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
Hisashi Ishida Japan Atomic Energy Agency

Authors
Hisashi Ishida *1, Mariko Higuchi *1, Yoshiteru Yonetani *1, Takuma Kano *1, Yasumasa Joti *2, Akio Kitao *2 and Nobuhiro Go *1

*1 Japan Atomic Energy Agency
*2 The University of Tokyo

The Earth Simulator has the highest power ever achieved to perform molecular dynamics simulation of large-scale supra-molecular systems. Now, 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.

To perform large-scale simulations rapidly and accurately, state-of-the-art algorithms, such as Particle-Particle Particle-Mesh (PPPM) which can compute the long-range Coulomb interaction accurately and efficiently, and dynamic load balance algorithm to optimize the load balance among the processors, are implemented. SCUBA uses a domain decomposition method which can achieve high parallelization. Moreover, and algorithm to calculate the interactions among atoms are intensively vectorized in order to improve the performance of SCUBA on the Earth Simulator. A benchmark test was carried out using the system of a RuvAB-Holliday junction complex which consisted of 546,725 atoms. At present, SCUBA has achieved a parallelization efficiency ratio of 75.8%, and a vectorization ratio of 96.2% even when 45 nodes (360 processors) are used. Finally, in order to elucidate how RuvA recognizes the Holliday junction and how branch migration occurs at the molecular level, molecular dynamics simulation of RuvA-Holliday junction DNA complex was performed.

Keywords: Large-scale supra-biomolecular MD simulation, Dynamic load balance, PPPM method, Holliday junction

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.

One way to achieve large-scale simulations is to utilize parallel supercomputers such as the Earth Simulator. The two main strategies that have been used in the parallelization of MD programs are called particle decomposition (PD) and domain decomposition (DD). In the PD method all processors know all the coordinates of all the particles in the system. MD programs such as AMBER [1], CHARMM [2], GROMOS [3] which were originally designed for serial computers, have been adapted for use on parallel computers by using this method. Adapting these programs in this way is straightforward from the point of view of coding; however, this method requires extensive communication between the processors because it collects the data for all the coordinates from all the processors and distributes the particles' new coordinates to all the processors. In the DD method, which we have employed, the volume of the physical system is divided into rectangular subcells with a length longer than the potential cutoff radius. The main advantage of this method is that each processor communicates with a limited number of neighboring processors. The DD method, therefore, offers the best theoretical scalability of both memory use and the amount of communication between the processors used for MD simulations of large molecular systems on parallel computers.

We are developing an integrated molecular simulation system for biological macromolecules, called SCUBA (Simulation Codes for hUge Biomolecular Assembly), (pre-

viously called PABIOS) which is designed to run a system composed of more than a million particles efficiently on parallel computers.

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 both the AMBER force field and the CHARMM force field. The parameter file for the topology of biomolecules can be obtained from the PDB file by using the parameter module in SCUBA.
3. A variety of simulation methods
   A variety of simulation methods, such as energy minimization, molecular dynamics, free energy calculations by free energy perturbation and thermodynamic integration, 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 [4]. This algorithm reduces the computational time required to calculate the electrostatic forces from the conventional $O(N^2)$ to O(NlogN).
5. Input and output compatibility
   SCUBA's input and output file format is currently compatible with those used by AMBER [1] and PRESTO [5] (It will be compatible with CHARMM format in the future). By using the same molecular dynamics output format, users can utilize the many analysis algorithms provided by SCUBA, AMBER and PRESTO.
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 for long time steps
   In this fiscal year, we have implemented SHAKE, and RATTLE which allow the time step taken to be larger by fixing the bond lengths and angles in the system.
9. Time-integral algorithms with high accuracy
   In this fiscal year, we have implemented the Martyna-Klein-Tuckermann (MKT) algorithm which produces the correct ensemble thermodynamically.
10. High performance
    SCUBA achieves both a high parallelization efficiency ratio and a high vectorization ratio. The techniques employed to achieve such a high performance are explained below.
11. Reducing the amount of communication
    In the DD method, 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. PABIOS employs the method for minimizing communication between processors proposed by D. Brown [6], 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
    Last fiscal year, to overcome the load imbalance associated with irregular atomic distribution, we implemented a dynamic load-balancing algorithm. However, the number of processors used to calculate the PM part of PPPM had to be fixed before the simulation, and the optimum number to minimize the computation had to be surveyed manually. This fiscal year, we have enhanced the algorithm of the dynamics load balance to optimize the number of processors used to calculate the PM part of PPPM by allowing the number to change during the simulation.

Next, Performance of SCUBA on the Earth Simulator is described. We have carried out a benchmark test of SCUBA on the Earth Simulator. The physical system for this test was chosen to be RuvAB-Holliday junction complex, as shown in Figure 1. The size of the system was 274.0 Å × 142.2 Å × 148.9 Å, and the cutoff length for the van der Waals interactions was chosen to be 8 Å. On the basis of the cutoff length the system was divided into 31 × 16 × 16 = 7,936 subcells.

