Project on Bio-IT for Next-generation Healthcare

● Tatsuhiro Yamashita ● Nozomu Kamiya ● Atsushi Tomonaga
● Shunji Matsumoto

As part of its computing and simulation technologies in science, Fujitsu has been engaging in R&D of IT-Soyaku (*in silico* drug design) that makes it possible to generate chemical compound information and search for potentially efficacious materials. A joint research study started in FY2011, conducted with the Research Center for Advanced Science and Technology of the University of Tokyo and some pharmaceutical manufacturers. It successfully composed a new compound with high activity that held promise in terms of having possible medical potency. This development was made possible by the improvement of computing capacities such as the K computer, together with enhanced computer simulation technology. New initiatives are already underway to realize the next-generation healthcare system. One such initiative is a project to develop a single-cell diagnostic system. It takes a minute sample of a single cell and manages to measure and analyze component materials, and it has potential to be applied to health check-ups and other areas. Another case is the development of analytical technology for genome data, which is expected to occupy a significant place in future personalized medicine. These projects aim to develop technologies by joint studies with academic and research institutes. In this paper, we introduce some initiatives at Fujitsu in relation to the development of bio-IT, and describe the potential to change future healthcare.

1. Introduction

The year 2014 brought a series of reports about incidents of infectious diseases not so familiar in Japan, such as Ebola fever and dengue fever. It is considered that global warming and other environmental changes as well as economic globalization are behind these emergencies. The rapidly aging population in society also adds to the changes in our lifestyles and the environment. These phenomena are giving rise to situations where medical care in the field needs to take different measures than conventional treatments. In the case of Ebola fever, a potentially potent drug was found from a certain influenza drug that was under development. However, there is no guarantee that efficacious drugs may always be found among known pharmacopoeia. Meanwhile, only one in thirty thousand new drug candidate compounds is likely to be successful,[1] and new drug development becomes more difficult because of the higher requirements that the drugs must exceed the efficacy of existing drugs for approval.

Fujitsu has been working to develop technology that can handle information pertaining to proteins, genes and pharmaceutical materials that control biological objects under the concept of "bio-IT." Aiming to make a breakthrough in drug discovery, we have developed technology for *in silico* drug design (hereafter "IT-Soyaku"),[2] in which candidate chemical compounds are virtually designed on computers and evaluated for their potential efficacy through computations. For the execution of IT-Soyaku, large-scale computing must be utilized with supercomputers. Recent years have seen improvements in computing capacities such as the K computer bringing about the foundation for practical applications of this technology based on large-scale computing. For example, we successfully composed a new compound with the efficacy we envisaged for targeting cancer in a joint research study conducted with the Research Center for Advanced Science and Technology (RCAST) of the University of Tokyo and some pharmaceutical manufacturers.

Meanwhile, there are high expectations among

medical care providers for personalized medicine that offers dedicated medical care which is suitable for individual innate characteristics and conditions of patients that change each day. Technologies behind personalized medicine include DNA sequencers, mass spectroscopes and a variety of sensors developed for wearable devices. The task at hand is to combine and leverage various types of information obtained through these appliances, for the realization of personalized medicine. Against this background, we embarked on the development of technology for information analysis and utilization in this field.

In this paper, we outline the concept of IT-Soyaku at Fujitsu, and our projects for it. We will also explain the development of data analysis technology that aims to bring about a next-generation healthcare system.

2. IT-Soyaku at Fujitsu

The concept of IT-Soyaku has gained wide recognition in recent years, referring to the idea of computer-based development of drugs. At Fujitsu, we employ a combination of two solutions in this area, namely, the Optimum Packing of Molecular Fragments (OPMF), which facilitates the virtual designing of pharmaceutical candidate compounds, and the Massively Parallel Computation of Absolute Binding Free Energy with well-Equilibrated System (MAPLECAFEE),[3] used to predict the activity of the designed compounds to estimate the levels of their potency (**Figure 1**).

2.1 Designing pharmaceutical candidate compounds with OPMF

There are tens of thousands of proteins in the human body, each performing vital functions to sustain the organism. The functions are executed as these proteins are bonded with other proteins or the like in specific sites. The basic principle of drug designing is to lock out such functions of proteins related to disease causation by binding the proteins with other materials. OPMF facilitates an automated designing of chemical compounds according to a three-dimensional configuration of the protein's relevant sites.

