Accelerating the Research in Drug Delivery System; A Challenge of the Earth Simulator to Medical Innovation

Project Representative
Atsushi Miyauchi Research Organization for Information Science and Technology

Author
Atsushi Miyauchi Research Organization for Information Science and Technology

Objectivity of our project is to accelerate the research of drug delivery system through computational scientific method. We have inspected the problem domain and figured out the required resources and the research strategy. Along with this strategy, we have developed the quantum chemistry program based on tight binding approximation. The program is vectorized to 96.3% overall and parallelized using MPI communication library. We have performed preliminary calculations and obtained fairly good performance in terms of vector facilities. However, parallel performance is revealed to be insufficient for our final goal. Further efforts to get higher parallel performance are now underway. We expect we will be able to perform larger scale calculations next year.

Keywords: drug delivery system, DNA, polymer, molecular orbital method, tight-binding approximation

1. Introduction

Since the end of the last decade nanotechnology has been grown at rapid pace and spread over many different fields out of its birthplace, material science. Now in medical science, nanotechnology is expected to open up the door to the innovative methods of treatments no one can imagine twenty years ago. One of its novel application is drug delivery system (DDS). The main purpose of DDS is to target the seat of a disease and carry drugs there precisely.

Our project aims to accelerate the research of DDS by fully exploiting the amazing computational ability of the Earth Simulator. The project is planned to continue at least three years. FY05, the first year of the project, we inspected the problem domain and figured out the required resources and research strategy.

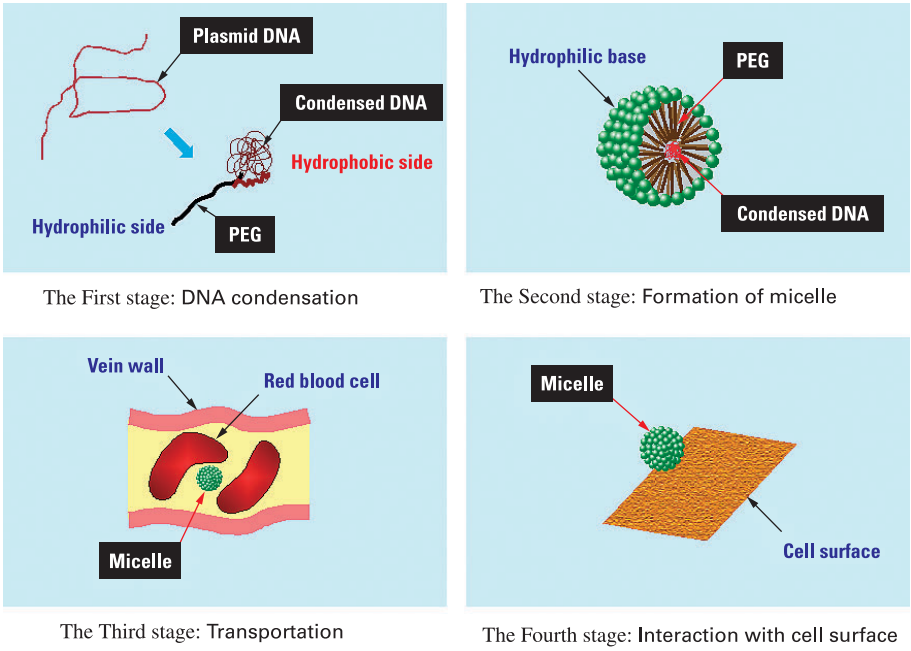
This report is organized as followings. The next section describes the physical aspects of DDS in detail and gives the computational requirements corresponding to each aspect. In the third section, tight binding approximation is described. We used it as our main workhorse to calculate the phenomena in the smallest physical scale. An example of tight binding calculation is shown in the fourth section. Final section summarizes this report.

2. Drug delivery system

Many kind of DDS technology has been proposed so far. Amongst them, our project devote much attention to the method using nanosized particle called micelle mainly comprised of poly-ethylene-glycol (PEG). This method is recently developed by Professor Kataoka of the University of Tokyo and expected to be promising in near future because of its relatively low impact on human body [1]. Although the whole process of PEG-motivated DDS is much complicated, we can recognize four characteristic stages in that. See figure 1.

