THE JOURNAL OF PHYSICAL CHEMISTRY B LETTERS

2007, *111,* 3073−3076
Published on Web 03/06/2007

Metadynamics in Essential Coordinates: Free Energy Simulation of Conformational Changes

Vojtěch Spiwok,* Petra Lipovová, and Blanka Králová

Department of Biochemistry and Microbiology, Institute of Chemical Technology, Prague, Technická 3, 166 28 Prague 6, Czech Republic

Received: December 14, 2006; In Final Form: January 30, 2007

We propose an approach that combines an extraction of collective motions of a molecular system with a sampling of its free energy surface. A recently introduced method of metadynamics allows exploration of the free energy surface of a molecular system by means of coarse-grained dynamics with flooding of free energy minima. This free energy surface is defined as a function of a set of collective variables (e.g., interatomic distances, angles, torsions, and others). In this study, essential coordinates determined by essential dynamics (principle component analysis) were used as collective variables in metadynamics. First, dynamics of the model system (explicitly solvated alanine dipeptide, Ace-Ala-Nme) was simulated by a classical molecular dynamics simulation. The trajectory (1 ns) was then analyzed by essential dynamics to obtain essential coordinates. The free energy surface as a function of the first and second essential coordinates was then explored by metadynamics. The resulting free energy surface is in agreement with other studies of this system. We propose that a combination of these two methods (metadynamics and essential dynamics) has great potential in studies of conformational changes in peptides and proteins.

Introduction

Accurate modeling of conformational motions in proteins is fundamental for understanding molecular recognition, induced fit, allosteric effect, function of molecular motors, and other processes. Molecular dynamics simulations (MDS) have been widely used to study conformational behavior of biomacromolecules since first applied to protein science.[1] However, there are several limitations of this method. Namely, currently accessible time scales of MDS (nano- to microseconds) are significantly shorter than the time scales of a majority of biologically interesting conformational changes. In other words, the probability of occurrence of these conformational changes within times of standard MDS is negligible. Another problem of MDS is that it does not directly provide quantitative information about a free energy surface (FES). In principle, it should be possible to evaluate FES using classical MDS because this method samples the studied system with canonical distribution. However, this approach generally fails due to the problem of trapping in free energy minima. Several methods aiming to enhance sampling of conformational space and/or to evaluate free energy changes have been developed in recent years.[2] Recently introduced metadynamics[3] explores FES as a function of a limited number of collective variables. Choice of these variables is crucial for its successful application. In some applications of metadynamics, the choice of collective variables is straightforward. However, in modeling conformational changes, it is often difficult to find a coordinate(s) that represents progress along a pathway of conformational change. Here, we assess an implementation of metadynamics in the space of essential coordinates determined by means of essential dynamics analysis.[4] This approach comprises two phases: first, a classical MDS is performed and essential coordinates are determined (further denoted as the essential coordinate determination run). Second, the FES is sampled using metadynamics with essential coordinates as collective variables. In the similar method presented by Lange and co-workers,[5] the potential energy is being flooded by a bias potential to avoid trapping in free energy minima.

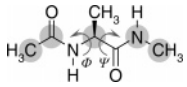

Figure 1. Covalent structure of the alanine dipeptide. Atoms selected for essential dynamics analysis and essential coordinate determination are highlighted by gray background.

The presented approach was tested on the alanine dipeptide (Ace-Ala-Nme, Figure 1). Amino acid dipeptides are the simplest model molecules for studying conformational behavior of peptides and proteins. Numerous free energy simulation methods were tested on this system (see ref 6 and references within).

Method

Trajectories of MDS are often analyzed by essential dynamics analysis[4] to trace major collective motions of the studied biomacromolecule. In this method, the first covariance matrix is calculated from superimposed coordinates **r** of N atoms:

$$C_{ij} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle,\ i, j = 1, 2 \ldots 3N \qquad (1)$$

where r_i is the i-th Cartesian coordinate of the system.

