JCTC Journal of Chemical Theory and Computation

Combining Elastic Network Analysis and Molecular Dynamics Simulations by Hamiltonian Replica Exchange

Martin Zacharias*

School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, D-28759 Bremen, Germany

Received September 4, 2007

Abstract: Coarse-grained elastic network models (ENM) of proteins can be used efficiently to explore the global mobility of a protein around a reference structure. A new Hamiltonian-replica exchange molecular dynamics (H-RexMD) method has been designed that effectively combines information extracted from an ENM analysis with atomic-resolution MD simulations. The ENM analysis is used to construct a distance-dependent penalty (flooding or biasing) potential that can drive the structure away from its current conformation in directions compatible with the ENM model. Various levels of the penalty or biasing potential are added to the force field description of the MD simulation along the replica coordinate. One replica runs at the original force field. By focusing the penalty potential on the relevant soft degrees of freedom the method avoids the rapid increase of the replica number with increasing system size to cover a desired temperature range in conventional (temperature) RexMD simulations. The application to domain motions in lysozyme of bacteriophage T4 and to peptide folding indicates significantly improved conformational sampling compared to conventional MD simulations.

Introduction

The limited time scale accessible during conventional classical molecular dynamics (MD) simulations is a major bottleneck to sample relevant conformational states of biomolecules. Conformational transitions between stable states of a biomolecules occur only rarely even on the time scale of tens to hundreds of nanoseconds that are currently possible.[1−12] Parallel tempering or replica exchange molecular dynamics (RexMD) simulations are now frequently used to enhance conformational sampling in Monte Carlo (MC)[13−16] and MD simulations.[5−8,10−12,17−24] During RexMD simulations several copies or replicas of a given system are simulated in parallel using classical MD or MC methods at different simulation temperatures. At preset intervals pairs of replicas (usually neighboring pairs) are exchanged with a specified transition probability. Instead of changing the temperature among the replicas it is also possible to modify the force field or selected force field contributions along the replicas (termed Hamiltonian-RexMD).[24] The exchanges allow conformations trapped in locally stable states (e.g., at a low simulation temperature) to escape by exchanging with replicas at higher simulation temperature (or running with a modified Hamiltonian). The RexMD method has been successfully applied in folding simulations of peptides and miniproteins[5−8,10−12,17−22] as well as for the folding of nucleic acid structural motifs.[25] Unfortunately, efficient exchange between replicas requires sufficient overlap of the energy distributions between neighboring replicas. As a consequence, in order to cover a desired temperature range, the number of required replicas grows approximately with the square root of the number of particles in the system.[24] A larger number of replicas in turn requires also increased simulation times for efficient "travelling" of replicas in the range of different temperatures.

Hybrid explicit/implicit solvent models have been suggested where the simulation of each replica is performed using an explicit solvent description and for each exchange part of the solvent is replaced by a continuum.[26] Another approach employs separate coupling of solute and solvent to different heat baths (target temperatures).[27] Only the solute reference temperatures are varied for each replica. In a further

* Corresponding author phone: ++49-421-200-3541; fax: ++49-421-200-3249; e-mail: m.zacharias@jacobs-university.de.

10.1021/ct7002258 CCC: $40.75
Published on Web 02/21/2008

extension of this approach the temperature of only selected collective degrees of freedom has been modified along a replica coordinate.[28] These methods reduce the effective system size compared at each attempted replica exchange. However, the artificial temperature gradient at the solute−solvent interface and the inclusion of nonphysical systems as replica runs may cause artefacts in the latter methods.

Alternatively, approaches that scale the Hamiltonian or energy function of the system along a replica-coordinate have been suggested.[24,29−33] Recently, a promising "Hamiltonian"-RexMD method has been developed where the solute−solute, solute−solvent, and solvent−solvent interactions are separately (linearly) scaled for each replica.[31] This approach can be used to "effectively" scale only the solute temperature along the replica coordinate. In case of no scaling of the solvent−solvent interactions the replica exchange probability becomes less dependent on the number of solvent degrees of freedom, and hence fewer replicas are required to cover a desired "effective" temperature range compared to standard temperature replica exchange. A similar approach where the nonbonded (Lennard-Jones and electrostatic) interactions within the solute as well as between solute and solvent have been scaled to various degrees has also been suggested.[32] Another method specifically designed for peptides and proteins employs a biasing potential for the peptide backbone to specifically lower the barriers for backbone dihedral transitions as replica coordinate.[33] The biasing potential is obtained from explicit solvent simulations of a model peptide. The method showed promising results during peptide folding simulations.[33]

In recent years it has been shown that soft normal modes obtained from Elastic network models (ENMs) of proteins frequently overlap with experimentally observed conformational changes in proteins.[34−39] In an Elastic network model a given structure of a protein serves as a reference structure, and the mobility of a residue or protein segment depends on the local density and number of short-range contacts (usually between Cα or heavy atoms of the protein). Collective degrees of freedom can be calculated very rapidly from an ENM model of a protein (within seconds on standard workstation computers) by a normal mode calculation (after diagonalization of the second derivative matrix). Due to the goarse-grained nature, ENMs may indicate directions of possible large scale conformational transitions of a biomolecule. In fact, often very significant overlap of the softest ENM modes with experimentally observed conformational changes in proteins has been found.[37] This has, for example, been explored in efficient flexible docking simulations of proteins.[40,41]

The idea of coupling ENM analysis of proteins and MD simulations has been explored by Zhang et al.[42] by separate temperature coupling of collective ENM degrees of freedom of a molecule and temperature control of the rest of the system in a single simulation. The motion along ENM degrees of freedom is amplified by increasing the temperature "along" the collective degrees of freedom of the molecule. The method allowed enhanced sampling of peptide and protein motion.[42] However, separate temperature coupling of different degrees of freedom corresponds to a nonphysical simulation system, and it is not clear if such a simulation produces conformations compatible with the desired simulation ensemble (e.g., a canonical ensemble). In addition, extended simulation runs with an increased temperature of the soft collective degrees of freedom of a system may lead to sampling of undesired conformations, e.g., unfolding of the protein (if the temperature of a collective degree of freedom is kept above the folding temperature of the protein).

In the present study an alternative "Hamiltonian" replica-exchange method is proposed that includes a biasing or flooding potential compatible with an ENM description of the protein/peptide as a replica coordinate. The purpose of the biasing or penalty potential is to drive the protein or peptide conformation away from its current state in directions compatible with the ENM model of the system. Penalty potentials with the purpose to drive structures away from a given state have already been used in conformational flooding[43] and metadynamics simulations.[44] However, the coupling of ENM derived penalty/biasing potentials and replica exchange simulations has not been tried. The level of biasing is gradually changed along the replicas (one replica runs at the original force field) such that frequent transitions are possible. Since the overall conformation of the biomolecule may change during the simulation, the ENM calculations can be repeated at preset intervals. Since exchanges between replicas depend only on different levels of a very soft potential, the method requires fewer replicas for efficient sampling compared to conventional temperature RexMD. As long as the ENM model is not updated the method simulates and exchanges between replicas of the same system with slightly different force fields not involving any sampling of an artificial nonphysical system (hence sampling the desired ensemble). In the present initial application of the method it has been tested on a peptide and a protein test case of very different size indicating in both cases significantly enhanced sampling compared to standard MD simulations. Possible modifications and improvements of the present initial setup of the method will be discussed.