In the first stage, relaxed DNA attached to a PEG starts to condense in the solution. Shortly after, DNA gets tangled to a small ball. Notice that PEG-DNA complex has hydrophilic end of PEG and hydrophobic end of DNA. In other word, it has amphiphathic property. In the next stage, hundreds of those PEG-DNA complexes in the water meet together and spontaneously form sphere called micelle in which each DNA heads for the center and PEG heads for the surface. The driving force of this self-organizing formation is amphipathic property mentioned above. In the third stage, micelles are carried through vein and capillary tube slipping through the blood cells. In the last stage, micelles reached the targeted portion of disease are attracted by the local gradient of ion concentration in the vicinity of the cell surface and then absorbed into it through the carrier or channel protein located at the membrane. These types of problem can be classified as multi-scale and multi-principle phenomena in contrast to single-scale and single-principle phenomena conventional science has dealt with.

To simulate such kind of complicated phenomena efficiently we must choose suitable computational methods corresponding to each stage. In the first stage, the electronic structure inside DNA is thought to play an important role for condensation. Therefore quantum mechanical consideration



The First stage: DNA condensation

The Second stage: Formation of micelle

The Third stage: Transportation

The Fourth stage: Interaction with cell surface

Fig. 1 Characteristic stages of drug delivery system.

is significant to calculate such phenomena. Density functional theory (DFT) is usually preferred for those cases. Although DFT can perform highly accurate calculation and give reliable results, its computational workload is immensely large, sometimes unacceptable. Therefore, DFT has been applied for relatively small systems up to few hundred of atoms even on the current top-rated supercomputers. However as least few thousands of atom are indispensable to simulate condensation of DNA in solution. We anticipate that very high accuracy like DFT offers is unnecessary to figure out the behavior or mechanism of DNA condensation. On the other hand, we suspect that molecular dynamics method based on heuristic potential is insufficient since it is unable to take complex configuration of local electrons into account. Consequently we adopt tight binding approximation as a baseline algorithm. By virtue of its thinner workload, tight binding approximation can be expected to handle more than ten thousands of atoms while it take local configuration into account adequately. In FY05 we developed a new quantum molecular dynamics program from scratch which is based on tight binding approximation and is able to calculate any type of atoms, since the former program we have used is unable to treat atoms other than carbon.

In second and fourth stages, number of atoms participated in the dynamic evolution amounts to a million or more. Therefore any quantum mechanical treatments are impossible even with tera-flops supercomputers. Our strategy is restricted to classical molecular dynamics method whether we like it or not. However, our classical molecular dynamics program is still amenable to be elaborated to this purpose. So, we went for tuning that program for further performance in FY05.

The third stage is thought to be genuinely classical phenomena. Fluid and solid mechanics govern the physical properties and behavior of the system at this stage. Numerical methods suitable for it are finite difference, finite element and discrete element method. We do not have adequate program of those kinds. So, we are still searching for the best method for the time being.

As described above, our project covers large part of entire DDS process. But actually some parts make progress faster, and others slower. In the next section, we focus on one of the smoothly progressing efforts.

3. Tight binding approximation

Quantum many body problem is one of the most difficult physical problem and many approximation method for it were proposed so far. Amongst them, three major methods are molecular orbital (MO) method, Hartree-Fock method and density functional theory (DFT). First we describe MO method since our tight binding approximation is strongly linked to it.

Imagine a hydrogen atom, which has single electron going around single ion. In this case, known as two-body problem, the equation of motion of electron can be solved explicitly. From its results, we can see there exists infinite orbitals and the electron is allowed to occupy one of those orbitals. Adding electrons to this system makes the problem extremely difficult. As is well known, even three body problem is analytically intractable as same as classical dynamics. However, if inter-electron interaction is sufficiently weak, we can expect the orbital picture still works for a multi-electron atom as a good building block of approximation. It is called the atomic orbital assumption.

Electrons are considered as independent each other, and

orbitals can be calculated from single electron hamiltonian. So far, single ion is in our mind, and now moves to two ions. If those ions locates close each other, they can be looked upon as doubly charged single ion: a fused ion. In this case, there is no doubt to apply atomic orbital assumption for it. Extrapolating this argument, we can assume the existence of molecular orbitals even if ions are separated. The molecular orbitals are calculated from single electron hamiltonian of molecule and expected to be distributed over entire molecule, not to be localized at specific ion. Practically, molecular orbitals are approximated as linear combination of atomic orbitals (LCAO) and their optimal coefficients are determined by variational principle.