* E-mail: spiwokv@vscht.cz.

10.1021/jp068587c CCC: $37.00

Diagonalization of the covariance matrix leads to eigenvalues and eigenvectors:

$$\mathbf{C} = \mathbf{E}\Lambda\mathbf{E}^{-1} \tag{2}$$

where $\mathbf{E}$ is a matrix of eigenvectors and Λ is a matrix of eigenvalues. Each eigenvector represents a certain collective motion of the molecule, and the corresponding eigenvalue represents an extent of this motion. The projection of the conformation at time t on the k-th eigenvector is defined as

$$p_k^t = (\mathbf{r}^t - \langle\mathbf{r}\rangle)\cdot\mathbf{e}_k \tag{3}$$

where $\mathbf{r}^t$ is the conformation at time t superimposed onto $<\mathbf{r}>$. Therefore, it is possible to study dynamics of the system in coordinates p_1 to p_n (i.e., essential coordinates). Analysis of trajectories by means of essential dynamics has been successfully applied in tracing collective motions in various proteins[7,8] and in data compression.[9] Methods, such as essential dynamics sampling,[10,11] dynamics with enhanced motions in directions of eigenvectors,[12] and a flooding-based method,[5] were developed on the basis of essential dynamics.

In this work, extraction of collective motions was combined with sampling of FES. Metadynamics (escaping free-energy minima, hills method) is a free energy calculation method recently introduced by Laio and Parrinello.[3] The physical basis of metadynamics is described in detail in ref 3. Briefly, first, a set of collective variables (p_1 to p_n) that are supposed to determine the FES of a studied system is selected. The gradient of free energy surface F_k is calculated using a short MDS run with a constraint applied to p_k:

$$F_k = -\frac{\partial A(\mathbf{p})}{\partial p_k},\ k = 1, 2\ldots \text{ number of collective variables} \tag{4}$$

The system could then "walk" in collective variables with coarse-grained dynamics described by the formula:

$$p_k^{t+1} = p_k^t + \delta p\,\frac{F_t^k}{||\mathbf{F}^t||} \tag{5}$$

where δp is size of a step. To avoid freezing in the nearest free energy minimum, the force is replaced by a time-dependent term:

$$F_k \rightarrow F_k - \frac{\partial}{\partial p_k} w \sum_{t'\le t}\prod_k \mathrm{e}^{-(|p_k - p_k^{t'}|^2)/(2\delta p^2)} \tag{6}$$

In every step of metadynamics, a Gaussian-shaped "hill" (of height w) is added to the FES. These hills stepwise "flood" free energy wells and thus allow the system to escape free energy minima. The sum of these Gaussians after completion of this "walking" approximates the FES of the studied system. This method was successfully applied in various fields (e.g., in conformation studies of biomolecules,[13] protein−ligand docking,[14] chemical reactions studies,[15] and even geophysics).[16] Metadynamics has great potential to solve the problem of the time-consuming nature of MDS because it can explore rarely occurring events while simultaneously providing quantitative information about FES.

Computational Details

All simulations were performed with the GROMACS package[17] using the AMBER95 all-atom force field,[18] implemented using AMBER ports.[19] One molecule of alanine dipeptide was placed in a periodic box with 313 water molecules (20.05 × 21.94 × 21.92 Å, TIP3P parameters).[20] The time step was 1 fs, and neither bond nor angle constraints were applied to the solute. The electrostatic cutoff was set to 1.0 nm. The temperature was kept constant at 300 K using a Berendsen thermostat. Metadynamics in essential coordinates was implemented using our in-house Python program. MDS with constraints applied in directions of eigenvectors was performed using the essential dynamics sampling feature of GROMACS.[10] The rms fit method developed by Coutsias, Seok, and Dill[21] was utilized.