As shown in **Figure 2**, OPMF first places a unique partial structure of a compound, called a fragment, into a stable site of the target protein, taking into consideration the interatomic force in relation to surrounding atoms. Benzene, with its hexagonal shape (six-membered ring) formed by having a carbon atom at each vertex, is an example of the fragments. The fragments thus positioned are joined together to form a compound. However, six-membered rings have great variations, with some of the component carbon atoms of benzene replaced with nitrogen or other elements. If all the variants were taken into consideration, the compound design would require a considerable computational effort. Therefore, OPMF employs abstract

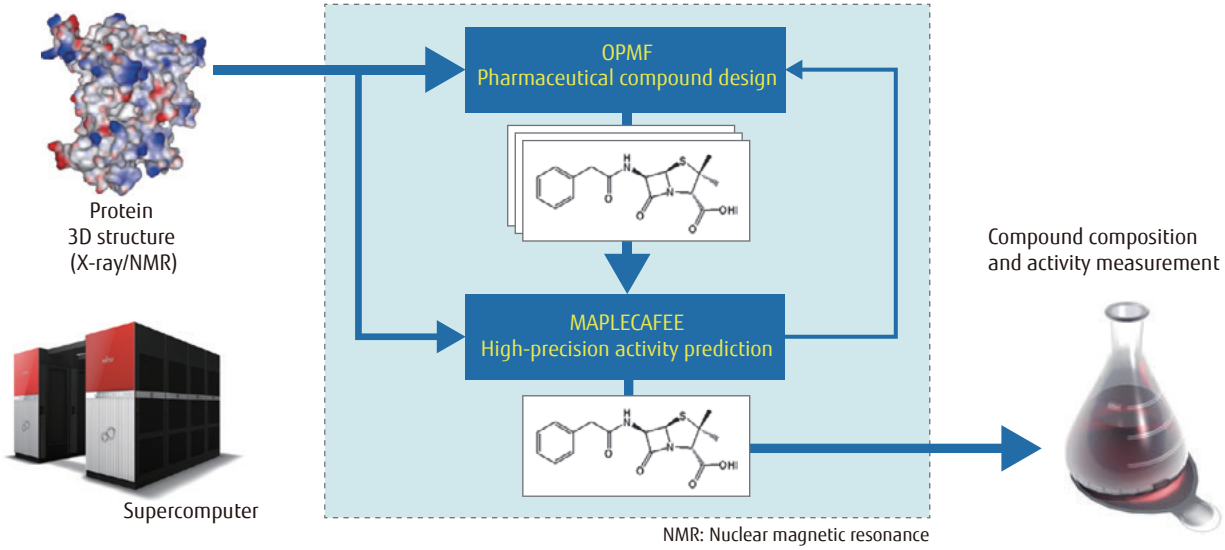

Figure 1
IT-Soyaku at Fujitsu.

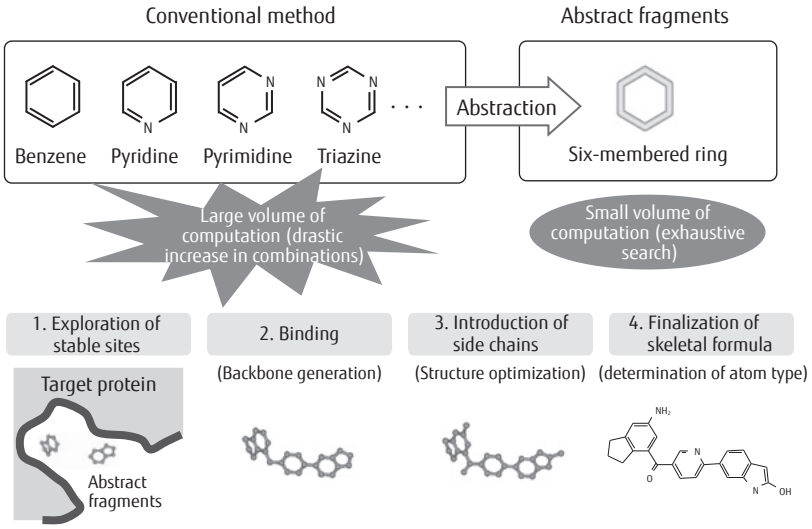

Figure 2
Compound designing process using OPMF.

fragments, without specifying component atoms, thereby reducing the computational efforts required and enabling exhaustive exploration of a diverse range of compound structures.

