Analysis of the Function of a Large-scale Supra-biomolecule System by Molecular Dynamics Simulation System, SCUBA (Simulation Codes for hUge Biomolecular Assembly)

Project Representative
Atsushi Matsumoto Japan Atomic Energy Agency

Authors
Hisashi Ishida Japan Atomic Energy Agency
Kei Yura Ochanomizu University
Atsushi Matsumoto Japan Atomic Energy Agency

The Earth Simulator has the highest power ever achieved to perform molecular dynamics simulation of large-scale supra-molecular systems. We are developing a molecular dynamics simulation system, called SCUBA, which is designed to run a system composed of more than a million particles efficiently on parallel computers. In order to understand how MutS recognizes mismatched DNA molecules and induces the repair at an atomic level, long-time molecular dynamics simulations of MutS in water with a verity of mismatched DNA molecules were carried out. The dynamics of the MutS-mismatched DNA complexes was analyzed.

Keywords: large-scale supra-biomolecular MD simulation, mismatch repair, MutS protein, mismatched DNA

1. Introduction

Molecular dynamics (MD) simulation not only provides dynamic descriptions of molecules on the atomic scale, but also provides valuable information for the interpretation of experimental data. The rapid development of computer power and the elucidation of the structures of biological macromolecules by X-ray crystallography and other experiments have increased the need for large-scale MD simulations in the field of biology.

We are developing an integrated molecular simulation system for biological macromolecules, called SCUBA (Simulation Codes for hUge Biomolecular Assembly), which is designed to run a system composed of more than a million particles efficiently on parallel computers [1].

SCUBA has several special features:

1. Topology of biomolecules
   The structure of SCUBA's program code is optimized for a system in which the topology of a biomolecular structure is considered.
2. A variety of force field parameters
   At present SCUBA can use the AMBER [2], CHARMM [3] and GROMOS [4] force field. The parameter file for the topology of biomolecules can be obtained from the PDB file by using the input module in SCUBA.
3. A variety of simulation methods
   A variety of simulation methods, such as energy minimization, molecular dynamics, free energy calculations, normal mode analysis, principal component analysis and so on, are included.
4. Algorithm for non-cutoff electrostatic interactions
   SCUBA utilizes the Particle-Particle Particle-Mesh (PPPM) algorithm, which efficiently calculates all the Coulomb electrostatic interactions [5]. This algorithm reduces the computational time required to calculate the electrostatic forces from the conventional $O(N^2)$ to O(Nlog$N$).
5. Input and output compatibility
   SCUBA's input and output file format is currently compatible with those used by AMBER [2].
6. Portability
   Written in Fortran90, SCUBA is designed to be easy to read, modify and extend. Users can easily maintain the existing code, and develop the current algorithms and integrate new ones.
7. Control files
   The control files for running SCUBA are described in a user-friendly manner.
8. Time-integral algorithms with high accuracy
   The Martyna-Klein-Tuckerman (MKT) algorithm which produces the correct ensemble thermodynamically is available. The MKT algorithm was extended to utilize the multiple time step (MTS) method, which increases the

time step length significantly [6, 7].

9. Time-integral algorithms for long time steps

SHAKE, and RATTLE which allow the time step taken to be larger by fixing the bond lengths and angles in the system are available. In this fisical year, SHAKE and RATTLE were incorporated in the MKT algorithm, which further increases the time step while maintaining the stability of the system. The user can freely choose the options which control the multiple time steps for a variety of interaction forces.

10. Analysis of ourput files

In this fisical year, SCUBA's function to analyse the trajectories obtained from MD simulations has been enhanced. Now, the analysis module in SCUBA can analyze a variety of physical quantities such as interaction energies and structural parameters from the trajectories.

11. Parallelization

SCUBA employs the domain decomposition (DD) method, which divides the volume of the physical system into rectangular subcells with a length longer than the potential cutoff radius. The processor assigned to a subcell needs to evaluate the interactions between the atoms in the subcell and between the atoms in 26 neighboring subcells. SCUBA employs the method for minimizing communication between processors proposed by D. Brown [8], which enables the number of processors between which data must be transferred to be reduced to only 7 of the neighboring subcells.

12. Vectorization

In order to improve the performance of SCUBA on the Earth Simulator, the algorithm to calculate the interactions among the atoms is intensively vectorized.