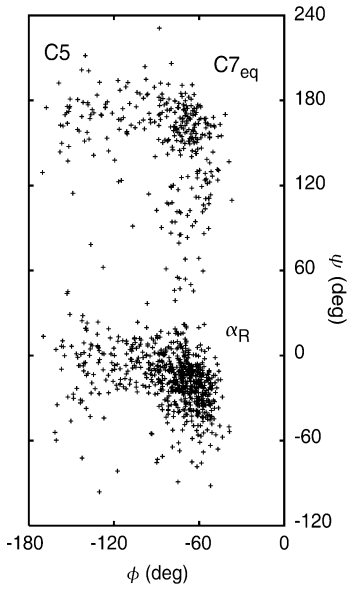

Figure 2. Trajectory of the essential coordinate determination run illustrated in space of $\Phi-\Psi$ dihedral angles. Conformations α_R, $C7_{eq}$, and C5 are indicated.

Results

Classical MDS in a duration of 1 ns (1000 snapshots) was performed in order to obtain essential coordinates describing the studied system (the essential coordinate determination run). Figure 2 illustrates that three major conformations of the alanine dipeptide (α_R, $C7_{eq}$, and C5) were explored during this short simulation. Conformations with a value of Φ greater than zero (such as $C7_{ax}$) were not observed during the essential coordinate determination run.

Five atoms of the alanine dipeptide molecule were selected for essential dynamics analysis of the trajectory (Figure 1). The selection was driven by two factors: first, these atoms had to be involved in major conformational changes in the studied molecule, and second, the molecule had to retain sufficient conformational freedom even with constraints applied to these atoms. Excessive freezing of the molecule could disproportionately increase rotational and translational motions compared to internal motions of the molecule and thus cause potential artifacts.

The essential dynamics analysis revealed two major collective motions present in the structure of the alanine dipeptide. Their eigenvalues account for 92% of the sum of all (fifteen) eigenvalues (see Supporting Information). The projection of the classical MDS trajectory onto the first and second eigenvectors is illustrated in Figure 3A.

The aim of this study was to test the possibility of the application of essential coordinates as collective variables in metadynamics. Metadynamics in the space of essential coordi-

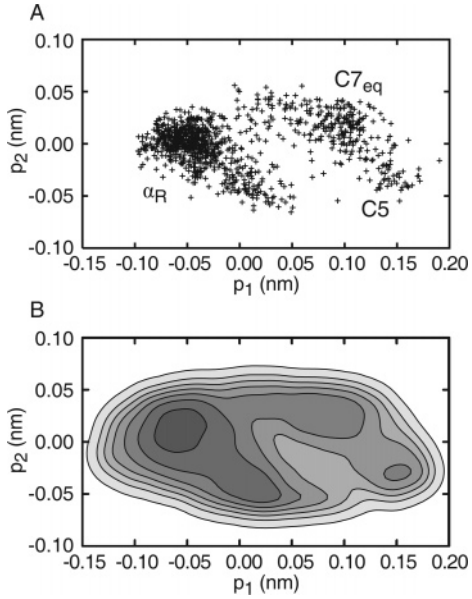

Figure 3. (A) Trajectory of classical MDS (essential coordinate determination run) displayed in essential coordinates. All three conformational families are indicated. (B) Results of metadynamics in essential coordinates after 10 000 steps (corresponding to 10 ns of classical MDS). Contours are plotted every 4 kJ/mol.

