Replica exchange molecular dynamics toolkit for drug–receptor docking

Masakatsu Ito, Takeshi Nishikawa and Umpei Nagashima

Grid Technology Research Center, National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central 2, Tsukuba, 305–8568 JAPAN; masakatsu-ito@aist.go.jp

ABSTRACT

Although the protein flexibility plays an important role in the drug–receptor docking process, the treatment of protein structural fluctuations remains a major challenge to docking simulations. To deal with this problem, we have developed a new software framework which enhances the efficiency of conformational sampling by replica exchange molecular dynamics (REMD) method. The effectiveness of the method was examined with the polypeptide molecule and it was shown that the trajectories in REMD method explored the energy region faster than the Monte Carlo moves in the ordinary replica exchange method.

Our toolkit employs object–oriented design to mask the complexity associated with biomolecules and parallelization, and to facilitate the adaptation to other software component.

Keywords: replica, exchange, MD, drug, protein

1 Introduction

In this post-genomic era, much attention is focused on the 3D structure and flexibility of protein molecules for the rational drug design. The structure information provides a good starting point for estimating the ligand–protein binding affinity, whereas the flexibility of molecules plays an important role in the docking process of a ligand into the active site of a target protein. Full consideration of such structural fluctuations is required to accelerate the screening process in drug–discovery. The statistical viewpoint of these fluctuations[1] is also essential for understanding the protein functionality which is associated with various structure changes such as collective movements of the protein backbone and side–chains during the ligand exchange process.

Nevertheless, the treatment of protein flexibility remains a major challenge. Although an accurate estimation of binding affinity can be achieved only by the rigorous basis on statistical mechanics, such sensible algorithms still have not been available in the commonly used software packages for drug–receptor docking. There are both theoretical and technical difficulties causing this situation. The computational cost becomes prohibitive large when the binding affinity is evaluated as the free energy change in the conventional Monte Carlo (MC) or molecular dynamics (MD) simulations with all atom potential energy functions. Although these potential functions, such as AMBER[2] and CHARMM[3], enable the detailed atomistic simulation, they cost much higher than the scoring functions used in most of the docking programs[4]. Moreover, software packages for biomolecular simulations are so large and complex that it is impractical for anyone but the original developers to extend or modify the code. Consequently, it often takes a long time until a new efficient algorithm becomes available to those who try to simulate biomolecules.

One possible resolution of the theoretical difficulty is suggested by exploiting the similarity of binding and folding[5]. Molecular recognition and protein folding processes share a number of statistical mechanic aspects; both processes involve accurately locating molecular fragments with respect to each other, reducing the configurational entropy, and simultaneously lowering the free energy by the exclusion of hydrophilic interactions, like hydrogen bonds and salt bridges[6]. For this description, the energy landscape theory attributes the subverted ergodicity to the ruggedness of the free energy surface that hinders the diffusion through an ensemble of partially folded states characterized by a few reaction coordinates (or order parameters). Therefore, as the generalized and extended ensemble methods have enhanced the efficiency of conformational sampling in the simulation of protein folding[7][8], these methods might be also advantageous to the docking simulation.

An obvious solution to the technical problem is software engineering. Object–oriented programming languages have demonstrated the effectiveness to handle the complexity of biomolecules and to facilitate the extension of functionalities and the addition of new algorithms, in recent simulation libraries, e.g. NAMD[9] and MMTK[10]. Such software components with high reusability are beneficial, particularly in the field of biophysics and structural biology where new simulation and analysis techniques are constantly being developed.

Therefore, we have developed a new software toolkit in C++ employing object–oriented design. The replica exchange molecular dynamics (REMD) toolkit implements the extended–ensemble method which can im-

prove the accuracy of docking simulation, structure refinement, and free energy calculation. This article gives an overview of REMD toolkit and the performance of REMD algorithm tested with polypeptide $(Ala)_{10}$.