Computational Methods

Test Systems and Simulation Conditions. The RexMD method with an ENM derived biasing potential (ENM-RexMD) was tested on two different biomolecular systems. In all cases MD simulations were performed employing the *Sander* module of the Amber8 package[45] in combination with the parm03 force field.[46] Studies on peptide folding and domain−domain motions were performed employing a generalized Born implicit solvent model as implemented in Amber8 using the pairwise descreening method by Hawkins et al.[47,48] (corresponding to igb=1 in the input of *Sander*). A Debye-Hückel term as implemented in Sander was used with a salt concentration of 1 M. The Settle algorithm[49] was used to constrain bond vibrations involving hydrogen atoms, which allowed a time step of 2 fs. Folding simulations were performed on a small β-hairpin forming chignolin peptide[50] (sequence: GYDPETGTWG, pdb1UA0). An initial extended structure (independent of the experimental structure) was generated using the *xleap* module of the Amber8 package. Five variants of the start structure were generated using short

(5 ps) MD simulations at 800 K with different initial velocities quenched to 280 K within additional 2 ps simulation time. All subsequent peptide folding simulations were performed at 280 K to stabilize folded structures of the peptide. Conventional MD simulations of up to 25 ns starting from the experimental NMR structure (first structure of the NMR ensemble of pdb1UA0) and employing the GB continuum model showed that at this temperature the folded peptide structure is indeed stable. It remained within 2 Å root-mean square deviation of heavy atoms ($Rmsd_{heavy}$) from the experimental start structure (not shown). In addition, under the above simulation conditions and employing the parm03 force field all five extended starting structures folded into structures close to experiment within <25 ns of conventional MD simulations.

In case of T4-lysozyme (T4-L) the high-resolution X-ray structure (pdb2LZM)[51] served as the start structure. The structure was first energy minimized (1000 steps) and subsequently heated to 310 K (37 °C) within 0.2 ns and equlibrated within 1 ns simulation time using the same generalized GB model as for the peptide simulations. The equilibrated structure served as the starting structure for conventional and ENM-RexMD simulations.

Distance-Dependent Biasing Potentials for Soft Degrees of Freedom. At the start and at preset time intervals of the RexMD simulations elastic network model (ENM) calculations on the peptide and proteins were performed following the approach by Hinsen.[36] In the ENM a given protein is assumed to be at an equilibrium (reference) state, and it is described as a set of centers (Cα atoms or heavy atoms) that are connected by harmonic springs. The energy change for any deformation is controlled by spring force constants associated with each pair in the structure. In the ENM of Hinsen the spring constant decays with the distance according to a Gaussian function.[36] For the small chigolin peptide the ENM calculations were performed using all heavy atoms, whereas for lysozyme the Cα backbone atoms were used. The calculation of the elastic network modes took only a few seconds and had a neglectable effect on the overall simulation time. The calculated linear independent and orthogonal eigenvectors and associated eigenvalues of the ENM can be used to calculate B-factors and other properties of the structure depending on Cartesian conformational fluctuations around the reference state. It is also possible to calculate average distance fluctuations compatible with deformations in each mode. As outlined below these distance fluctuations were used to construct a biasing (or penalty/flooding) potential that drives the structure away from the current (reference) structure along directions compatible with the ENM model of the structure.

The ENM analysis and recalculation of the penalty potential was performed at intervals of 15−20 ps. The deformability of a structure in normal modes is given by the corresponding eigenvalue. To calculate distance fluctuations the protein or peptide structure was deformed in each mode i by a factor proportional to $(1/\text{eigenvalue}(i))^{0.5}$ followed by calculation of the interatomic distance variance (change of the square of interatomic distances). That is the soft modes (with small eigenvalue) contribute most to the distance variances. Summation over a set of modes gives the average distance fluctuations compatible with the collective motions of the system. The excitation (or deformation) in each mode was scaled such that the average distance fluctuation (summed over all included modes) did not exceed 2 Å. For the present simulations this value was chosen since it corresponds approximately to the motion of atoms in between recalculation of the ENM modes.

From the distance fluctuations (Δd_{ij}) of a given distance between an atom pair i, j a distance (d_{ij}) dependent penalty potential similar to a Gaussian function (but much less costly to calculate) was constructed:

$$V(d_{ij}) = k([d_{ij} - d_{ij0}]^2 - \Delta d_{ij}^{\,2})^2, \quad \text{if } |d_{ij} - d_{ij0}| \le \Delta d_{ij}$$

$$V(d_{ij}) = 0, \quad \text{otherwise} \tag{1}$$

This penalty potential has its maximum at the distance (d_{ij0}) in the reference state (the structure for which the ENM calculation was performed) and decreases both at smaller and larger distances such that it approaches zero when the change in distance approaches the distance fluctuation (Δd_{ij}) derived from the ENM calculation. It was implemented as an optional distance restraining potential in the disnrg.f routine of the Amber8 package.

To limit the number of added distance dependent penalty potentials only pairs with distance fluctuations 50% larger than the average distance fluctuations were included. The penalty potential basically acts as "flooding" potential to drive the structure away from the current state toward regimes that are compatible with the ENM derived collective degrees of the system. Since in the present implementation fluctuations have been calculated by summation over all mode contributions (according to their eigenvalue), the distances in the distance dependent perturbation potential are not projected onto each mode (invariant to the direction of each mode). One could call the relevant distances also soft distances. A projection on one mode direction is, however, possible if one is for example specifically interested in one selected soft mode direction. The advantage of using a distance-dependent restraining potential is that it is invariant under rotation and can therefore be applied directly during the simulation without a special treatment of rotational components of motion.