2.2 Prediction of activity with MAPLECAFEE

An important indicator of efficacy is whether the designed compound stays bound with the target protein and prevents the protein's activities. The binding free energy, which is the energy required for separating the compound combined with the protein, can be calculated by experiment. Instead of the experiment, MAPLECAFEE calculates the binding free energy based on the Molecular Dynamics (MD) Simulation, and predicts the activity between the compound and protein.

To determine the binding free energy accurately, it is necessary to take into consideration not only the protein and compound, but also the behavior of water molecules existing in the immediate vicinities to reconstruct the actual conditions within the organism. The simulation involves a considerable volume of computation, for it is necessary to cover all interatomic forces in operation. We addressed this issue by arranging a parallel computation in MAPLECAFEE, to achieve the calculation to be performed in a few days instead of a few years with a single computer (**Figure 3**). This computational method is suited to a supercomputer such as the K computer. The accuracy of activity prediction is further enhanced by adding a set of parameters

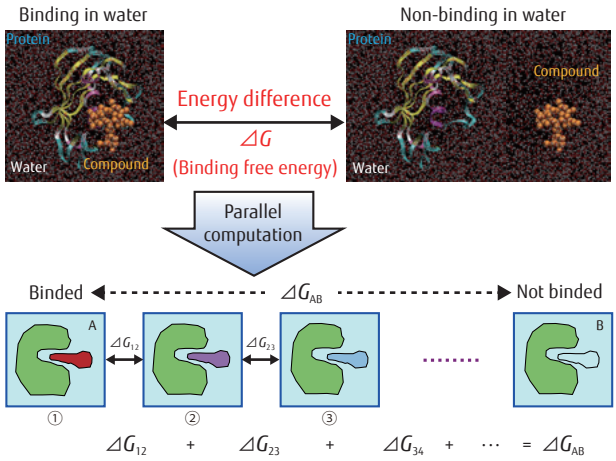

Figure 3
Computation of binding free energy using MAPLECAFEE.

that were uniquely developed by Fujitsu Laboratories, as force field parameters to define the interatomic forces used in the MD simulation. To generate the parameters, we employed the Force Field Formulator for Organic Molecules (FF-FOM), a precision force field assignment program that was also developed by Fujitsu Laboratories.

Figure 4 depicts the results from a blind test conducted in 2009, for the activity prediction using MAPLECAFEE. In the test, we used five existing compounds (with the activity for proteins verified) provided by a pharmaceutical company, and compared the data between values obtained from experiments and

predictions by MAPLECAFEE. The horizontal axis represents the values from experiments, and the vertical axis stands for the predicted values. Accurate prediction would result in the computational results being plotted between the two dotted lines. Initially, one of the five compounds yielded a predicted value that was outside the experimental value range. However, it was suspected that the experiment had not been accurately conducted for this particular compound. Therefore, a re-experiment was performed following the results of this blind test. The result was closer to the predicted values. Consequently, the pharmaceutical company was highly impressed by our MAPLECAFEE.

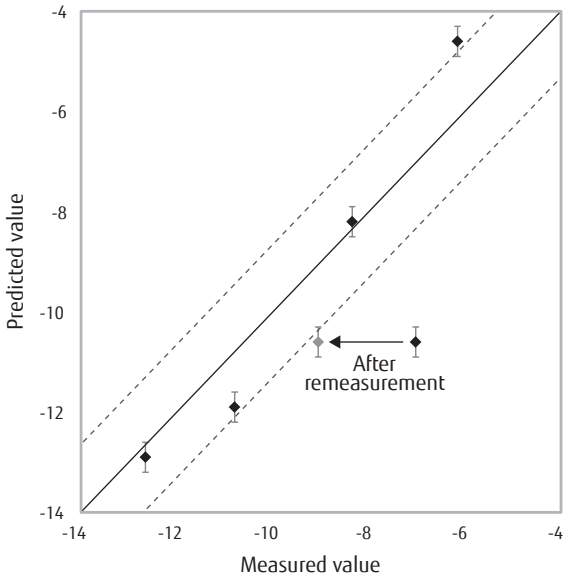

Figure 4
Blind test results.

3. IT-Soyaku joint research

Aiming to establish the validity of OPMF and MAPLECAFEE for IT-Soyaku, around 2010 Fujitsu started selecting target proteins and trials to synthesize and assay the chemical compounds designed using these systems. Building upon this development, Fujitsu and RCAST embarked upon joint research projects on IT-Soyaku with each of three different pharmaceutical companies (**Figure 5**). The joint research projects are operated on the following themes:

1) Construction of an IT-Soyaku platform that integrates a large-scale computing environment and software programs required for conducting the IT-Soyaku R&D
2) Creation of new candidate pharmaceutical compounds (lead compounds) using the platform

In the following, we describe major activities in these joint research projects.