13. Dynamic load balance

To overcome the load imbalance associated with irregular atomic distribution, a dynamic load-balancing algorithm is implemented. Moreover, the number of processors used to calculate the PM part of PPPM can be optimized to minimize the computation by allowing the number to change.

14. High performance

By intensive parallelization and vectorization, and by using the dynamic load balance mentioned in 10, 11, and 12, SCUBA has achieved both a high parallelization efficiency ratio and a high vectorization ratio. SCUBA has achieved a parallelization efficiency ratio of 75.8 %, and a vectorization ratio of 96.2 % even 45 nodes (360 processors) were used to perform an MD simulation for a system of RuvAB-Holliday junction complex which consisted of 546,725 atoms. The performance of SCUBA on the Earth Simulator is shown in Fig. 1.

15. Optimization of memory use

The arrays used in the program of SCUBA are intensively optimized to reduce the amount of memory use. This optimization enables SCUBA to perform molecular dynamics simulations of large-scale supra-molecular systems comprised of more than a million atoms on the Earth Simulator.

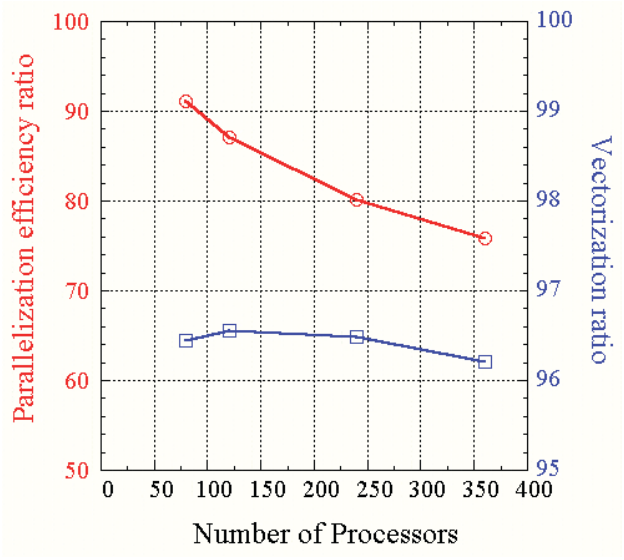

Fig. 1 Present performance of SCUBA on the Earth Simulator. The parallelization efficiency ratio and vectorization ratio are drawn in red and blue, respectively.

2. Molecular dynamics simulation of MutS

2.1 Introduction

DNA mismatch repair (MMR) maintains genome stability by repairing mismatches that arise through DNA replication errors and during recombination. Defects in MMR result in a significant increase in the spontaneous mutation rate and predispose humans to cancer.

In *E. coli*, the proteins MutS, MutL and MutH are responsible for the MMR. MMR is initiated by MutS, which functions in the homodimer form. MutS recognizes and efficiently binds to mispaired bases such as G-T, A-C, G-G, A-G, A-A, T-T, T-C, or unpaired bases in DNA duplexes (see Fig. 2). It is thought that the ATPase activity of MutS plays a role in proofreading to verify mismatch binding and authorize the downstream excision [9]. MutL, which functions in the homodimer form, interacts with the MutS-mismatched DNA complex. The MutS-MutL-mismatched DNA complex recruits and activates the endonuclease MutH. MutH, which functions as a monomer, removes the mismatch and replaces the mismatch with the correct base.

2.2 Materials and Methods

In our study, to understand how MutS recognizes a mismatched DNA and triggers repair, molecular dynamics (MD) simulations were performed on MutS in water with and without the mismatched DNA. We used six different