2 REMD toolkit

Replica exchange molecular dynamics method was developed [8] to overcome the multiple–minima problem of flexible molecules such as protein folding where the conventional simulation methods fails to obtain thermodynamic quantities because of the subverted ergodicity.

For evaluating a thermodynamic quantity, REMD simulation treats a compound system consisting of N non–interacting replicas of a molecule, each at different inverse temperature β_i. A state of this system is specified by a set of molecular states $\{\mathbf{x_1}, \mathbf{x_2}, .., \mathbf{x_N}\} = \mathbf{X}$ where $\mathbf{x_i}$ is a pair of molecular coordinate vector $\mathbf{q_i}$ and momentum vector $\mathbf{p_i}$. In the canonical ensemble at β, each molecular state $\mathbf{x} = (\mathbf{q}, \mathbf{p})$ is weighted with the Boltzmann factor,

$$w_B(\mathbf{x}; \beta) = e^{-\beta H(\mathbf{q},\mathbf{p})} \quad (1)$$

where $H(\mathbf{q}, \mathbf{p})$ is the Hamiltonian of the molecule. The canonical ensemble is generated with MD algorithm while MC algorithm is used in the original version of replica exchange method (REM). Since the replicas are non–interacting with each other, the weight factor for a compound state $\mathbf{X}$ is given by the product of the Boltzmann factors for each replica,

$$W_{REM}(\mathbf{X}; \beta_1, \beta_2, \ldots \beta_N) = \exp(-\sum_i^N \beta_i H(\mathbf{x_i})), \quad (2)$$

whose summation over all states, i.e. the partition function specifies the extended ensemble of the compound system.

An extended ensemble is obtained numerically by constructing a Markov–chain until the corresponding probability distribution approaches to the thermodynamical equilibrium. Such a Markov–chain can be generated by the stochastic sampling process where the transition probability $w(\mathbf{X_{old}} \rightarrow \mathbf{X_{new}})$ satisfies the detailed balance condition. Now suppose that the transition to a new state of the system is caused by the exchange of molecular states , $\mathbf{x_i}$ and $\mathbf{x_j}$, and then the replica exchange part of transition probability is obtained as

$$w(\mathbf{X_{old}} \rightarrow \mathbf{X_{new}}) = w(\mathbf{x_i} \leftrightarrow \mathbf{x_j}) = min[1, e^{-\Delta}] \quad (3)$$

where

$$\Delta = (\beta_i - \beta_j)(H(\mathbf{x_i}) - H(\mathbf{x_j})). \quad (4)$$

Without lack of generality we can assume that $\beta_1 < \beta_2 < \ldots < \beta_N$. An actual REMD simulation is then realized by performing the following two steps alternately.

1. Each replica in the canonical ensemble is simulated simultaneously and independently for given MD steps.

2. A pair of neighboring molecular states are exchanged with the probability in Eq. (3)

2.1 Object–oriented representation of statistical mechanics concept

REMD toolkit consists of four main component objects, each being the intuitive realization of the theoretical entity in replica exchange method : (1) simulator, (2) extended ensemble, (3) ensemble, and (4) walker objects. The UML diagrams in Fig.1 and 2 describes the organization and relationships of the component objects.