3.1 Construction and utilization of large-scale computing environment

The MD simulation takes into account a diverse range of forces in operation between the target molecules and atoms, and estimates molecular/atomic position changes through calculation. The result is attained by following minute changes occurring in one femtosecond (one millionth of one billionth of

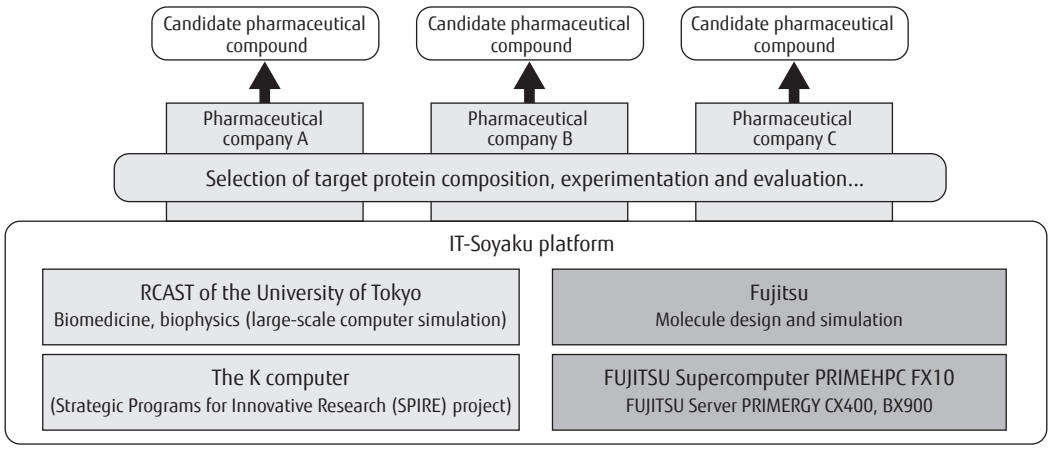

Figure 5
Joint research formation.

a second). Observations of proteins from a macro perspective require the passage of one microsecond (one millionth of a second) or one millisecond (one thousandth of a second), which makes it necessary to repeat the computation one billion or trillion times. For more efficient drug discovery, computations must be executed on several compounds simultaneously. Therefore, establishing a large-scale computing environment is essential for pursuing the research and development of IT-Soyaku.

Being selected for the K computer projects for the Strategic Program for Innovative Research hosted by the High Performance Computing Infrastructure, RCAST is pursuing activity predictions through large-scale simulation and evaluation. Fujitsu has reinforced its in-house computational environment for IT-Soyaku, securing a supercomputer capable of performing 100 TFLOPS (floating-point operations per second), one-hundredth the capacity of the K computer. Consequently, it is now possible to run computations for more than 10 compounds simultaneously.

3.2 Application to research themes targeting cancer

Each pharmaceutical company offered a theme for the joint research, and they all had cancer as a target disorder. The activity prediction of MAPLECAFEE is based on the difference in energies between two states of the protein and compound, where they are bound and separated. It requires an accurate conformation for the binding of these two components, representing the state of efficacy in the body, in order to predict the activity with high accuracy.

For existing compounds, there are ways to verify the combining state, such as the X-ray crystallography, or estimations based on published data. However, IT-Soyaku deals with virtually designed compounds, for which these methods are not available. In fact, some compounds handled in our joint research projects were found to be of no activity in experiments, despite the predicted high activity. Nevertheless, these results also contributed to building up know-how on designing new chemical compounds and improving the system, owing to the improvement in large-scale computing, which also made it possible to explore the reasons behind these "failures."

3.3 Discovery of new compounds

Through a joint research program with Kowa Company, Ltd. (hereafter "Kowa"), 100 candidate compounds that Fujitsu designed in pursuit of IT-Soyaku were narrowed down on the basis of certain criteria such as results feasibility, and eight were chosen to be actually synthesized by Kowa. One of them did achieve the target level of activity In conjunction with the efforts for IT-Soyaku, we conducted an orthodox exploration to find high activity compounds among known varieties. The result showed that only 0.4% of the investigated compounds achieved the target activity. This serves as strong evidence for the astonishing success rate of IT-Soyaku (one in eight). The other projects, likewise, succeeded in discovering compounds that met the target level of activity among a few, designed by IT-Soyaku.