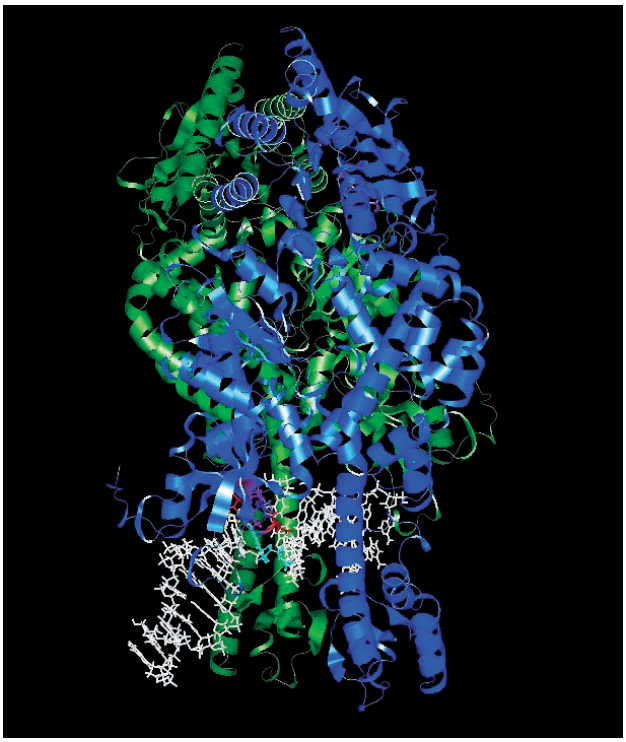

Fig. 2 The structure of MutS. The MutS dimer is shown in blue (subunit A) and green (subunit B). MutS shown in blue interacts with the mismatch region of the mismatched DNA. The mismatch base-pair is shown in red. The rest of the DNA is shown in white.

system of *Escherichia. coli* MutS in bound state with ADP and a system in bound state with ATP. The five ADP-bound state systems are those with G-T, C-A, A-A, G-G mismatches and with a single unpaired thymidine. The other ATP-bound system is that with G-T mismatch. The atomic structures for these six systems were determined by X-ray crystallography (PDB code: 1E3M [10], 1OH5-8 [11], 1W7A [12] respectively). Using SCUBA, we have carried out MD simulations of the MutS dimer with and without the mismatched DNA. Each system comprised about 200,000 atoms including water molecules. The MD simulations were carried out at a constant pressure of one bar and a temperature of 300 K for several tens of nanoseconds in total.

2.3 Results

We analyzed the conformation of the MutS dimer complexed with and without the mismatched DNA. In order to check the stability of the conformations of the MutS-mismatched complexes and the uncomplexed MutS in the MD simulations, the root mean square of the atomic displacements (RMSD) of the atomic positions of the systems with respect to the energy-minimized coordinates was calculated during the MD runs. The conformations of MutS-mismatched complexes were stable, and the RMSD values were about 3Å. On the contrary, the RMSD value for the uncomplexed MutS was more than 3Å. These results indicate that the complexed MutS fully encircles the mismatched DNA by DNA-binding domains from two MutS molecules, maintaining the structures stabley, while the uncomplexed MutS molecules were mobile.

In order to analyze the flexibility of the MutS-mismatched complexes, the averaged root mean square of the fluctuations (RMSF) for all the heavy atoms of the complexes were calculated. In general, the thermal fluctuation was rather low at domain I (residues 1 – 115) of subunit A, which lodges in the widened minor groove where the mismatched base is recognized. The thermal fluctuation at domain IV (residues 444 – 503) of subunit B, which stabilizes the kinked DNA by interactions with the opposite narrowed major groove, was also low. Domains II (residues 116 – 266) and III (residues 267 – 443 + 504 – 567), which form connecting domains and are called a lever arm, was moderately mobile. This flexibility may be important for activating ATPase of MutS thorough the DNA binding domain after MutS recognizes mismatched DNA.

Although a similar tendency in the thermal fluctuations mentioned above was observed in all MutS - mismatched DNA complexes, we observed differences in the dynamics of the disordered loop (residues 661 – 667) of domain V (residues 568 – 800) of subunit B and the variable loop in both subunits between ADP-bound and ATP-bound states. The disordered loop region is thought to be essential for MMR, as the structural ordering of this loop, depending on the type of the nucleotide, may provide inter-subunit interactions between subunits A and B, and may play an important role in regulating mismatch binding and in mediating the interaction between MutS and MutL [9]. Therefore, the difference in the dynamics of the loop observed in the MD simulations may imply a subtle adjustment of MutS to activate the next step of MMR.

3. Conclusion

We developed an MD simulation system,SCUBA, which achieves a high parallelization efficiency ratio and a high vectorization ratio on the earth simulator. A long-time simulation of MutS for several tens of nanoseconds was successfully carried out. Our simulations of MutS showed which domains of MutS were flexible. Next we are going to analyse how the structurally separate regions of the binding site of the mismatched DNA and the activation site of the ATPase are functionally related to each other, and then plan to analyse how the activated MutS induce interaction with MutH at the next step of MMR.

References

[1] H. Ishida and S. Hayward, Path of nascent polypeptide in exit tunnel revealed by molecular dynamics simulation of ribosome, *Biophysical J*. **95**, 5962–5973 (2008).

