

# Activity Report on Bio Molecular Simulation Consortium

Consortium Representative

**Toshikazu Takada** Fundamental and Environmental Research Laboratorie, NEC Corporation

Author

**Toshikazu Takada** Fundamental and Environmental Research Laboratorie, NEC Corporation

## 1. Outline of Consortium

Bio Molecular Consortium was established in March, 2003. The main mandate of the consortium is to evaluate proposals proposed by the members of the consortium for Earth Simulator projects among themselves from the professional point of view and then to recommend proper proposals with certification of the consortium as candidates of Earth Simulator projects. As mentioned below, through the evaluations, these proposals are brushed up to have a high quality of research as projects on Earth Simulator. The number of the members is now more than 40 and they are the top scientists in this field of Japan.

## 2. Goals of Consortium

As an important research area, bio technology is being recognized recently. Thus, our objectives are:

- 1) To understand the mechanism of efficient biological systems at molecular levels
- 2) To apply knowledge obtained from simulations to industrial activities such as drug and enzyme design

Basically, all the codes of bio molecular simulations used here are originally developed in order to achieve these objectives with high performances on Earth Simulator and related simulations are carried out to demonstrate their usefulness as the second objective.

Another is to recommend proposals to Earth Simulator, which are reviewed to be proper projects to be carried on it. To be distinct, these proposals are sent to all the consortium members for peer review. Then, after such selections, the chair will recommend them with certification of the consortium. Therefore, the research quality of the proposals which are recommended, are kept to be high from the professional point of view.

## 3. Activities of Consortium

As mentioned before, the major activity of this consortium is to recommend bio-related material simulations to Earth Simulator. The other is to provide these researchers with a place for sharing information of simulations related to biology, since these areas have been developing rather independently. For detailed understanding of the mechanism of biological systems, collaborations by these researchers in

different areas such as genome informatics, protein folding, molecular dynamics and so on, are clearly needed. In the consortium, this is achieved by circulating their proposals before applying to Earth Simulator.

## 4. Sub-Themes of Project

This year, the following four projects were selected:

- 1) "Protein Folding Simulations from the First Principles", directed by Yuko Okamoto of Nagoya University
- 2) "Large Scale MD Simulations of Proteins on the Earth Simulator: MD simulations of hemoglobin in water", directed by Minoru Saito of Hirosaki University
- 3) "Analysis of the Function of a Large-scale Supra-biomolecule System by Molecular Dynamics Simulation System, SCUBA (Simulation Codes for hUge Biomolecular Assembly)", directed by Hisashi Ishida of Japan Atomic Energy Agency
- 4) "A Novel Phylogenetic Classification Method of Genomic Sequence Fragments from Uncultured Microbe Mixtures in Environmental and Clinical Samples", directed by Takashi Abe of National Institute of Genetics and SOKENDAI

The activity reports of these researches are given below.

## 5. Cross Relationship of Sub-Themes

To understand mechanism of biological systems at molecular levels, it is needed to hybridize many of individual theories and computational technologies which have been developed in separated areas such as genome informatics, protein folding, molecular dynamics simulation and so on. Therefore, Bio Molecular Simulation Consortium now conducts 4 sub-themes given above as the Earth Simulator projects.

### <Sub-Theme 1>

In this sub-theme, a powerful molecular dynamics simulation algorithm called Replica-Exchange Molecular Dynamics (REMD) has been applied to the ab initio folding simulation of a rather small protein, protein G, in explicit water molecules. The numbers of amino acids and surrounding water molecules are 56 and 17,187. The total number of replicas is 224 in this simulation using successfully 112 nodes of Earth Simulator. A more powerful simulation algo-

rithm (MUCAREM) than REMD for folding simulation of protein G in explicit water are ready in order to further enhanced sampling.

#### <Sub-Theme 2>

The purpose is to computationally demonstrate and visualize the structural changes of hemoglobin using COSMOS90 which is able to efficiently simulate protein motions in water with all degrees of freedoms and long-range Coulomb interactions. Using COSMOS90, molecular dynamics simulations of hemoglobin in water are carried out this year. That is, 85 simulation jobs of hemoglobin were submitted to the earth simulator and executed in about four months. The obtained results are consistent with the experimental hypothesis, i.e., the subunit  $\alpha 1\beta 1$  relatively rotates against to the  $\alpha 2\beta 2$  subunit like two stack dumbbells.