In what follows, computational aspects of our program are described in brief. A main constituent in calculation of variational principle is hamiltonian matrix. Generally hamiltonian matrix has dense elements and turns out to require plenty of arithmetic operations, while tight binding approximation restricts this matrix elements to the nearest atoms around each atom [2]. By virtue of this decimation, operational count of tight binding approximation becomes significantly smaller than fully banded hamiltonian. In addition, it needs to handle only valence orbitals and electrons. On the other hand, Hartree-Fock method inevitably requires full orbitals and electrons due to totally antisymmetric property of Slater determinants. Furthermore, tight binding method calculate each matrix element from a few parameters sometimes called Slator-Koster parameters instead of time-consuming exact integration [3]. Using sparse hamiltonian matrix above, energy of electrons can be obtained from the density of states (DOS) of electrons, which is determined by applying spectrum theorem for Green function.

Electron Energy is defined as

$$E_{elec} = \int^{E_f} n_0(E)\, E dE,$$

where E_f denotes Fermi energy. Spectrum theorem assures a following relation

$$n_0(E) = -\frac{1}{\pi} \lim_{\varepsilon \to +0} \mathrm{Im} G_{00}(E + i\varepsilon).$$

Green function is defined as a kind of inverted hamiltonian matrix. Problem here is to invert matrix directly requires extremely large workload and usually exhausts majority of computational time.

Instead of direct inversion, our program adopt Lanczos iteration which can construct Green function very efficiently [4] through continued fraction,

$$G_{00}(E) = \left\langle f_0 \left| \frac{1}{E-H} \right| f_0 \right\rangle = \cfrac{1}{E - a_0 - \cfrac{b_1^2}{E - a_1 - \cfrac{b_2^2}{E - a_2 - \cdots}}},$$

where f_0 denotes a seed state. The coefficients a_n and b_n are obtained through sequential tridiagonalization of hamiltonian matrix starting from the seed state,

$$a_n = \langle f_n | f_n \rangle,$$
$$b_n = \langle F_n | F_n \rangle,$$
$$|F_n\rangle = H|f_n\rangle - a_n|f_n\rangle - b_n|f_{n-1}\rangle,$$
$$|f_n\rangle = b_n^{-1/2}|F_n\rangle.$$

So far, each ion is assumed to be fixed under Born-Oppenheimer approximation. However, our physical interest concerns with dynamic behavior of molecules. Thus forces exert on each ion are needed for that purpose. Those forces are obtained by applying Hellmann-Feynman theorem [5].

$$F = -\left\langle \psi \left| \frac{\partial E}{\partial x} \right| \psi \right\rangle$$

Finally, fourth order Runge-Kutta method is employed to integrate the equation of motion of ions numerically. Here closes single computational cycle.

The major restriction to tight binding approximation is in lack of inter-electron interaction in principle, in other word, no exchange and correlation terms in contrast to Hartree-Fock or DFT. Certainly we are aware that highly accurate energy calculation is beyond our scope, we concentrate exclusively on the change of configuration or long term behavior of large system of molecules quantum mechanically intractable ever. We strongly believe those kind of simulation will come to play a crucial role in future bio-polymer science.

4. Computational example

A preliminary calculation has performed to validate our program. A small segment of DNA comprised of 888 atoms, 28 residues was chosen for this purpose. Its coordinate data were taken from predefined PDB data set in Nucleic Acid Builder (NAB) developed by the Scripps Research Institute [6]. Data set was slightly modified from the standard PDB format to acceptable format of us. Our program is written in C++ programming language and compiled with Earth Simulator's C++ compiler wrapping MPI. We used -Kexceptions option, which was necessary to get exception handling mechanism available. Any other option about performance was not used. To promote vectorization, source codes were substantially rewritten. Some functions were inlined manually, some loops were merged to extend vector-length and so on. Eventually our program achieved 96.3% of overall vectorization ratio after those aggressive vectorization efforts.