RexMD Using an ENM Derived Biasing Potential. In standard RexMD, copies or replicas of the system are simulated at different temperatures (T_0,T_1,T_2,..,T_N). Each replica evolves independently, and after 500−1000 MD-steps (∼1 ps) an exchange of pairs of neighboring replica is attempted according to the Metropolis criterion:

$$w(x_i \rightarrow x_j) = 1 \quad \text{for } \Delta \le 0;$$

$$w(x_i \rightarrow x_j) = \exp(-\Delta) \quad \text{for } \Delta > 0$$

where

$$\Delta = (\beta_i - \beta_j)\,[E(r_j) - E(r_i)] \tag{2}$$

with $\beta = 1/RT$ (R: gas constant and T: temperature), and $E(r)$ representing the potential energy of system for a given

configuration. Instead of modifying the temperature it is also possible to scale the force field (or part of it) along the replica coordinate. In the present case a distance-dependent potential as described in the last paragraph was added to the force field. Each replica runs at a different level of added biasing potential (the first replica runs with the original force field). Note, that the ENM derived distance dependent perturbation potential was only calculated for the structure of the replica that runs at the original force field. Exchanges at every 750 steps (1.5 ps) between neighboring biasing levels were attempted according to[24]

$$w(x_i \rightarrow x_j) = 1 \quad \text{for } \Delta \leq 0;$$

$$w(x_i \rightarrow x_j) = \exp(-\Delta) \quad \text{for } \Delta > 0$$

where

$$\Delta = \beta[(E^j(r_j) - E^j(r_i)) - (E^i(r_j) - E^i(r_i))] \qquad (3)$$

In this case, the Metropolis criterion involves only a single β or temperature and the energy difference between neighboring configurations using the force field for replica j (E^j) minus the same difference using force field for replica i (E^i). Compared to temperature RexMD the energy differences are only affected by the force field term that changes upon going from one replica to another replica run. For all present test cases 5 replicas were used with different levels of the biasing potential. The levels of the biasing potential can be adjusted using the factor k in eq 1. Test calculations indicated that for the replica with highest penalty level a penalty maximum of $20 \times RT$/(number of distances) and appropriate scaling of the intermediate replicas resulted in an acceptance probability for replica exchanges of ~20−30% for the present systems. However, the scaling needed to be adapted for each system in test calculations.

Results

Application of the ENM-RexMD Method to T4-Lysozyme. Lysozyme from the *Escherichia coli* bacteriophage T4 (T4-lysozyme:T4-L) is one of the best studied proteins with >200 T4-L crystal structures of wild-type and mutants available in the protein data bank.[52] The protein consists of two domains (N-terminal and C-terminal domains) that are both involved in substrate binding in a cleft between the domains. Analysis of different crystal forms[52] and structure determination by NMR spectroscopy[53] as well as computer simulation studies[42,54] indicate that the protein can undergo hinge-bending (opening-closing) motions of the two domains. The ENM-RexMD method was applied to T4-lysozme starting from the experimental crystal structure (pdb2LZM). A GB continuum model was used (see the Methods section). For comparison a conventional MD simulation starting from the same start structure and applying the same simulation conditions was also run.

The conventional MD simulation resulted in a backbone Rmsd of ~3.5 Å from the start structure during ~3.2 ns simulation time. No tendency of unfolding was observed during the simulation time (Figure 1). The Rmsd of the N-terminal and C-terminal domains that encompass the

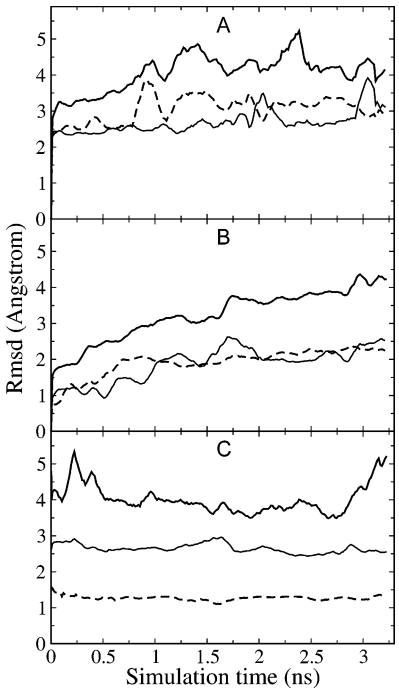

Figure 1. (A) Time course of the T4 lysozyme backbone (Cα) Rmsd (for the replica with the original force field and with respect to the start structure) during the ENM-RexMD (complete structure: continuous bold line) of the N-terminal (residues 15−54: dashed line) and C-terminal (residues 82−152: thine line) domains, respectively, that encompass the ligand binding site. (B) The same Rmsd plots for a conventional MD simulation starting from the same start structure as the ENM-RexMD. (C) Rmsd time courses (same as in A, B) for a conventional MD simulation that started from an open T4 lysozyme structure (obtained during ENM-RexMD after ~1 ns simulation time). Rmsd moving window averages with a window size of 0.1 ns are plotted.

enzyme active site showed a smaller Rmsd of ~2 Å indicating the larger Rmsd of the complete protein is mainly due to a rearrangement of the two domains. In addition to the Rmsd, the distance between centers of mass of atom groups of the N-terminal and C-terminal T4-lysozme domains was also recorded. This distance can be used as a measure of the relative domain motion or the opening and closing of the active site region (Figure 2). During the first 1−2 ns simulation time the domain−domain distance decreased from ~15 Å to ~11.5 Å and stayed at this level throughout the rest of the simulation (Figure 2, green curve). The final distance is slightly smaller than in an X-ray structure of a T4L mutant that is considered to represent one of the most closed forms of the protein (pdb152L).[55] This closing motion is also the reason for the relatively large average Rmsd from the start structure observed during the MD simulation (Figure 1b).

The ENM-RexMD simulation produced an Rmsd time course (for the replica that runs at the original force field) that increased more rapidly at the beginning but with a similar Rmsd toward the end of the simulation. In contrast to the continuous MD simulation, the domain−domain distance flipped many times between several states that caused increased fluctuations of the Rmsd compared to the

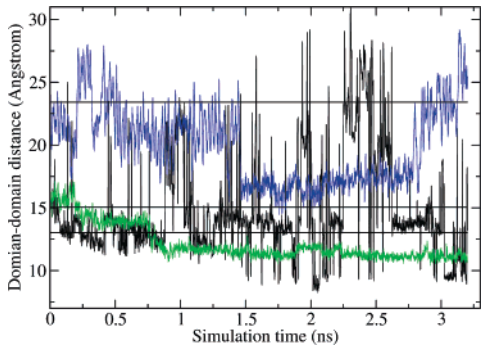

Figure 2. The hinge-bending motion of T4 lysozyme was monitored using the distance (domain−domain distance) between the centers of mass of the Cα backbone atoms of residues 18−26 (belonging to the N-terminal domain) and of residues 138−147 (of the C-terminal domain). These residues form the binding cleft between the N- and C-terminal domains of the enzyme. The distance was recorded for the ENM-RexMD (black line; for the replica with the original force field), during a conventional MD simulation starting from the same start structure as the ENM-RexMD (green line) and during the simulation that started from an open T4 lysozyme structure (blue line). For comparison the corresponding distance in the most closed experimental conformer (pdb152L, smallest distance), the start structure (pdb2LZM), and in one of the most open experimentally determined structures (pdb172L, largest distance) are shown as horizontal (black) lines.

continuous simulation (note, that in Figure 1 these fluctuations are partially damped by plotting a window average of the backbone Rmsd). Similar to the conventional MD simulation the Rmsd of the N- and C-terminal domains was smaller than for the complete protein (Figure 1a). The sampled states included significantly more open conformations than the start structure, states that are close to the start structure, and closed states with a domain−domain distance similar to the states sampled during conventional MD simulations (Figure 2, black line; snapshots shown in Figure 3). Interestingly, several of the sampled open structures showed good agreement (in structure and domain−domain distance) with another X-ray structure of a T4L variant (pdb172L)[52] that has been discussed as a representative conformation for an open T4-lysozyme state.[52,53] A superposition of one selected snapshot on the pdb172L structure (Rmsd(Cα)=2.8 Å) is shown in Figure 3d.

