Innovation in Drug Discovery through Information Technology

● Shunji Matsumoto

(Manuscript received March 5, 2008)

Traditional drug discovery begins by identifying a biological substance (target protein) related to a disease and finding small chemical compounds that can control that function in a library of known chemical compounds. This limits the search space from the start, making it difficult to create chemical compounds with new structures (new drugs). In recent years, improved performance in computer processing has made it possible to research and develop drugs using computers and information technology (IT) in a process called "IT-based drug design" that goes beyond existing chemical compounds and expands the search space dramatically. It uses the three-dimensional structures of proteins and chemical compounds and molecular simulations to enable the design of drug candidates that include unknown virtual substances. This paper provides an overview of IT-based drug design that makes breakthroughs in drug design possible through exhaustive, high-speed search techniques and molecular simulations.

1. Introduction

Despite the desire to "continue with a healthy life forever", modern human beings cannot easily avoid a life that in some way involves cancer, diabetes, or other diseases. Recent progress in medical technologies has been remarkable, but the mystery of life, which cannot be completely unraveled due to the diversity and unpredictability of living organisms, can sometimes appear to be mocking advanced science.

At the same time, the computer, which was used for calculating ballistic trajectories for missiles in the early days, has come to be applied to diverse purposes. Today, in addition to making the work of business more efficient, it enables people throughout the world to exchange information over the Internet and enjoy realistic and dynamic simulations of events and high speeds with game machines that can fit into the palm of a hand. Behind all this lie scientific theories and advanced technologies for high-speed processing of large volumes of information.

Attempts at applying the simulation abilities of computers to the world of living organisms have begun.[1] Using actual data based on biochemical experiments and computer simulations based on scientific theories, ambitious national projects are now underway that aim to reproduce the human body on computers, explain the mechanisms behind diseases, predict the functions and side effects of drugs, develop new methods of treatment, and so on.

As part of this movement, Fujitsu has been researching and developing drugs using computers and information technology (IT) in a process called "IT-based drug design". This is the world of high-speed simulation that pursues high precision, high speed, and reality at the atomic level.

In this paper, we first describe the current state of drug discovery and surrounding issues. We then describe the core technologies of IT-based

drug design and their actual practice and touch upon future directions.

2. Current status of drug discovery

The patents of many drugs in the market on sale today are scheduled to expire around 2010. These will give way to generic drugs that have recently become ubiquitous on television commercials and that will pose a threat to the source of profits of the pharmaceutical industry that developed the original business for drugs. Yet, it is also said that the development of new drugs as follow-on products is not proceeding as expected, thereby raising questions about the "2010 problem".

Traditional drug discovery begins by identifying a biological substance (target protein) having a deep relationship with a certain disease and then finding a chemical compound that can control that function, generally by looking in a library of known chemical compounds. A synthetic chemist then works with this chemical compound to form a new substance. This process makes the search space highly dependent on the chemical-compound library and the experience of the synthetic chemist, which are thought to be the main reasons why it is becoming increasingly difficult to create new substances (new drugs).

The traditional flow of drug discovery is shown in **Figure 1**. It generally takes ten years and several hundred million dollars before a single drug is developed from a chemical-compound library of several million entries. Furthermore, the preparation of a new substance is no guarantee of success because most drug candidates eventually fail for reasons such as having no medicinal effect or major toxicity or side effects. Although it is easy to think that "If only the medicinal effects, toxicity, and side effects could be accurately predicted at an early stage, then …". However, the diversity and unpredictability of living organisms does not allow this wish. Huge profits from successful drugs are based on these risks.