While these compounds found through IT-Soyaku are exciting because of their new discovery status and high activity, there is no means to evaluate their pharmacological appeal. We are of the opinion that drug discovery should combine the IT-based approach and more conventional, experiment-based methods, leveraging the advantages of both. MD simulation makes it easier to visualize the behaviors of proteins and compounds through the process of IT-Soyaku, providing for an understanding of the variety and severity of forces operating on atoms in certain positions. Information like this gives researchers insight into ways to enhance compounds.

We will continue to work on improving the newly discovered compounds with promising results to enhance the activity, together with improving the system. There is still a long way to go before IT-Soyaku reaches the stage of being able to create a new drug. We will strive to develop further new themes, and garner more partners for collaboration in IT-Soyaku in the future.

4. New initiatives for the next-generation healthcare system

There are particularly high expectations for the realization of personalized medicine as the next-generation healthcare system. This is partly explained by the recent development of the next-generation high-speed DNA sequencers and other devices that altogether have made individuals' genomic information more accessible and affordable. The future challenge is to establish

ways to process and leverage the diverse range of data, including genomic information, that are generated by these devices. In other words, the challenge is to develop technology for medical data analysis. With this in mind, we have started the following two new projects from the perspective of "bio-IT" in the field of life information science:

1) Development of technology to analyze material information using mass spectroscope

A mass spectroscope is an instrument to measure the molecular weight (mass) of materials, for example, contained in a blood sample, to analyze the composition and proportion of component materials. Single Cell Mass Spectrometry, developed by RIKEN, is a method that makes it possible to analyze component materials within one single cell, by extracting them under a microscope. It is capable of carrying out analyses from a small sample of blood, and can be completed in a few minutes. By virtue of this, there are a variety of possible ways to apply this method. We developed an analytical technology to process the data output of this device, under the concept of a Single Cell Molecular Evaluation System through joint research efforts.

2) Development of technology to analyze genomic information

Pieces of genomic information, ones relevant to diseases such as cancer in particular, have been accumulated to build up databases around the world. Furthermore, there are epigenomes which contain information about changes that occur to genomes a posteriori, and some consider that they hold certain significance with regards to cancer and other diseases. These are important data for personalized medicine. However, the amount of genomic data, including epigenomes, is extraordinarily large. It is not easy to identify how certain segments of the data relate to other parts, or how they influence each other. In this sense, some old technologies such as data mining and machine learning are having their potential rediscovered. Fujitsu is working to develop technology to analyze genomic information using its proprietary technologies in these fields, through joint research efforts with specialists on various disorders.

5. Conclusion

This paper described some projects in IT-Soyaku and the development of medical data analysis technology at Fujitsu. Bio-IT deals with elements such as proteins, DNA and pharmaceutical agents, which are invisible to the naked eye but influential on organisms. We aspire to visualize and quantify them in order to establish an information base for the next-generation healthcare system. It is our hope that the information thus derived will contribute to the future of a healthy society.

The Single Cell Molecular Evaluation System that is referred to in this paper was developed through the 2013 Adaptable and Seamless Technology Transfer Program (five-year program) provided by the Japan Science and Technology Agency (JST).

References

1) Japan Pharmaceutical Manufacturers Association: Data book 2012.
2) S. Matsumoto: Innovation in Drug Discovery through Information Technology. *FUJITSU Sci. Tech. J.*, Vol. 44, No. 4, pp. 481–488 (2008). *http://www.fujitsu.com/global/documents/about/resources/publications/fstj/archives/vol44-4/paper17.pdf*
3) H. Fujitani et al.: Direct calculation of the binding free energies of FKBP ligands. J. Chem. Phys., 123, 084108 (2005).

Tatsuhiro Yamashita
Fujitsu Ltd.
Mr. Yamashita currently engages in development of bio-IT-related technology.

Atsushi Tomonaga
Fujitsu Ltd.
Mr. Tomonaga currently engages in R&D of technology for IT-Soyaku.

Nozomu Kamiya
Fujitsu Ltd.
Mr. Kamiya currently engages in R&D of technology for IT-Soyaku.

Shunji Matsumoto
Fujitsu Ltd.
Mr. Matsumoto currently engages in development of bio-IT-related technology.