In order to control if the open structures produced during the ENM-RexMD run represented also stable states during conventional MD simulations one such open structure (from a simulation time of ∼1 ns of the ENM-RexMD) was used as a start structure for a conventional MD simulation. During this simulation the Rmsd of the complete protein with respect to the start structure remained at ∼4 Å, and smaller Rmsds of the individual N- and C-terminal domains (∼1 Å and ∼2.5 Å, respectively, Figure 1c) were observed. Several open forms with different domain−domain distances were sampled (blue curve in Figure 2). A transition to a closed form was not observed during the 3.2 ns simulation time. However, an extension of the simulation to 10 ns resulted in a closed structure (not shown).

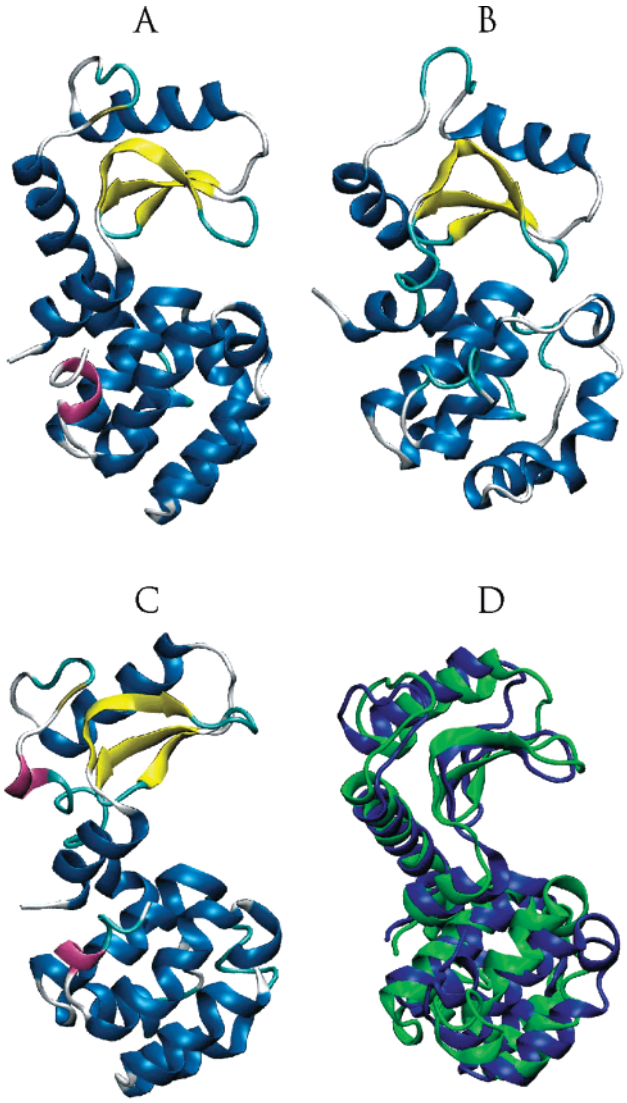

Figure 3. Snapshots of the T4 lysozyme structure observed during the ENM-RexMD simulations: (A) start structure (pdb2LZM), (B) a structure with a more closed cleft between N- and C-terminal domains compared to the start structure, and (C) an open protein structure observed during ENM-RexMD after ∼1 ns simulation time and used as a start structure for a continuous MD simulation. Structures are shown as a cartoon representation with a color coding according to a secondary structure. (D) Superposition of an open T4 lysozyme structure observed during the ENM-RexMD (green cartoon) and the experimental X-ray structure pdb172L (blue) which has been considered as one of the most open available experimental conformers.

The result indicates that the ENM-RexMD shows significantly improved sampling of open and closed states with many (>20) sampled transitions compared to the continuous MD simulations on a relatively short time scale of 3.2 ns. In contrast, not a single complete open−close or close−open transition was sampled during the same simulation time in the conventional MD simulations.

Folding Simulations on a β-Hairpin Forming Peptide. The ENM-RexMD approach was further evaluated on the small 10 residue chignolin β-hairpin forming peptide. The structure of this peptide was recently determined by NMR

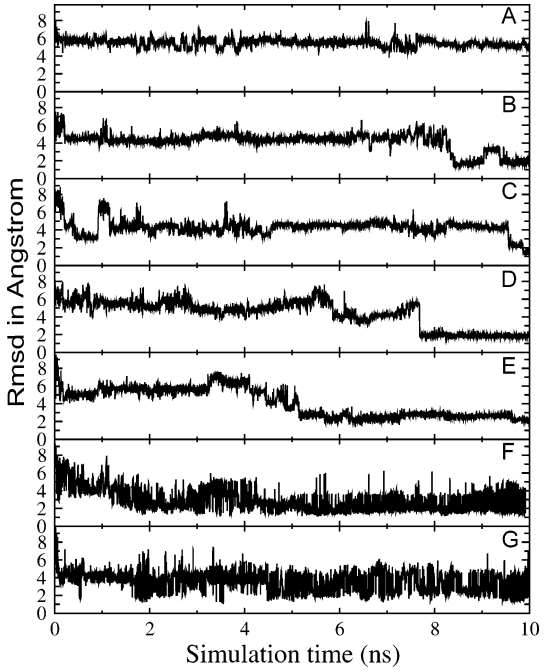

Figure 4. Rmsd (heavy atoms) from the experimental structure of the chignolin peptide (first model of pdb1UA0) observed during standard MD simulations starting from five different extended start structures (A−E). (F) Rmsd (heavy atoms) observed during a 5 replica ENM-RexMD (for the replica with the original force field) starting from five different start structures used in the standard MD simulations A−E. (G) was the same as in (F) but starting from the same start structure in all replicas (the start structure used in the first (A) standard MD simulation).

experiments.[50] It has also been demonstrated that extensive conventional temperature RexMD simulations (using 16 replicas) of more than 100 ns can lead to a folded structure very similar to the experimental NMR structure.[10]

Furthermore, a recent backbone dihedral biasing the potential RexMD approach (termed BP-RexMD) also achieved folding of the peptide to structures in close agreement with experiment within ∼10 ns MD simulations in explicit solvent and required only 7 replicas.[33]