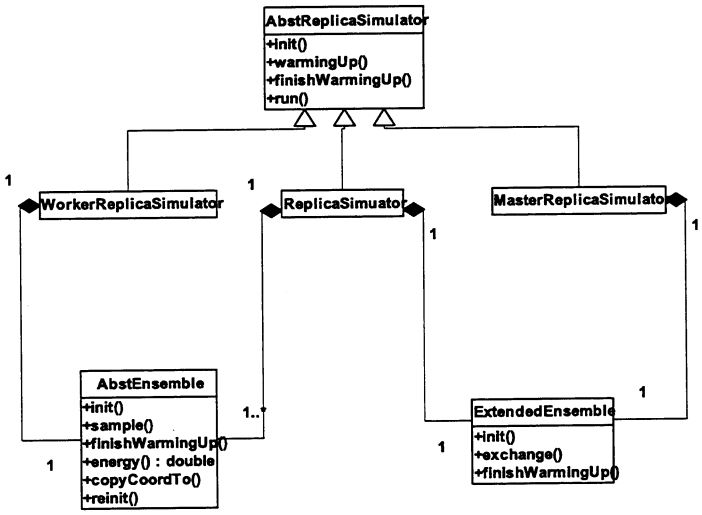

Figure 1: UML class diagram for replica exchange method. (Rectangles represent software classes, arrows indicate the generalization relationships, and the lines with diamond ends indicate composition relationships.)

Figure 1 shows a top layer of REMD toolkit. An `ExtendedEnsemble` object exchanges the molecular states of ensemble objects. `AbstReplicaSimulator` class is a generalization of three different classes, each being specialized for a different type of simulation process ; a `ReplicaSimulator` object run on a single workstation, whereas a `MasterReplicaSimulator` object and a set of `WorkerReplicaSimulator` objects are distributed over a workstation network to perform a parallelized simulation. Each worker object performs an MC or MD simulation in parallel and send the resultant energy to the master object. The master calculates the exchange probability from the collected energies, and controls the workers to exchange their molecular states each other. These three classes inherit the common interface from their base class and hide the implementation details of

MPI parallelization to assist the adaptation to other network environment, e.g. Grid.

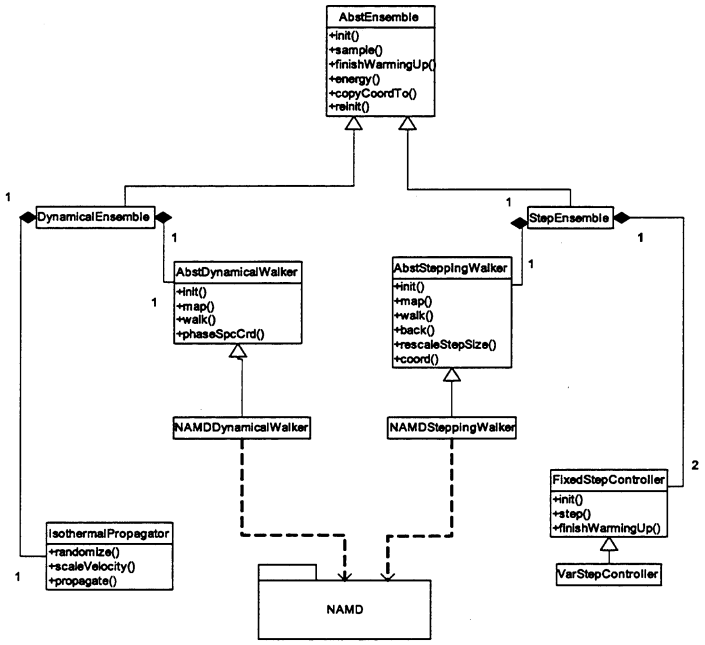

Figure 2: UML class diagram for MC and MD simulations

Figure 2 shows component classes responsible for MC and MD simulations. `AbstEnsmeble` class is the generalization of various statistical ensembles and defines the common interface to handle ensemble properties. `StepEnsemble` and `DynamicalEnsemble` class represent an ordinary canonical ensemble and an MD trajectory at constant temperature, respectively.

A molecular structure is represented by a set of coordinates which a walker object holds. `AbstSteppingWalker` is a realization of "random walker" in statistical mechanics whose MC move is controlled by `FixtedStepController` or `VarStepController`. `IsothermalPropagator` class integrates equations of motion with the Gaussian thermostatting technique[11] to move a walker of `AbstDynamicalWalker` class.

The adaptor classes to NAMD molecular dynamics package[9] also exploit the concept of inheritance and polymorphism so that they are guaranteed to cooperate with other component objects in the same way as their base classes. Thus this feature would also facilitate the extension and adaptation of walker classes to other molecular dynamics or molecular orbital packages, while maintaining all the functionality and data structure of the other components.