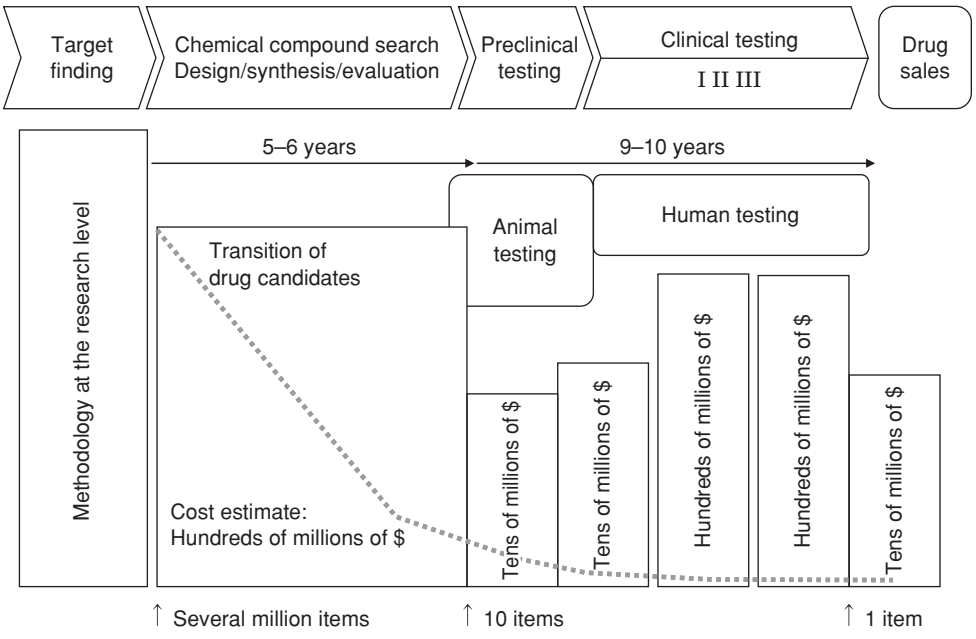

Figure 1
Traditional flow of drug discovery.

3. Core technologies of IT-based drug design

3.1 Target finding

The first important step in drug discovery is to see what substances react in what way where and with what in the body, what kind of phenomena arise as a result, and how this is all related to disease. While this is simple to write, reactions in living organisms are not something that can be systematically understood at a glance. This step actually requires the efforts of scientists from around the world in interpreting hypotheses and the results of animal testing that have been accumulated over many years, in forming new hypotheses, and in retesting those hypotheses through experiments. This can be an enormous amount of work in terms of both time and effort, and people feel the need to establish technology that can support scientists in this endeavor through the use of IT.

In this regard, we propose using text mining technology[2] to automatically extract relationships between substance-interaction information and diseases from medical-literature databases managed by the National Center of Biotechnology Information (NCBI) in the U.S. and have begun sales of dynamic pathway analysis software called Genesphere. Currently this software has extracted over 4 million entries of interaction information from about 17 million registered medical references. It facilitates a systematic understanding of reactions within living organisms by sorting and relating information in accordance with research objectives.

This software makes it possible to efficiently construct original drug-design concepts by negating the need to read a huge amount of medical literature and by providing visual renderings of complex relationships that could not be inferred from only individual reference materials. It can also enable side effects to be estimated to some extent. Of course, any new hypothesis must still be tested by actual experiment.

3.2 Chemical compound design

A disease can be suppressed by controlling the function of a protein identified as a target. This is one basic function of a drug and the familiar "key and keyhole relationship" is well known from hay fever pamphlets and other sources of information. If a causative agent (key) binds with a highly reactive part (keyhole) of an allergy-related protein, substances that trigger itching will be produced. Consequently, if a substance that can strongly bind with that protein and cover the keyhole before the causative agent can be found, a hay fever drug can be obtained.

IT-based drug design makes use of three-dimensional structural information (coordinates of individual atoms, etc.) about protein and chemical compounds and molecular-simulation technology to enable the design of chemical compounds, including unknown virtual substances. Nevertheless, figuring out how to design a chemical compound efficiently and effectively when more than 10^{26} trial calculations could theoretically be required is not without complications.

In response to this situation, we have established Optimum Packing of Molecular Fragments (OPMF)[3] technology that divides the molecules making up a drug into partial structures and treats these structures as abstract fragments in order to search for a massive assembly by assembling a relatively small number of abstract fragments. Individual abstract fragments are formed by a landmark technique as follows. First, a stable position where the target protein can easily bind is determined, and next, a chemical-compound scaffold is formed by interconnecting multiple abstract fragments. These abstract fragments are then replaced by specific atoms to return the substance to the real world. During this design process, multiple molecular simulation techniques are used depending on specific objectives, and drug candidates are sorted, taking into account suitability as a drug.

In this way, an exhaustive virtual search of chemical compound space can be performed while checking binding conditions with the target protein at the atomic level, and the number of drug candidates can be narrowed down to a few hundred or even a few chemical compounds.

3.3 Prediction of medicinal effects

The ability to precisely predict the extent to which a designed chemical compound will be effective can contribute to decision making with regard to actual synthesis. However, "high medicinal effect" simply means "strongly reacts at low concentrations", and it is extremely difficult to predict reactions precisely in what is mostly a water state.

To overcome this problem, we have developed and begun to market MAPLECAFEE (Massively Parallel Computation of Absolute Binding Free Energy with well Equilibrated System)[4] using molecular-dynamics (MD) simulations. This software takes the protein, chemical compound, and water—the participants in the reaction—and rigorously calculates the contributions that the atomic force, electrostatic force, and entropy of all atoms make to the energy term. These calculations, which would theoretically take several tens of years by simple MD calculations, can be performed in a realistic period of time by using massively parallel MD calculations and statistical estimation. Calculations that require several thousand jobs can be accomplished using a large-scale personal computer cluster environment and an efficient job-management mechanism.