Test calculations using an implicit GB continuum model as implemented in Amber8 (igb=1) and the parm03 force field indicated that it is also possible to achieve folding of this peptide in implicit solvent to structures in close agreement with experiment (see also the Methods section). The folded conformation was the most populated structure during 25 ns MD simulations at 280 K staying within 2 Å ($Rmsd_{heavy}$) of the experimental start structure (not shown). In order to test the ENM-RexMD approach five different chignolin start structures were generated by short MD simulations at high temperature starting from a fully extended conformation (see the Methods section). As indicated in Figure 4 conventional MD simulations lead to conformations close to experiment after 5−10 ns for four out of five start structures (Rmsd ∼2 Å, Figure 4). However, for the first start structure significantly longer simulations (>25 ns) were necessary to achieve a transition to structures in close agreement with experiment (presumably because of an altered Pro backbone conformation compared to the other start

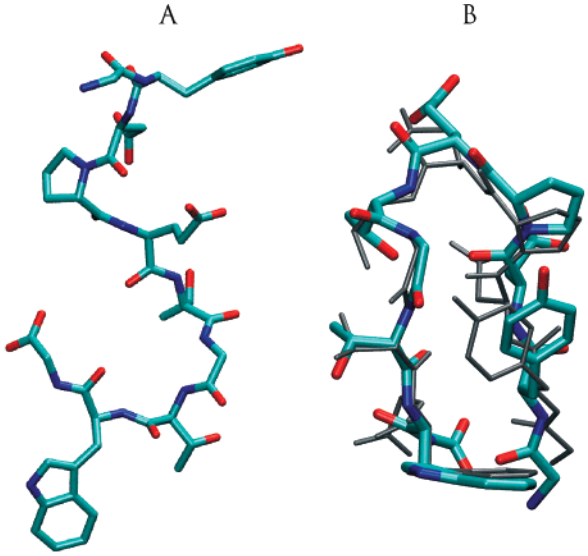

Figure 5. (A) Stick model (atom color-coded) of an extended start structure of the ENM-RexMD simulations. (B) Superposition of a typical near-native structure sampled during ENM-RexMD (stick model, atom color coded) and the experimental NMR structure (first structure of NMR ensemble in pdb1UA0, thin gray stick model). The heavy atom Rmsd between the two structures was ∼1.6 Å.

structures that corresponded to an altered rotation of the Pro psi dihedral angle).

For comparison, two ENM-RexMD simulations with 5 replicas were set up. In contrast to the T4L simulations ENM calculations were performed using all heavy atoms of the peptide structure and were recalculated every 15 ps. One simulation started from all five different start structures (one for each replica and the first most difficult start structure assigned to the replica with the original force field). The second ENM-RexMD simulation started from the first start structure (the most difficult one) assigned as start for all replicas. In both ENM-RexMD simulations transitions to a conformation in good agreement with experiment (Figure 5) were seen already after ∼1.5−2 ns (Figure 4, last two plots) significantly faster than in the conventional MD simulations. These structures also quickly evolved to the most populated conformational states (Figure 4).

The mean potential energy of the structures (averages over 0.4 ns) dropped much faster already during the first 2 ns of the ENM-RexMD compared to the conventional MD simulations (Figure 6). Also, the energy probability distribution obtained during the first 10 ns conventional MD simulations (of the 5 different start structures) are all shifted to higher energies compared to the distribution obtained from the ENM-RexMD simulation (Figure 6b). However, the energy distribution obtained after 80 ns conventional MD was in close agreement with the result of 40 ns ENM-RexMD (for the replica run at the original force field, Figure 6c) for each start structure except for the first start structure (thin line in Figure 6c). As mentioned above, in this case transitions to near native conformations (combined with a drop of the potential energy) were observed only after simulation times >25 ns which is a likely reason for the shift of the distribution curve to slightly higher energies. The close

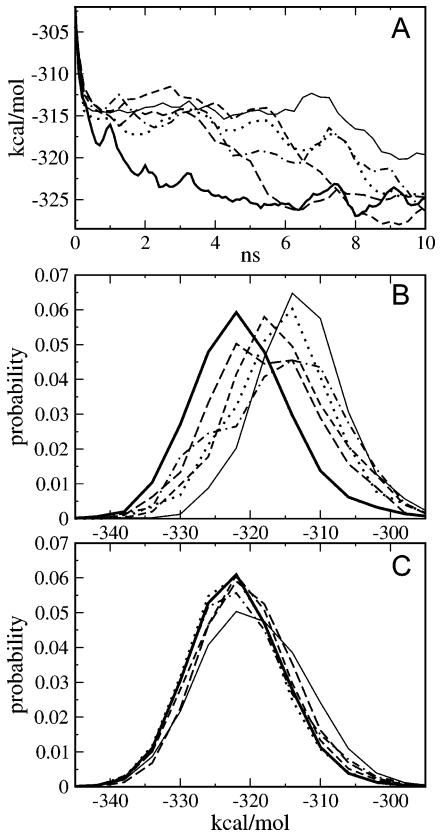

Figure 6. (A) Decrease of the mean potential energy (averages of 0.4 ns) during the first 10 ns of the ENM-RexMD (for the ENM-RexMD that started from 5 different start structures and the replica run with the original force field: bold continuous line). For comparison the decrease in mean potential energy is also shown for 5 conventional MD simulations starting from the 5 different start structures used also for the ENM-Rex-MD (five different line types). (B) Energy probability distribution obtained during the first 10 ns simulation time of the ENM-RexMD (replica run with the original force field: bold continuous line) and of the 5 conventional MD simulations (starting from the same 5 start structures, same line types as in A). (C) Energy probability distribution obtained during entire ENM-RexMD (40 ns; replica run with the original force field: bold continuous line) and 80 ns conventional MD simulation starting from each of the 5 start structures (same lines types as in A, B).

agreement of the final energy distribution of the ENM-RexMD and the conventional MD simulations indicates that the same energetic states are sampled and that the recalculation of the ENM from time to time during the RexMD simulation has only a minor influence on the calculated ensemble of states. The energy distribution obtained during the first 10 ns ENM-RexMD (bold curve in Figure 6b) is already very similar to the complete 40 ns ENM-RexMD (bold curve in Figure 6c) indicating a very rapid relaxation of the sampled ensemble of states. To further check if ENM-RexMD and extensive conventional MD simulations result in similar sampled conformations the Rmsd probability distribution (heavy atoms with respect to experimental structure) of conformations from all 5 conventional MD simulations was compared with those obtained during ENM-RexMD (Figure 7). During the first 10 ns the ENM-RexMD sampled an Rmsd distribution quite similar to the distribution of the entire 40 ns ENM-RexMD (bold curves in Figure 7a,b). However, the Rmsd distribution of the conventional MD simulation was shifted to higher Rmsd values during the initial 10 ns but was very similar to the result of the ENM-RexMD in case of using conformations sampled during all complete conventional MD simulations (5×80 ns; Figure 7b). In addition, a cluster analysis with respect to the backbone Rmsd of sampled structures was performed with an Rmsd cutoff of 2 Å and using the *kclust* program of the MMTSB tools.[57] For the cluster analysis 10 000 structures from the ENM-RexMD and 10 000 structures from the final 20 ns of the conventional MD simulations were analyzed. The first 5 clusters with highest population (representing in both cases ~60% of the sampled conformations) showed high overlap. In both cases the highest populated cluster corresponded to structures in close agreement with the experimental structure (Figure 8a). The close correspondence of representative conformations of highly populated clusters from the ENM-RexMD and the conventional MD is shown in Figure 8 indicating that extensive MD simulations starting from different start structures and ENM-RexMD simulation sample similar conformational ensembles.