3 Results and Discussion

The effectiveness of REMD toolkit was examined for the system of polyalanine $(Ala)_{10}$. The force field parameters were taken from X-PLOR[12] and the code of Mindy is used to deal with the molecular topology and that potential energy function. Mindy is a serial MD program derived from NAMD package[9]. The MD time step was set to 1.0 fs and an MD simulation of 2.0×10^5 steps (100 ps) is performed for each replica. We assigned eight temperatures to the replicas : 565.7 , 475.7, 400.0, 336.4, 282.8, 237.8, 200.0, and 168.2 K, which are distributed exponentially, following the annealing schedule of simulated annealing simulations and the distribution of the first REMD simulation[8].

3.1 Energy fluctuations of polypeptide

In REMD algorithm, an MD simulation is used for generating a statistical ensemble. For conformational sampling of a large molecule, it is often more advantageous to construct a trial move that consists of a sequence of MD steps, rather than to randomize velocity vector at each steps or to construct MC moves. Since such frequent randomization shortens the velocity correlation time, the diffusion constant in the configurational space becomes quite low. Therefore it is noted that for a large molecule, an MD simulation would be more efficient for conformational sampling than a hybrid and ordinary MC simulations.

We examined this effect in the comparison between the potential energy fluctuations in the REMD simulation and those in the REM simulation. MC step size was adjusted to maintain the averaged acceptance probability around 0.3 in the preliminary simulation and was fixed while taking the data.

The fluctuations of the replica at 237.8 K are shown in Fig.3. The energy of the REMD simulation fluctuated much faster than that of the REM simulation, so that the wider range of energy region was explored along the MD trajectory. On the other hand, the step size of MC move cannot be too large, because then the acceptance would become small and energy fluctuations and the sampling efficiency would be suppressed. Moreover, the acceptance probability of MC moves with the constant step size decreases with the system size N , because the statistical energy fluctuation (root–mean–square error) increases with $N^{1/2}$. MD does not suffer from this sort of problem, that is, the numerical stability of integration algorithms does not be affected. Hence, for larger systems, the performance of REMD simulation would be better than REM.

4 Summary and Future Direction

We have developed the parallelized object–oriented toolkit for REMD and REM algorithms, whose component can be adaptable to other C++ package for biophysical simulations. The effectiveness of REMD simulation was examined with the polypeptide $(Ala)_{10}$. It

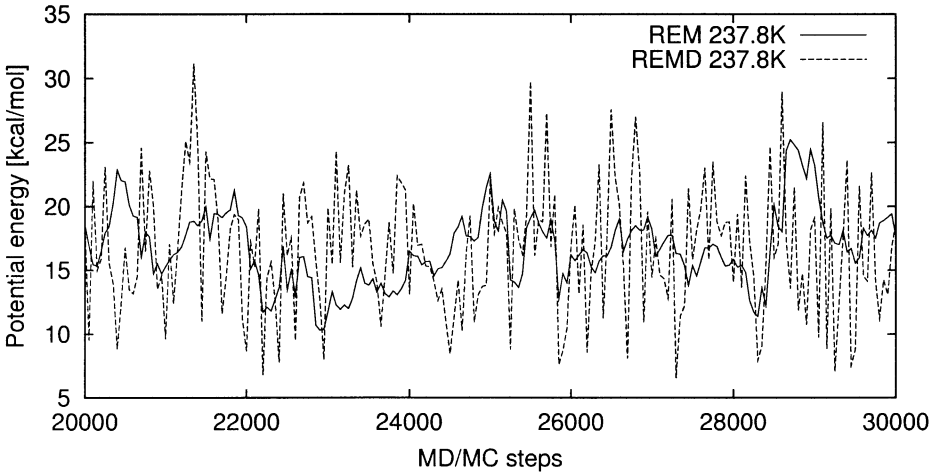

Figure 3: Potential energy fluctuations at 237.8 K in the REM (solid line) and REMD (short–dashed line) simulations

was shown that the trajectories in REMD simulation were exploring the wider range of energy region at faster rate than the MC moves in REM. Thus REMD toolkit can be used for receptor–drug docking processes where the full consideration of protein flexibility is required.