[2] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.R. Ross, T.E. Cheatham, III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp. Phys. Commun*. **91**, 1–41 (1995).

[3] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations, *J. Comp. Chem*. **4**, 187–217 (1983).

[4] W. F. van Gunsteren and H. J. C. Berendsen, GROMOS: GROningen MOlecular Simulation software. Technical report, Laboratory of Physical Chemistry, University of Groningen, Nijenborgh, The Netherlands, (1988).

[5] R.W. Hockney and J.W. Eastwood, Computer Simulation Using Particles. McGraw-Hill, NY, (1981).

[6] G. J. Martyna, M. E. Tuckerman, D. J. Tobias and M. L. Klein, Explicit reversible integrators for extended systems dynamics, *Mol. Phys*. **87**, 1117–1157 (1996).

[7] R. Zhou, E. Harder, H. Xu and B. J. Berne, Efficient multiple time step method for use with Ewald and particle mesh Ewald for large biomolecular systems, *J. Chem. Phys*. **87**, 1117–1157 (1996).

[8] D. Brown, J. H. R. Clarke, M. Okuda and T. Yamazaki, A domain decomposition parallelization strategy for molecular dynamics simulations on distributed memory machines, *Comp. Phys. Commun*. **74**, 67–80 (1993).

[9] M. S. Junop, G. Obmolova, K. Rausch, P. Hsieh and W. Yang, Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair, *Mol Cell* **7**, 1–12 (2001).

[10] M. H. Lamers, A. Perrakis, J. H. Enzlin, H. H. Winterwerp, N. De Wind, T. K. Sixma, The Crystal Structure of DNA Mismatch Repair Protein Muts Binding to a G·T Mismatch, *Nature* **407**, 711–717 (2000).

[11] G. Natrajan, M.H. Lamers, J.H. Enzlin, H.H. Winterwerp, A. Perrakis, T.K. Sixma, Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates, *Nucleic Acids Res*. **31**, 4814–4821 (2003).

[12] M.H. Lamers, D. Georgijevic, J.H. Lebbink, H.H. Winterwerp, B. Agianian, N. de Wind, T.K. Sixma, ATP increases the affinity between MutS ATPase domains, Implications for ATP hydrolysis and conformational changes, *J. Biol. Chem*. **279**, 43879–43885 (2004).

分子動力学シミュレーションシステムSCUBA(Simulation Codes of hUge Biomolecule Assembly)を用いた大規模生体超分子系の機能解析

プロジェクト責任者
松本　淳　日本原子力研究開発機構
著者
石田　恒　日本原子力研究開発機構
由良　敬　お茶の水女子大学
松本　淳　日本原子力研究開発機構

地球シミュレータは従来にはない大規模生体超分子系の分子動力学シミュレーションを可能とする計算能力をもつ。我々は生体超分子系を扱う大規模な分子動力学シミュレーションシステムSCUBAを開発している。SCUBAは長距離相互作用を高速かつ高精度に計算するPPPM計算法や温度、圧力を一定に保つ様々な時間積分アルゴリズムなど、最新のアルゴリズムを採用した計算性能に優れたシミュレーションシステムである。SCUBAは機能別にモジュール化することでメモリ使用量を最適化しており、百万原子以上からなる系でも分子動力学シミュレーションが実行可能である。現在SCUBAは、地球シミュレータ360プロセッサ使用時でベクトル化率95%以上、並列化効率50%以上の優れた性能を達成している。本年度は、短距離相互作用と長距離相互作用の変化に対応する時間スケールを切り分けることで、時間ステップを大きくするマルチ時間ステップ法をSCUBAに装備した。さらに、SHAKE、RATTLE法をマルチ時間ステップ法に組み入れることで、系の安定性を向上させた。これにより長時間にわたる高精度なシミュレーションが可能となった。

本年度は、ミスマッチDNAに結合してミスマッチ修復を誘導するタンパク質MutSの系について長時間にわたる分子動力学シミュレーションを地球シミュレータ上で実行した。様々なミスマッチDNAと結合したMutSとの複合体の系についてMutSの動的構造を解析した結果、MutSがミスマッチを認識する部位とMutSがミスマッチ修復を誘導する部位におけるダイナミクスが互いに関連していることが示唆された。

キーワード：大規模生体超分子動力学シミュレーション, 長時間シミュレーション, DNA修復, MutSタンパク質, ミスマッチ