A parallelization efficiency ratio of 75.8 % was achieved even when 45 nodes (360 processors) were used as shown in

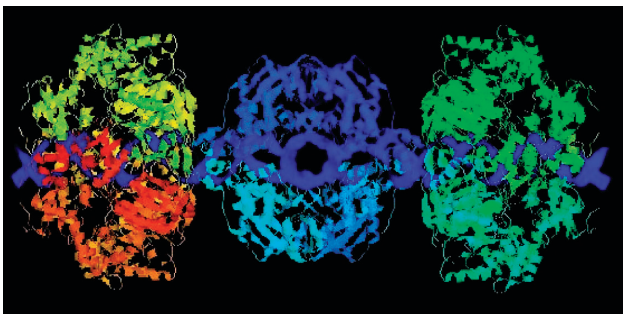

Fig. 1 The system for the benchmark test is RuvAB - Holliday junction complex, which is a biomolecular complex consisting of four strands of DNA and twenty protein molecules and executes recombination of homologous DNA strands. The size of the system and the number of atoms in the system are 274.0 Å × 142.2 Å × 148.9 Å and 546,725 atoms, respectively.

Figure 2. Last fiscal year, the parallelization efficiency ratio was less than 50.0% when more than 15 nodes were used to perform an MD simulation for a system which consists of 166,177 atoms. It shows that the computational performance of SCUBA improves as the number of atoms in the system increases. The vectorization ratio of 96.2% was achieved even when 45 nodes were used due to the intense vectorization for the calculation of atomic interactions.

To elucidate how RuvA maintains the Holliday junction, conventional molecular dynamics simulations are performed both for Holliday junction DNA complexed with RuvA tetramer and for an uncomplexed Holliday junction DNA. Only in the simulation of complex, it is observed that the initial X-shaped form of Holliday junction DNA is maintained and the hole at the junction center is kept wide, indicating that the HhH motifs and the acidic pins of RuvA tetramer keep the junction center wide and mobile to facilitate branch migration. Next, to see how branch migration occurs, umbrella sampling simulations are performed for complexes of RuvA tetramer and Holliday junction DNA. In the initial state of simulation taken from the X-ray structure of the complex, hydrogen bonds between two base-pairs are disconnected. As the branch migration starts, hydrogen bonds between the last two base-pairs of the shrinking stems become additionally disconnected. Then, four out of eight bases from four unpaired base-pairs become to interact with acidic pins from RuvA. Finally these four bases move away from pins and form hydrogen bonds of two new base-pairs of growing stems. The free energy profile along this reaction path is analyzed. Two free energy barriers are observed before and after the metastable state where four bases interact with pins. Even though the accuracy of the computed free energy barrier is limited due to sampling inefficiency, it is estimated to be about 10~15 kcal/mol.

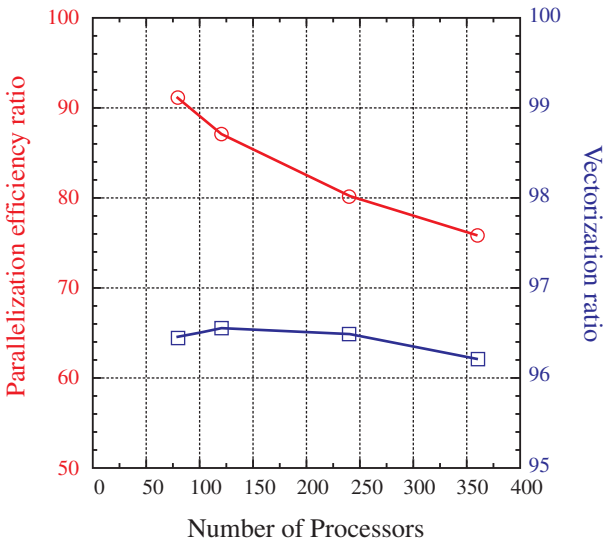

Fig. 2 Present performance of SCUBA on the Earth Simulator. The parallelization efficiency ratio and vectorization ratio are drawn in red and blue, respectively.

In conclusion, we have developed an MD simulation system for large-scale supra-biomolecules, SCUBA, which achieves a parallelization efficiency ratio of more than 50.0% and a vectorization ratio of more than 95.0% even when 45 nodes (360 processors) are used. SCUBA was used to perform an MD simulation of the Holliday junction and the dynamic properties of the Holliday junction were analyzed.

Bibliographies

1) 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).
2) 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).
3) 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).
4) R.W. Hockney and J.W. Eastwood, Computer Simulation Using Particles. McGraw-Hill, NY, (1981).
5) K. Morikami, T. Nakai, A. Kidera, M. Saito and H. Nakamura. PRESTO : A vectorized molecular mechanics program for biopolymers. Comput. Chem. 16, 243–248 (1992).
6) 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).

分子動力学シミュレーション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
石田　恒　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

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。そこで、我々は数百万原子の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（旧名PABIOS）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、系のエネルギー、温度、圧力を一定に保つ様々な時間積分アルゴリズム、原子結合長を固定することにより時間ステップの増大を可能とするSHAKE, RATTLEアルゴリズム、系の原子分布異方性により引き起こされる並列化効率の悪化を防ぐための動的ロードバランスなど、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は空間分割法による高い並列化効率をもち、かつ地球シミュレータ上で高速計算をするためのベクトルチューニングがなされた。本年度は、約55万原子の系（RuvAB-Holliday分岐DNA）について分子動力学シミュレーションを実行した結果、地球シミュレータで360個のプロセッサを用いてもSCUBAはベクトル化率95％以上、並列化効率50％以上の優れた性能を達成することに成功した。そして、ホリデイジャンクションモデル（RuvA-Holliday分岐DNA）の長時間分子動力学シミュレーションを実行することで、Holliday分岐DNA結合タンパク質RuvAがDNAを組み換えるメカニズムを分子レベルで明らかにすることができた。

キーワード：大規模生体超分子動力学シミュレーション, 空間分割法, PPPM法, 動的ロードバランス, ホリデイジャンクション