Figure 2 shows the snapshot after 200 computational steps. White, gray, blue, red and green balls designate hydrogen, carbon, nitrogen, oxygen and phosphorus atoms respectively. It revealed that the average time consumption for

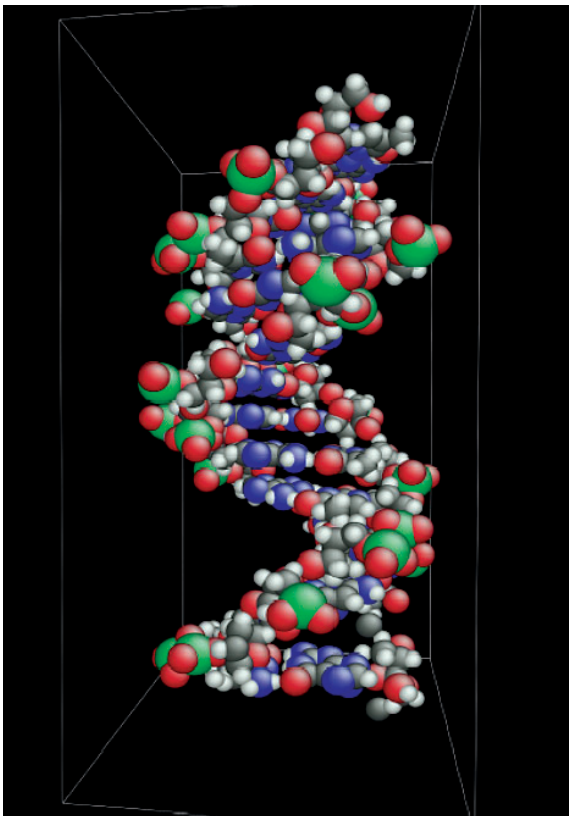


Fig. 2 A snapshot of DNA after 200 computational steps, 888 atoms included.

each computational step per atom was 0.013sec. From these results, we can predict promising large-scale simulation in near future. For example, utilizing 1000 processor elements, approximately 360 hours run makes us to calculate the duration of 100psec with 100 thousands of atoms. In addition to vectorization, parallelization has done simultaneously. By virtue of order N property of the program, minimal works were sufficient to make it parallel. Strategy of parallelization we adopt was rather simple: atomic decomposition through MPI functions. However its parallel performance was still insufficient for our final goal. According to the criterion the Earth Simulator Center proposed, our estimation revealed the number of processor elements to be allowed to assign would be at most around few hundreds currently. Now we still continue to try to get higher efficiency at least a thousand of processor elements to be allowed.

5. Summary

To make DDS to be realized as early as possible, we try to accelerate it by using computational scientific method. As the first year of our project, we have developed quantum chemistry program based on tight binding approximation. Fairly good performance in terms of vector facilities was obtained. However, parallel performance is left under satisfactory level for our final goal. We still continue to try for higher performance and hope to accomplish a large-scale simulation in the next fiscal year.

References

[1] K. Kataoka, A. Harada, D. Wakebayashi and Y. Nagasaki, "Polyion Complex Micelles with Reactive Aldehyde Groups on Their Surface from Plasmid DNA and End-functionalized Charged Block Copolymers", Macromolecules, vol.32, pp.6892, (1999).

[2] G. Grosso and G. P. Parravicini, *Solid State Physics*, Academic Press, 2000.

[3] J. C. Slator and G. F. Koster, "Simplified LCAO Method for the Periodic Problem", Phys. Rev., vol.94, pp.1498, (1954).

[4] R. Haydock, "The Recursive Solution of the Schrodinger Equation", Solid State Physics, vol.35, pp.215, (1980).

[6] P. Atkins and R. Friedman, *Molecular Quantum Mechanics*, Oxford University Press, 2005.

[7] http://www.scripps.edu/mb/case/

地球シミュレータを利用したドラッグデリバリシステムの研究：革新的医療への挑戦

プロジェクト責任者
宮内　敦　財団法人高度情報科学技術研究機構
著者
宮内　敦　財団法人高度情報科学技術研究機構

我々の目的は計算科学的手法を通してドラッグデリバリシステムの研究を加速することにある。プロジェクト初年度となる本年度は、問題領域を調査し必要となる研究項目を洗い出し、それに対する戦略を策定した。さらに、この戦略に沿って強結合近似に基づく量子化学計算プログラムを開発した。このプログラムは地球シミュレータ上で96.3%までベクトル化されている。同時に、MPIライブラリを用いて並列化も施した。プログラム動作を確認するために小規模のテスト計算を行い、十分なベクトル性能を有することを確認した。しかしながら、並列性能は利用可能ノード数が数十ノード程度となり、我々の最終目標を下回っていることが判明した。現在、より高い並列性能を得るべくチューニング作業を継続している。次年度にはより大規模な計算を可能にしたい。

キーワード：ドラッグデリバリシステム, DNA, 高分子, 分子軌道法, 強結合近似