The above approach enables the prediction of binding energy comparable to the measured value of an actually synthesized compound. Unlike other techniques that are limited to measuring relative energies between similar chemical compounds, this technique features precise calculation of absolute binding energy even without the presence of related chemical compounds. This is a crucial technology for the design of chemical compounds in IT-based drug design.

4. Practice of IT-based drug design

To test the usefulness of the IT-based drug design, we attempted to reproduce the existing drugs listed in **Table 1** for which structure and bioactivity have been published. As reflected by the results for Gleevec shown in **Figure 2**, we found that a structure showing very good agreement with the actual structure could be created for all the existing drugs in the list.

These results show that the proposed technique can be applied even for proteins having different properties. They also show that the technique can be used to find a chemical compound that, while having a structure different than that of an existing drug, can be predicted to have similar bioactivity. This means that the technique is used not only for designing new chemical compounds but also for modifying existing drugs. This characteristic is worth studying as a means of solving the "2010 problem" described at the beginning of this paper.

Table 1
Existing drugs targeted for reproduction experiment.

Existing drug	Target disease	Target protein
Gleevec	leukemia	Bcr-Abl
Iressa	lung cancer	EGFR
Viagra	ED	phosphodiestrease5
Kaletra	HIV	HIV protease

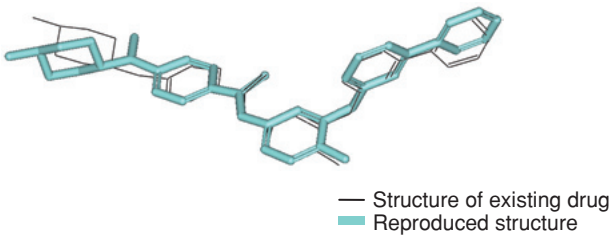

Figure 2
Reproduction of Gleevec.

5. Future directions

Our group is currently involved in drug-design projects targeting allergosis, breast cancer, and common diseases.

The drug-design project for allergosis is a collaborative business with a company that can perform chemical-compound synthesis, evaluation tests, and animal testing. As there is no existing drug for the target protein in question, a few pharmaceutical companies are very interested in this project.

The drug-design project for breast cancer is an independent study on our part. Although there is no approved drug as yet, five drugs have previously been announced by other companies as being under development, so we must clarify the differentiating features of our product.

The drug-design project for common diseases is a joint study with a university. The target of this project is a novel protein that we have identified using our Genesphere software. The university side of the project will examine the validity of this protein as a target and analyze its crystal structure before beginning with chemical-compound design.

Fujitsu has been developing and marketing computational-chemistry software packages since 1983. The IT-based drug design technique introduced here is a culmination of profound technologies achieved by great efforts and know-how cultivated over many years.

Looking to the future, Fujitsu plans to go beyond the sale of software and hardware and move in the direction of the intellectual property business by forming alliances with experimental partners (joint research or joint ventures) and licensing chemical compounds that we ourselves have designed.

6. Conclusion

We see our mission as bringing about breakthroughs in drug discovery through the use of computers and contributing either directly or indirectly to society as a bridge between science and industry. It would also give us great pleasure if our achievements could impress upon experimental researchers the usefulness of computers and help accelerate their research.

References

1) The Center for Advanced Medical Engineering and Informatics, Osaka University: 2007 Global COE program. *http://www.mei.osaka-u.ac.jp/gCOE/english/index.html*
2) S. Kinoshita et al.: BioCreAtIvE Task1A: entity identification with a stochastic tagger. BMC Bioinformatics 6 (Suppl. 1): S4, 2005. *http://www.biomedcentral.com/1471-2105/6/S1/S4*
3) Advances in Structure-Based Drug Design (SBDD) + Solutions Guide, 2007 Edition: T. Kaminuma (editor), Chem-Bio Informatics Society, March 16, 2007 (in Japanese). *http://www.cbi.or.jp/cbi/advertise/ad_solutionSBDD.html*
4) H. Fujitani et al.: Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys*., Vol.123, Issue.8, 084108 (2005).

Shunji Matsumoto
Fujitsu Ltd.
Mr. Matsumoto received the B.S. degree in Applied Physics from the University of Tokyo, Tokyo, Japan in 1983. He joined Fujitsu Ltd. in 1983 and has been engaged in research and development of application software for artificial intelligence (AI) and bio-chemical sciences. He is a member of the Information Processing Society of Japan.