#### <Sub-Theme 3>

A molecular dynamics simulation system, called SCUBA, is now under development, which is designed to handle more than one million particles efficiently on parallel computers. To perform such large-scale simulations, Particle-Particle-Mesh (PPPM) which is able to treat long-range Coulomb interactions accurately and efficiently, is adopted. A benchmark is taken using the system of a RuvAB-Holliday junction complex consisted of 546,725 atoms. 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 is underway.

#### <Sub-Theme 4>

Self-Organizing Map (SOM) is an effective tool for clustering and visualizing high-dimensional complex data by mapping onto a two-dimensional map. The conventional SOM to genome informatics is modified, by making the learning process and resulting map be independent of the order of data input, and a novel bioinformatics tool is developed for phylogenetic classification of sequence fragments which are obtained from pooled genome samples of microorganisms in environmental and clinical samples. The SOM to classify sequence fragments derived from the Sargasso Sea near Bermuda is used, resulting the Sargasso sequences are effectively visualized on a single map.

These projects are considered to be leading edges.

### 6. Future Plans and Scopes

Recently, Bio and Nano Simulations are expected to produce next generation technologies for material developments in industries, which will be considered to be safe and costless for the environments. To accomplish this, computer codes are to be developed in Japan originally. This is one of the goals of Bio Molecular Simulation Consortium.

For it, it is strongly required to hybridize those which have different functionalities using component based programming. One issue is how to share data generated by the components, for example, with XML. Thus, it is clearly needed to have collaborations between computer scientists and computational scientists.

# 平成17年度地球シミュレータ研究プロジェクト成果報告書

コンソーシアム責任者

高田 俊和 日本電気(株)基礎・環境研究所

著者

高田 俊和 日本電気(株)基礎・環境研究所

## 1. コンソーシアムの説明

バイオシミュレーション研究者の会は、2003年3月に設立された任意の団体である。設立の目的は、地球シミュレータへの申請課題が、当該分野の専門家による相互の事前評価を通じ、地球シミュレータでの計算に相応しい充分高い研究内容であることを確認し、研究者の会として、一括して申請を行うことである。以下で述べるように、申請課題については、会員間で相互レビューを厳格に行っており、研究内容の水準の確保には充分注意を払っている。現在、会員数は40名を越しており、生体分子シミュレーションに関わる国内の研究者の多くが会員として登録されている。

## 2. コンソーシアムの目的

次世代の重要分野として、バイオテクノロジーが注目されている中、

- 1) 生体機能の発現メカニズムを、分子レベルで明らかにする。
- 2) それらの知見・技術を、創薬など産業上の活用に結びつける。

ことを目的として、必要な生体分子シミュレーション技術の構築とプログラムの自主開発を進め、それらの有効性を実証するための大規模計算を地球シミュレータで行う。また、次の活動内容で述べるように、バイオ分野の地球シミュレータへの課題申請の際、その課題の研究水準が専門的見地からみても充分高いことを、コンソーシアム内でのサーキュレーションにより事前に確認し、コンソーシアムとして地球シミュレータに推薦することも、重要な活動目的のひとつである。

## 3. コンソーシアムの活動内容

バイオシミュレーション研究者の会の主たる活動内容は、次の二つである。第1の活動は、例年行われる地球シミュレータへの課題申請の際、その研究内容が専門的見地から見ても充分高いものであり、且つ地球シミュレータでの計算に相応しいことを、事前審査することである。そのプロセスは、1) 課題申請希望者は、決められたフォーマットに研究内容を記載し、バイオシミュレーション研究者の会の事務局に提出する、2) 事務局より、全ての会員に提案書が送付され、会員のレビューを受ける、3) 会員の支持を受けた研究課題のみが、地球シミュレータへサブテーマとして申請される、である。このように会員同士の厳しい相互レビューの下、申請が行われている。

第2の活動は、研究者間の情報交換の場の提供である。生体分子の発現メカニズムを解明するには、ゲノム情報の解析、蛋白質のフォールディング、蛋白質の熱運動解析、電子状

態の計算など多岐にわたる技術が必要である。これらの計算手法は、これまで独立な研究分野として発展してきた経緯があるが、これらの分野の技術交流の活性化による相互認識の向上を図る必要がある。申請課題の相互レビューは、この点においても充分役割を果たしていると考えている。

## 4. サブグループ名

平成17年度は、以下の4課題が、サブテーマとして実施されている。

サブテーマ1:

「第一原理からのタンパク質の折り畳みシミュレーション」

サブテーマ責任者 岡本祐幸 名古屋大学

サブテーマ2:

「蛋白質の高次構造変化のリアルなシミュレーション」

サブテーマ責任者 斎藤 稔 弘前大学

サブテーマ3:

「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」

サブテーマ責任者 石田 恒 日本原子力研究所

サブテーマ4:

「環境中ならびに臨床検査サンプル中の微生物混合ゲノムに由来するゲノム断片配列についての新規な系統分類法」

阿部貴志 国立遺伝学研究所

次のセクションで、これらのテーマの目的などについて、述べる。

## 5. グループ間の連携

生体分子の機能発現のメカニズムを解明するのは、これまで独立した研究分野として発展してきている多方面の技術を連携させる必要がある。バイオシミュレーションの会では、このような認識の下に、1) ゲノム情報解析、2) 蛋白質の *ab initio* フォールディングシミュレーション、3) 蛋白質の水溶液中での熱運動シミュレーション、の3分野でのシミュレーションを、地球シミュレータで行っている。平成17年度、上記4サブテーマについて研究を進めてきたが、それらは、サブテーマ4が1)に、サブテーマ1が2)に、サブテーマ2, 3が3)、に分類される。それらの研究内容の概略を、次に示す。

### <サブテーマ1>

タンパク質の自然の立体構造は、アミノ酸配列の情報及び周りの溶媒環境のみで決まっており、自由エネルギーの最小状態に対応すると広く信じられている。しかしながら、蛋白質の3次元構造を、線形に引き伸ばした状態から、シミュレーションだけから予測することは、エネルギー最適化問題の中

でも最も難しい課題のひとつである。独自に開発したレプリカ交換分子動力学(REMD)により、アミノ酸数56個のタンパク質であるProtein Gにおいて、ab initioフォールディングシミュレーションに始めて成功している。今後は、よりサンプリング効率の高いMUCAREMを適用していく予定である。

#### ＜サブテーマ2＞

本研究課題の目的は、生命を維持する上で重要な蛋白質であるヘモグロビンの立体構造変化(アロステリック効果)を、分子動力学シミュレーションプログラムCOSMOS90を用いて再現することである。COSMOS90は、水中の蛋白質における長距離クーロン力をカットオフせずに高速にシミュレートするプログラムであり、その理論も独自に開発されている。ヘモグロビンの分子動力学シミュレーションを行うため、COSMOS90を地球シミュレータ上で高速に稼動できるように、ベクトル化と並列化を行った。その結果、分子動力学シミュレーションプログラムとしては、世界的に見ても最も高速性能を実現しているプログラムのひとつとなっている。

#### ＜サブテーマ3＞

数百万原子からなる生体分子を扱うことのできる並列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は長距離クーロン相互作用を精度良く高速に計算するParticle-Particle Particle-Mesh(PPPM)法など、最新のアルゴリズムを搭載している。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結果、約55万原子の系(RuvAB-Holliday分岐DNA)について、45ノード(360プロセッサ)を用いてもベクトル化率95%以上、並列化効率50%以上の性能を達成することに成功した。

#### ＜サブテーマ4＞

多様な地球環境由来の難培養性微生物のゲノムは、新規性の高い遺伝子を含む可能性が高く産業的に関心を集めており、ヒトの体内環境の難培養性微生物については、医薬学

的にも注目を集めている。ゲノム配列解析用の一括学習型SOMは、新規性の高い配列の系統分類を可能にする、革新的なバイオインフォマティクスの技術である。SOMは生物種に関する予備知識なしに、断片配列の大半を生物系統に分類(自己組織化)が可能であり、環境由来のゲノム配列の新規な情報の解析に適用している。

バイオシミュレーション研究者の会から申請されている研究課題は、その分野のリーディングエッジに位置する研究水準を維持していると考えている。

## 6. 今後の計画・展望

将来のHPC(High Performance Computing)の分野における重要な領域として、マテリアルシミュレーションが注目されており、それらは、大まかに言って、ナノとバイオに分類される。何れも、次世代の産業を支える基本物質の開発に繋がるとの期待から、欧米ではプログラム開発やシミュレーションが活発に行われている。期待されているように、マテリアルシミュレーションが産業競争力の強化に有効であるならば、それらのプログラムの国産化を進めなければならない。また、使い勝手の良いシミュレーション環境も用意する必要もある。

そのためには、上記サブテーマで開発されている様々な計算機能が自由に連成され、複雑な事象についてのシミュレーションが簡便に行えるようにする必要がある。その実現には、これらのプログラムをコンポーネント(部品)化して、更に、それらのコンポーネント間で授受されるデータ書式を標準化することが、有効であると考えている。地球シミュレータを拠点として、生体分子シミュレーション関連のプログラムのコンポーネント化とデータの標準化が進むように、この分野の研究者間の交流を活性化していきたいと考えている。