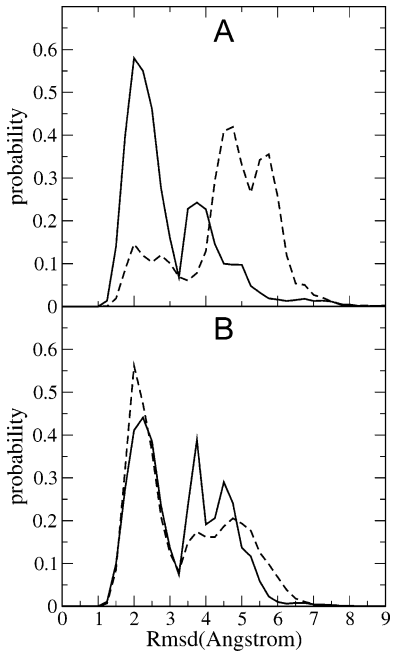

Figure 7. (A) Heavy atom Rmsd probability distribution (Rmsd with respect to experimental reference structure: the first model of pdb1UA0) observed during the first 10 ns of ENM-RexMD for the simulation that started from 5 different structures and for the replica run with the original force field: continuous line) and combined result for all 5 conventional MD simulations (starting from the same 5 different start structures: dashed line). (B) The same result as in (A) but for the entire ENM-RexMD (40 ns) and all conventional MD simulations (5×80 ns).

Discussion

To improve conformational sampling RexMD simulations have evolved as an important tool to study biomolecular

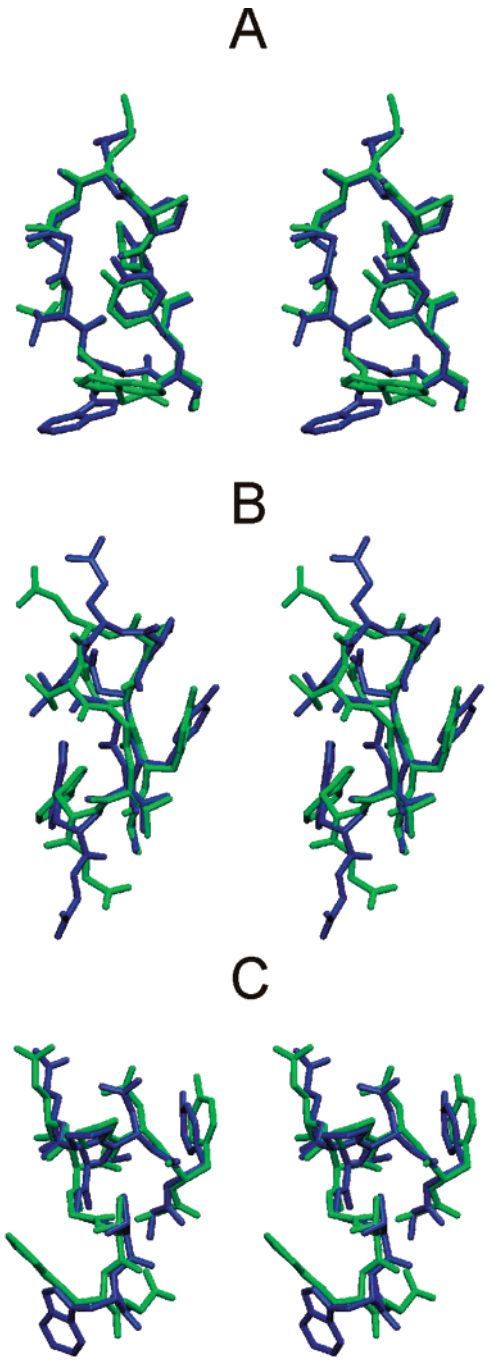

Figure 8. Comparison of highly populated conformational clusters observed during ENM-RexMD (for the replica run with the original force field: green stick models) and during the final 20 ns of all conventional MD simulations (blue structures). Cluster analysis was performed using *kclust* of the MMTSB tools.[56] (A) Superposition (stereoview) of the best representative conformation (closest backbone Rmsd(Cα) to the cluster centroid) for the largest populated cluster observed during ENM-RexMD (green) and conventional MD (blue) with an Rmsd(Cα) of 0.5 Å. (B) Superposition of the best representative of the second most populated cluster of ENM-RexMD and the third most populated cluster of conventional MD (Rmsd(Cα)=0.3 Å). (C) Same as for cluster 5 of ENM-RexMD and cluster 2 of conventional MD (Rmsd(Cα)=0.4 Å).

structures.[5−8,10−33] A major disadvantage of conventional temperature RexMD is the rapid increase of the required number of replicas to cover a desired temperature range.[24] The ratio of the standard deviation of the system potential energy (a measure of the energy fluctuation) vs average energy decreases with the square-root of the system size. Hence, to achieve sufficient overlap of the energy distributions between replicas run at different temperatures (required to achieve a reasonable exchange acceptance ratio) the temperature "spacing" between neighboring replicas is required to decrease with system size. Another drawback of large numbers of replicas is the need to run longer simulations (or more exchanges) to allow sufficient "travelling" or exchanges between high- and low-temperature replicas compared to a small number of replicas. Approaches that circumvent these drawbacks have been developed that include separate temperature coupling during RexMD of solute and solvent[27] or separate temperature coupling of essential degrees of freedom of the system.[28]

In Hamiltonian Rex-MD simulations the potential energy function (Hamiltonian) is scaled along the replicas.[24,25,29−33] A critical issue in the application of such Hamiltonian replica exchange methods is the choice and magnitude of force field energy terms to be scaled along the replicas. Force field terms that drastically change the energy of the simulation system may require many replicas with intermediate levels of the selected force field term to allow efficient exchanges between replicas.