REMD toolkit is being integrated into the rational drug design system[13], which aims for the prediction of the binding affinity of several HIV protease inhibitors(Fig.4).

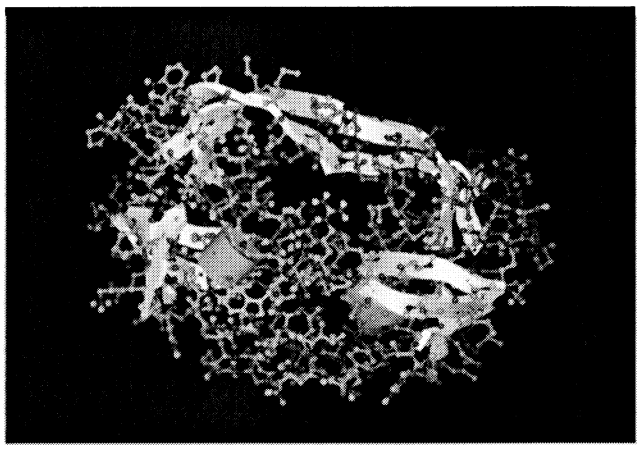

Figure 4: HIV-1 Protease complexed with Viracept(AG134)

REFERENCES

[1] B. Ma, M. Shatsky, H.J. Wolfson, R. Nussinov, Protein Sci. **11**, 184 (2002); I. Luque and E. Freire, Proteins, **S4**, 63(2000); D. M. Taverna and R. A. Goldstein, Proteins, **46**, 105(2002)

[2] P.K. Weiner and P. A. Kollman, J. Comp. Chem. **2**, 287 (1981)

[3] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comp. Chem. **4**, 187 (1983)

[4] G. Jones, P. Willet, R.C. Glen, A.R. Leach, R. Taylor, J. Mol. Biol. **267**, 727 (1997); B. Kramer, M. Rarely, T. Lengauer, Proteins, **37**, 228 (1999); T.J. Ewing, S. Makino, A.G. Skillman, I.D. Kuntz, J. Comput. Aided Mol. Des. **15**, 411, (2001)

[5] P. A. Rejto, and G. M. Vehrkhivker, Proc. Natl. Acad. Sci. USA, **93**, 8945 (1996) ; G. M. Verkhivker, D. Bouzida, D. K. Gehlhaar, P.A. Rejto, S. T. Freer , and P. W. Rose , Cuur. Opin. Struct. Biol. **12**, 197 (2002)

[6] J. H. Ha, R. S. Spolar, and M.T. Jr. Record, J. Mol. Biol. **209**, 801 (1989) ; R. S. Spolar, M.T. Jr. Record, Science, **263**, 777 (1994)

[7] U. H. E. Hansmann, Chem. Phys. Lett. **281** 140 (1997)

[8] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999)

[9] L Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Grusoy, N. Karawets, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten, J. Comp. Phys. **151**, 283 (1999)

[10] K. Hinsen, J. Comp. Chem. **21**, 79 (2000)

[11] F. Zhang, J. Chem. Phys. **106**, 6102 (1997)

[12] A. T. Brünger , X–PLOR, version 3.1, A System for X–Ray Crystallography and NMR, The Howard Hughes Medical Insititute and Department of Molecular Biophysics and Biochemistry, Yale University, 1992 .

[13] H. Chuman, C. Feng, Y. Inagaki, T. Nishikawa, U. Nagashima, H. Goto, T. Boku, and M. Sato , 6th World Conference of the Theoretcally Oriented Chemists (WATOC), Abstract Book, PC448, 2002.