nates was implemented using our in-house program written in Python language. Values of δp and w were set to 0.01 nm and 0.1 kJ/mol, respectively (see eqs 5 and 6). Force was evaluated using 1 ps microscopic MDS (plus 0.2 ps of equilibration phase) within each step of metadynamics. The conformation belonging to the α_R cluster was selected as the starting structure. FES calculated by 10 000 steps of metadynamics in essential coordinates is illustrated in Figure 3B. This corresponds to 10 ns of classical MDS (12 ns including equilibration). In agreement with other studies,[3,6,22,23] the α_R conformation was predicted to be more favorable than two other nearly isoenergetic states $C7_{eq}$ and C5. Free energy minima of $C7_{eq}$ and C5 states were calculated as 7.3 ± 2.6 and 8.8 ± 3.0 kJ/mol, respectively, relative to α_R. The free energy barrier of transition from α_R to $C7_{eq}$ was calculated as 11 kJ/mol. Free energy values and confidence intervals were calculated from 10 000 to 20 000 steps (see Supporting Information). These values are in agreement with computational[3,6,22] as well as experimental studies[23] of the alanine dipeptide. It must be acknowledged that during the metadynamics run also the state α_L was sampled and essential coordinates of α_L state were overlaying with α_R. However, Gaussians corresponding to α_L do not dramatically contribute to the free energy of α_R state (by approximately 1.2 kJ/mol). This issue is illustrated in Supporting Information.

In order to test the predictive potential of metadynamics in essential coordinates a parallel metadynamics calculation was performed with a different initial conformation. The free energy surface calculated by metadynamics starting from the $C7_{eq}$/C5 conformation was in agreement with that starting from α_R. Free energy values for $C7_{eq}$ and C5 were calculated as 6.8 ± 3.8 and 8.2 ± 3.5 kJ/mol, respectively, relative to α_R. The results of metadynamics in essential coordinates starting from $C7_{eq}$/C5 conformation are presented in Supporting Information.

Also metadynamics in the space of dihedral angles Φ and Ψ was performed. The resulting free energy surface is in agreement with the free energy surface obtained by metadynamics in essential coordinates. The conformation α_R was more stable than C5/$C7_{eq}$. The free energy value of C5/$C7_{eq}$ was calculated as 5.6 ± 2.3 kJ/mol (these two conformations were not sufficiently resolved). The results are presented in the Supporting Information.

Discussion

Metadynamics brings several advantages to the exploration of free energy surfaces of molecular systems. It quantitatively evaluates free energy changes and simultaneously avoids trapping the system in free energy minima. There are also other advantages of this method such as a possibility of using multiple metadynamic walkers.[24] The key step of metadynamics is dimensionality reduction represented by a choice of collective variables. This step is often rather intuitive and a matter of experience. Cartesian coordinates of protein atoms with thousands of degrees of freedom cannot be used as collective variables because of the vast dimensionality of the problem. Recently reported metadynamics studies of conformational changes in proteins have utilized distance-dependent terms of selected pairs of amino acid residues as collective variables.[12] However, this approach requires certain experience in the studied system. The approach presented here uses a general dimensionality reduction method for a definition of collective variables. We believe that this approach permits more complex dynamics of biomacromolecules to be studied.

The disadvantage of the presented approach is due to fact that some knowledge of the dynamics of the studied system is necessary in order to determine essential coordinates. In the presented study, essential coordinates were determined by classical MDS. This method is often insufficient due to trapping the system in free energy minima, and it therefore cannot account for the full complexity of the dynamics. It was illustrated that essential coordinates obtained from relatively short MDS trajectories of compact folded proteins can provide a statistically reliable definition of an essential subspace in long timescales (>100 ns).[25] However, this is not likely to be case for more complex conformational changes or folding/unfolding processes. When a specific collective motion is not observed during the determination of essential coordinates, it can then act as a hidden degree of freedom. This was probably the reason why the α_L conformation was not resolved from the α_R conformation in the presented study.

Alternatively, other strategies may be employed to determine essential coordinates. For example, if the native structure of the studied protein/peptide is known, it is possible to determine essential coordinates from high-temperature unfolding trajectories. High-temperature unfolding trajectories were shown to approximate normal temperature folding/unfolding processes.[26] Also knowledge of experimental structures (X-ray or NMR) could be utilized for the determination of essential coordinates. Moreover, it is possible to combine the presented method with other methods that enhance sampling of FES. Future application of this approach will elucidate its advantages and disadvantages in comparison with other free energy calculation methods. The performance of the presented method when applied to a wide scope of biomacromolecules will be the focus of future research.