One possibility is to design biasing force field terms that depend specifically on soft degrees of freedom of the system. An advantage is that the magnitude of the biasing potential can be kept small (relative to the total energy of the system) since motions in soft degrees of freedom naturally require smaller driving forces than hard degrees of freedom. Second, the likelihood to observe large scale conformational changes is also enhanced because these often involve motions along soft degrees of freedom of the molecule. The coarse-grained nature of the ENM description has the advantage that it provides soft degrees of freedom of a molecule at a smoother and more long-range level than the atomistic force field description. Hence, the ENM model of a protein molecule can look "beyond barriers" present at the level of a molecular mechanics force field. As already mentioned, the idea to couple ENM analysis to MD simulation has been explored by increasing the effective temperature of motion along soft ENM modes.[42] This resulted in an enhanced conformational sampling, however, with the drawback that the simulation is performed on a nonphysical system (separate temperature coupling of different degrees of freedom of the system).

In the present ENM-RexMD approach an ENM derived distance-dependent penalty or biasing potential is added at various levels for each replica and acts to help to drive the conformation toward regimes compatible with the ENM description. As long as the ENM description is not changing, each replica samples a canonical distribution of conformations compatible with the force field description for each replica. Hence, exchanges are performed between canonical ensembles such that detailed balance conditions are fulfilled (in the present method at no point a nonphysical system is sampled). Since the structure of the peptide or protein is changing, it is necessary from time to time to recalculate

the ENM of the system and to adapt the biasing potential (this could also be done using a Monte Carlo scheme to adapt or reject changes in the biasing potential). For the present simulation systems the adaptation was done every 15−20 ps. The initial application of the method resulted in successful folding of a β-hairpin forming peptide in close agreement with experiment and faster than observed during standard MD simulations. For this peptide, folded structures in close agreement with experiment were observed as the most populated state already after <2 ns starting from several different start structures. In order to control if the recalculation of ENM modes from time to time during the ENM-RexMD has an influence on the energies and types of sampled structures, the sampled conformations were compared with trajectories obtained from extended (80 ns) conventional MD simulations. Very good agreement of the potential energy distribution, Rmsd distribution (with respect to the experimental structure), and the most populated conformational clusters between ENM-RexMD and conventional MD at the final stage of the simulations was obtained. This result shows that the updating of ENM-modes has only a small influence on the canonical equilibrium sampling of conformational states. However, much smaller differences in the energy and Rmsd distributions during the early stage vs complete ENM-RexMD simulation were found indicating a much faster relaxation toward an equilibrium distribution compared to conventional MD simulations.

It needs to be emphasized that in the present initial application of the method no extensive optimization of the simulation parameters was performed. It is likely that further simulation parameter optimization can lead to a further improvement of the conformational sampling of the approach. Simulation evaluation and trajectory analysis has been performed only using the unbiased replica run with the original force field (although after proper reweighting the other replicas could also be included).

Application to T4 lysozyme, a two domain protein, resulted in significantly enhanced sampling of hinge-bending motions. During the ENM-RexMD many transitions between open and closed structures (detected by calculating domain−domain distances) were observed. The Rmsd of the individual N- and C-terminal domains showed a deviation of ∼2.5 Å from the experimental structure that was generally smaller than the Rmsd of the complete protein from experiment (∼3.5−4 Å). The sampled domain−domain distance fluctuations reached ∼10−15 Å indicating domain−domain rearrangements that cannot be explained by intramolecular changes within each domain. In addition, sampled open structures were in quite good agreement with the X-ray structures of an open T4L conformation structure (a snapshot with an Rmsd(Cα)=2.8 Å from the open T4L structure in pdb172L is shown in Figure 3d) although the simulation started from a different more closed experimental structure. Conventional MD simulations of the same structures generated only conformations relatively close to the start conformation on the same time scale as the ENM-RexMD runs. This was achieved with a small number of 5 replicas for both the β-hairpin and the T4L system.

In the present implementation the ENM derived distance dependent biasing potential was only calculated for the structure from the replica that runs with the original force field (reference replica). This means that the potential drives the structures in each replica run away from the structure that runs with the original force field (to offer alternative low-energy structures that can then exchange with the reference replica run). In the T4L case the structures in each replica run differ mainly by the degree of opening/closing of the enzyme active site, and the soft mode directions of these structures are presumably quite similar. However, in the peptide folding case the structures in each replica run may differ significantly, and the corresponding mode directions of the structure in the reference replica run may also significantly differ from those of the reference structure. Hence, one may think that in this case the ENM derived potential in each replica acts mainly as a random perturbation potential. However, frequently the structures in each replica run contain at least segments that are similar to segments of the reference structure. Since the biasing potential has been expressed in terms of intramolecular distances, any such segment will be perturbed by the added perturbation potential (in a nonrandom fashion). It should be emphasized that other more sophisticated coupling schemes of ENM and RexMD might be possible that include for example ENM derived biasing potentials calculated separately for the structures of each replica run.

As has been pointed out by Huang et al.[57] a RexMD method that scales only part of the Hamilitonian of a system to reduce the number of required replicas in a RexMD simulation may not cover a large range of different conformations simultaneously (e.g., both unfolded and completely folded structures). It should therefore be emphasized that the current ENM-RexMD method primarily enhances the conformational sampling locally around a reference state. Therefore, it is necessary to recalculate the ENM at frequent intervals to adapt to a new reference state of the system.

A drawback of the ENM-RexMD approach is that for the current setup several parameters need to be adjusted. This includes the scaling of the distance fluctuations obtained from the ENM analysis, the maximum height and levels of the biasing potential along the replicas, and the frequency of recalculating the ENM description of the molecule. Future work will focus on a systematic evaluation of the parameters used during the ENM-RexMD approach and a possible automatic setup for a given simulation system.

Acknowledgment. This work was performed using the computational resources of the CLAMV (Computational Laboratories for Animation, Modeling and Visualization) at Jacobs University Bremen and supercomputer resources of the EMSL (Environmental Molecular Science Laboratories) at the PNNL (Pacific Northwest National Laboratories; grant gc11-2002).

References

(1) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Reversible peptide folding in solution by molecular dynamics simulation. *J. Mol. Biol.* **1998**, *280*, 925−932.

(2) Duan, Y.; Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **1998**, *282*, 740−744.

(3) Roccatano, D.; Amadei, A.; Di Nola, N.; Berendsen, H. J. C. A molecular dynamics study of the 41−56 β-hairpin from b1 domain of protein G. *Protein Sci.* **1999**, *10*, 2130−2141.

(4) Pande, V. S.; Roshkar, D. S. Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment of protein G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062−9067.

(5) Garcia, A. E.; Sanbonmatsu, K. Y. Exploring the energy landscape of a β-hairpin in explicit solvent. *Proteins: Struct., Funct., Bioinf.* **2001**, *42*, 345.

(6) Zhou, R.; Berne, B. J.; Germain, R. The free energy landscape for β-hairpin folding in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931−14937.

(7) Simmerling, C.; Strockbine, B.; Roitberg, A. E. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **2002**, *124*, 11258−11259.

(8) Rao, F.; Caflisch, A. Replica exchange molecular dynamics simulations of reversible folding. *J. Chem. Phys.* **2003**, *119*, 4035−4042.

