaux P. Pathway complexity of Alzheimer's beta-amyloid Aβ16–22 peptide assembly. *Structure* 2004;12:1245–1255.
- Olsen RA, Kroes GJ, Henkelman G, Arnaldsson A, Jonsson H. Comparison of methods for finding saddle points without knowledge of the final states. *J Chem Phys* 2004;121:9776–9792.
- Lavery R, Parker I, Kendrick J. A general approach to the optimization of the conformation of ring molecules with an application to valinomycin. *J Biomol Struct Dyn* 1986;4:443–462.
- Lavery R, Zakrzewska K, Sklenar H. JUMNA (Junction Minimization of Nucleic-Acids). *Comput Phys Commun* 1995;91:135–158.
- Rohs R, Etchebest C, Lavery R. Unraveling proteins: a molecular mechanics study. *Biophys J* 1999;76:2760–2768.

23. Navizet I, Cailliez F, Lavery R. Probing protein mechanics: residue-level properties and their use in defining domains. *Biophys J* 2004;87:1426–1435.
24. Gabb HA, Prevost C, Bertucat G, Robert CH, Lavery R. Collective-variable Monte Carlo simulation of DNA. *J Comput Chem* 1997;18:2001–2011.
25. Bastard K, Thureau A, Lavery R, Prevost C. Docking macromolecules with flexible segments. *J Comput Chem* 2003;24:1910–1920.
26. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE III, DeBolt S, Ferguson D, Seibel GL, Kollman PA. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp Phys Commun* 1995;91:1–41.
27. Case DA, Pearlman DA, Caldwell JW, Cheatham TE III, Wang J, Ross WS, Simmerling SC, Darden TA, Mer KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Tsui V, Gohlke H, Radmer RJ, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC, Weimer PK, Kollman PA. AMBER7. <http://its.unc.edu/hpc/applications/science/amber/index.html>. University of California, San Francisco; 2002.
28. Cheatham TE 3rd, Cieplak P, Kollman PA. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J Biomol Struct Dyn* 1999;16:845–862.
29. Hingerty B, Richie RH, Ferrel TL, Turner JE. Dielectric effects in biopolymers: the theory of ionic saturation revisited. *Biopolymers* 1985;24:427–439.
30. Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* 2000;56:275–291.
31. Santini S, Mousseau N, Derreumaux P. In silico assembly of Alzheimer's A β 16–22 peptide into beta-sheets. *J Am Chem Soc* 2004;126:11509–11516.
32. Wei G, Mousseau N, Derreumaux P. Sampling the self-assembly pathways of KFFE hexamers. *Biophys J* 2004;87:3648–3656.
33. Derreumaux P. From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential. *J Chem Phys* 1999;111:2301–2310.
34. Derreumaux P. Generating ensemble averages for small proteins from extended conformations by Monte Carlo simulations. *Phys Rev Lett* 2000;85:206–209.
35. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22:2577–2637.
36. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structure. *J Mol Graphics* 1996;14:51.
37. Ferrara P, Apostolakis J, Caffisch A. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J Phys Chem B* 2000;104:5000–5010.
38. Tjernberg L, Hosia W, Bark N, Thyberg J, Johansson J. Charge attraction and beta propensity are necessary for amyloid fibril formation from tetrapeptides. *J Biol Chem* 2002; 277:43243–43246.
39. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* 2002;47:409–443.
40. May A, Zacharias M. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim Biophys Acta* 2005 Oct5, Epub ahead of print.
41. Evans SP, Bycroft M. NMR structure of the N-terminal domain of *Saccharomyces cerevisiae* RNase HI reveals a fold with a strong resemblance to the N-terminal domain of ribosomal protein L9. *J Mol Biol* 1999; 291:661–669.
42. Bertini I, Del Bianco C, Gelis I, Katsaros N, Luchinat C, Parigi G, Peana M, Provenzani A, Zoroddu MA. Experimentally exploring the conformational space samples by domain reorientation in calmodulin. *Proc Natl Acad Sci USA* 2004;101:6841–6846.
43. Kim MK, Jernigan RL, Chirikjian GS. Rigid-cluster models of conformational transitions in macromolecular machines and assemblies. *Biophys J*. 2005;89:43–55.
44. Fezoui Y, Connolly PJ, Osterhout JJ. Solution structure of alpha t alpha, a helical hairpin peptide of de novo design. *Protein Sci* 1997;6:1869–1877.
45. Fischer S, Karplus M. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem Phys Lett* 1992;194: 252–261.
46. Czerminski R, Elber R. Reaction path study of conformational transitions in flexible systems: application to peptides. *J Chem Phys* 1992;92:5580–5601.