Conclusion

The potential of metadynamics was expanded by implementation of metadynamics in essential coordinates. This approach was tested on the conformational behavior of the alanine dipeptide. Essential coordinates were determined by means of essential dynamics analysis of a 1 ns MDS trajectory. Its free energy surface was than sampled by metadynamics with essential coordinates used as collective variables. The resulting free energy surface was in agreement with computational as

well as experimental studies of the model system. This method combines an efficient sampling and a self-healing nature of metadynamics together with a general dimensionality reduction method. The proposed approach represents a promising strategy for computational modeling of thermodynamics and kinetics of conformational changes in complex biomacromolecules.

Acknowledgment. The authors would like to thank Dr. Oliver F. Lange for the documentation of essential dynamics sampling features of the Gromacs package and Benjamin Kukull for careful reading of the manuscript. The work was supported by the Academy of Sciences of the Czech Republic (GAAV KJB 500500512) and the Ministry of Education, Youth, and Sports (MSM 6046137305).

Supporting Information Available: Illustrations addressing development of the free energy surface during the metadynamics run, the issue of sampling of α_L conformation, and development of the free energy surface during the metadynamics run starting from $C7_{eq}$/C5 conformation are available as Supporting Information. This material can be obtained free of charge via Internet at http://pubs.acs.org.

References and Notes

(1) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.

(2) van der Vaart, A. *Theor. Chem. Acc.* **2006**, *116*, 183.

(3) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.

(4) Amadei, A.; Linssen, A. B.; Berendsen, H. J. *Proteins* **1993**, *17*, 412.

(5) Lange, O. F.; Schafer, L. V.; Grubmuller, H. *J. Comput. Chem.* **2006**, *27*, 1693.

(6) Ensing, B.; de Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. *Acc. Chem. Res.* **2006**, *39*, 73.

(7) Young, M. A.; Gonfloni, S.; Superti-Furga, G.; Roux, B.; Kuriyan, J. *Cell* **2001**, *105*, 115.

(8) Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. *Cell* **2003**, *112*, 859.

(9) Meyer, T.; Ferrer, C.; Pérez, A.; Rueda, M.; Bidón-Chanal, A.; Luque, F. J.; Laughton, C.; Orozco, M. *J. Chem. Theor. Comput.* **2006**, *2*, 251.

(10) Amadei, A.; Linssen, A. B.; de Groot, B. L.; van Aalten, D. M.; Berendsen, H. J. *J. Biomol. Struct. Dyn.* **1996**, *13*, 615.

(11) de Groot, B. L.; Amadei, A.; van Aalten, D. M.; Berendsen, H. J. *J. Biomol. Struct. Dyn.* **1996**, *13*, 741.

(12) Chen, C.; Xiao, Y.; Zhang, L. *Biophys. J.* **2005**, *88*, 3276.

(13) Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V.; Gervasio, F. L.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 2705.

(14) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600.

(15) Stirling, A.; Iannuzzi, M.; Laio, A.; Parrinello, M. *ChemPhysChem* **2004**, *5*, 1558.

(16) Oganov, A. R.; Martoňák, R.; Laio, A.; Raiteri, P.; Parrinello, M. *Nature* **2005**, *438*, 1142.

(17) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.

(18) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(19) Sorin, E. J.; Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2516.

(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(21) Coutsias, E. A.; Seok, C.; Dill, K. A. *J. Comput. Chem.* **2004**, *25*, 1849.

(22) Bolhuis, P. G.; Dellago, C.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5877.

(23) Madison, V.; Kopple, K. D. *J. Am. Chem. Soc.* **1980**, *102*, 4855.

(24) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *J. Phys. Chem. B* **2006**, *110*, 3533.

(25) Lange, O. F.; Grubmuller, H. *J. Phys. Chem. B* **2006**, *110*, 22842.

(26) Day, R.; Bennion, B. J.; Ham, S.; Daggett, V. *J. Mol. Biol.* **2002**, *322*, 189.