(9) Roccatano, D.; Nau, W. M.; Zacharias, M. Structural and dynamic properties of the CAGQW peptide in water: A molecular dynamics simulation study using different force fields. *J. Phys. Chem.* **2004**, *108*, 18734−18742.

(10) Seibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. Reproducible Polypeptide Folding and Structure Prediction using Molecular Dynamics Simulations. *J. Mol. Biol.* **2005**, *354*, 173−183.

(11) Nguyen, P.; Stock, G.; Mittag, E.; Hu, C-K.; Li, M. S. Free energy landscape and folding mechanism of a β-hairpin in explicit water: A replica exchange molecular dynamics study. *Proteins* **2006**, *61*, 795−806.

(12) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. Peptide folding simulations. *Curr. Opin. Struct. Biol.* **2003**, *15*, 168−175.

(13) Swendsen, R. H.; Wang, J. S. Replica Monte Carlo simulations of spin glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.

(14) Okamoto, Y. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J. Mol. Graphics Modell.* **2004**, *22*, 425−439.

(15) Predescu, C.; Predescu, M.; Ciobanu, C. V. J. On the Efficiency of Exchange in Parallel Tempering Monte Carlo Simulations. *J. Phys. Chem. B* **2005**, *109*, 4189−4196.

(16) Okabe, T.; Kawata, M.; Okamoto, Y.; Mikami M. Replica-exchange Monte Carlo method for the isobaric−isothermal ensemble. *Chem. Phys. Lett.* **2001**, *335*, 435−439.

(17) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(18) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **2001**, *60*, 96−123.

(19) Sanbonmatsu, K. Y.; Garcia, A. E. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 225.

(20) Zhou, R.; Berne, B. J. Can a continuum solvent model reproduce the free energy landscape of a β-hairpin folding in water. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777−12782.

(21) Zhou, R. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins: Struct., Funct., Bioinf.* **2003**, *53*, 148−161.

(22) Nymeyer, H.; Garcia, A. E. Simulation of the folding equilibrium of α-helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934−13939.

(23) Yoshida, K.; Yamaguchi, T.; Okamoto, Y. Replica-exchange molecular dynamics simulation of small peptide in water and in ethanol. *Chem. Phys. Lett.* **2005**, *41*, 2280−2284.

(24) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. *J. Chem. Phys.* **2002**, *116*, 9058−9065.

(25) Kannan.; S.; Zacharias, M. Folding of a DNA hairpin loop structure in explicit solvent using replica-exchange molecular dynamics simulations. *Biophys. J.* **2007**, *93*, 3218−3228.

(26) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. J. Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. *J. Chem. Theory Comput.* **2006**, *2*, 420−433.

(27) Cheng, X.; Cui, G.; Hornak, V.; Simmerling, C. Modified Replica Exchange Simulation Methods for Local Structure Refinement. *J. Phys. Chem. B* **2005**, *109*, 8220−8230.

(28) Kubitzki, M. B.; de Groot, B. L. Molecular dynamics simulations using temperature-enhanced essential dynamics replica exchange. *Biophys. J.* **2007**, *92*, 4262−4270.

(29) Jang, S.; Shin, S.; Pak, Y. Replica-exchange method using the generalized effective potential. *Phys. Rev. Lett.* **2003**, *91*, 58305−58309.

(30) Zhu, Z.; Tuckerman, M. E.; Samuelson, S. O.; Martyna, G. J. Using Novel Variable Transformations to Enhance Conformational Sampling in Molecular Dynamics. *Phys. Rev. Lett.* **2002**, *88*, 100201.

(31) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. A. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749−13754.

(32) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. A Novel Hamiltonian Replica Exchange MD Protocol to Enhance Protein Conformational Space Sampling. *J. Chem. Theory Comput.* **2006**, *2*, 217−228.

(33) Kannan, S.; Zacharias, M. Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 697−706.

(34) Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.

(35) Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Des.* **1997**, *2*, 173−181.

(36) Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins: Struct., Funct., Bioinf.* **1998**, *33*, 417−429.

(37) Tama, F.; Sanejouand, Y. H. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* **2001**, *14*, 1−6.

(38) Tobi, D.; Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18908−18913.

(39) Bahar, I.; Rader, E. J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586−592.

(40) May, A.; Zacharias, M. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim. Biophys. Acta* **2005**, *1754*, 225−231.

(41) May, A.; Zacharias, M. Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 794−809.

(42) Zhang, Z.; Shi, Y.; Liu, H. Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophys. J.* **2003**, *84*, 3583−3593.

(43) Grubmüller, H. Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys. Rev. E* **1995**, *52*, 2893−2906.

(44) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *99*, 12562−12566.

(45) Case, D.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *Amber 8*; University of California: San Francisco, CA, 2003.

(46) Duan, Y.; Wu, A.; Chowdhury, C. S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.

(47) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric continuum. *Chem. Phys. Lett.* **1995**, *246*, 122−129.

(48) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based onpairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824−19839.

(49) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(50) Honda S.; Yamasaki K.; Sawada Y.; Morii, H. 10 residue folded peptide designed by segment statistics. *Struct. Fold. Des.* **2004**, *12*, 1507−1518.

(51) Weaver, L. H.; Matthews, B. W. Structure of bacteriophage T4 lysozyme refined at 1.7 Angstrom resolution. *J. Mol. Biol.* **1987**, *193*, 189−199.

(52) Zhang, X.-J.; Wozniak, J. A.; Matthews, B. W. Protein flexibility and adaptability seen in 25 crystal forms of T4 lyzozyme. *J. Mol. Biol.* **1995**, *250*, 527−552.

(53) Goto, N. K.; Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. What is the average conformation of bacteriophage T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR. *J. Mol. Biol.* **2001**, *308*, 745−764.

(54) de Groot, B. L.; Hayward, S.; van Alten, D. M. F.; Amadei, A.; Berendsen, H. J. C. Domain motions in bacteriophage T4 lysozyme: A comparison between molecular dynamics and crystallographic data. *Proteins: Struct., Funct., Bioinf.* **1998**, *31*, 116−127.

(55) Pjura, P. E.; Matsumura, M.; Wozniak, J. A.; Matthews, B. W. Structure of a thermostable disulfide-bridge mutant of phage T4 lysozyme shows that an engineered cross-link in a flexible region does not increase rigidity of the folded protein. *Nature* **1990**, *345*, 86−89.

(56) Feig, M.; Karanicolas, J.; Brooks, C. L. MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graphics Modell.* **2004**, *22*, 377−395.

(57) Huang, X.; Hagen, M.; Kim, B.; Friesner, R. A.; Zhou, R.; Berne, B. J. Replica exchange with solute tempering: efficiency in large scale systems. *J. Phys. Chem. B* **2007**, *111*, 5405−5410.

CT7002258