Impact of the Novel Protein Structure Prediction on R&D Process in Drug Discovery and Development

Yoko Ishino
Graduate School of Innovation & Technology Management, Yamaguchi University, Ube, Japan, 7558611
(E-mail: ishino.y@yamaguchi-u.ac.jp)

Abstract: In general, R&D process for drug discovery and development can be classified into two phases, a search phase and a development phase. Since the search phase is the most creative step playing a meaningful role in the drug innovation cycle, this study deals with the search phase, where unknown conformations of drug-target proteins have long desired to be elucidated with accuracy. This study focuses on G-protein coupled receptors (GPCRs) known as highly promising drug-target proteins. To date, the 3D structures of many of GPCRs have not been resolved yet. Therefore, we propose a new prediction method for conformations of active-state GPCRs, which are difficult to predict accurately by existing prediction methods. The proposed method was validated using human leukotriene B4 receptor BLT1 as a sample GPCR. The obtained accurate information on structures of GPCRs is expected to accelerate the research activity by being used in the computer-assisted drug design.
Key words: R&D Process; Drug discovery and development; Protein structure prediction; GPCR

1 Introduction

In the pharmaceutical industry, the annual revenue can exceed $1 billion when a new breakthrough drug is successfully developed. Such large-scale drugs are called “blockbusters.” However, this type of spectacular success is achieved very rarely. In general, considerable time and expense are required to develop a new drug. If a new breakthrough drug is successfully developed, “hitting the jackpot,” it can accumulate sufficient profit to support the company’s operational needs for years; at times, even decades. However, in most cases, companies are unable to recover their investment costs. Based on this type of relationship between investment and returns, the development of pharmaceutical products is often like a treasure hunt and is called a “high risk, high return” style of business. This study focuses on the upstream part of the pharmaceutical R&D process that has the foregoing characteristics.

In addition, pharmaceutical products can be broadly classified into two categories, prescription drugs and over-the-counter drugs. The former is the subject of discussion in this study. Furthermore, this study discusses the R&D process of new pharmaceutical products containing a novel active compound, which has a chemical structure different from existing drugs.

2 R&D Process in Drug Discovery

In order to receive decisive approval as a new drug, a drug candidate should finally meet the various requests, including efficacy, safety, pharmacokinetics (absorption, distribution, metabolism, and excretion), physical properties, stability, and economical efficiency. Many challenges are to be expected during the R&D process. A survey conducted by the Japan Pharmaceutical Manufacturers Association (http://www.jpma.or.jp/) indicates that the probability of successfully bringing a drug to market (the ratio of the compounds that finally being brought to market against the candidate compounds that existed at the starting point of R&D process) is less than 0.01%, an extremely low number. This is caused due to the difficulty in accurately assessing the efficacy and safety of a drug candidate when research starts. Therefore, the entire R&D process can take as long as ten to twenty years and requires enormous expenses, costing tens of billions of yen.

In general, the R&D process for drug discovery and development can be broadly classified into two phases: a “search phase,” in which not a few compounds are created or discovered as candidates for a new drug, and a “development phase,” in which the selected candidate is further evaluated and developed into a product and then clinical investigation is performed [1]. In the search phase prior to entering into the development phase, a literal "search" is performed to find candidate compounds. Specifically, an evaluation system called a screening system is first constructed to screen for compounds by assessing presence and extent of the targeted biological activity, and the screening information obtained is then used in order to discover several compounds that exhibit some activity. They are called hit compounds. Next, a lead compound is identified using the structural information of the hit

compounds. The chemical structure of a lead compound is used as a starting point for chemical modifications in order to improve potency or pharmacokinetic parameters. This process is called the lead compound optimization.

On the other hand, in the development phase, sequential preclinical and clinical studies are conducted using the compound selected in the search phase. In preclinical studies, criteria like efficacy, pharmacology, pharmacokinetics, and safety are tested in laboratory animals. If pharmaceutical action in humans can be expected and prospects of safe use are established based on the results of the preclinical studies, clinical studies are conducted to examine the effects in humans. Clinical studies have three sequential phases (Phase I–Phase III), with the study scale increasing from one phase to the next. After the conclusion of clinical trials, the drug candidate is submitted to the Ministry of Health, Labor, and Welfare, and put on the market as a new drug after approval.

A comparison of research processes in the search and development phases bespeaks the following differences. The search phase lasts for approximately 2 to 5 years, while the development phase lasts for 7 to 15 years. In the search phase people have a large degree of freedom in conducting research and a hope of possibly being able to reduce the amount of time and money, although establishing a preliminary forecast is difficult at this stage of R&D activity. On the other hand, in the development phase people can forecast the research progress with comparative ease because each test is strictly defined beforehand. However, the development phase has little room to save on time or money. In terms of the probability of a compound progressing from the current stage to the next stage, the search phase has a much higher amount of uncertainty than the development phase. Additionally, success in the search phase mostly depends on an individual's ability, experience, and luck, but success in the development phase is associated with a systematic operation, where people act in a group.

Because of these characteristics, research management in the search phase might be considered difficult. However, the search phase is of critical importance, as indicated by Alfonso Gambardella in his book reviewing pharmaceutical research in the US [2], "The generation of new drugs depends in large measure on activities that occur at the outset of the R&D process. Early research stages play a more meaningful role than in other industries, and they are the most creative steps of the drug innovation cycle." Therefore, this study focuses on the search phase to propose a point of view for improving the efficiency and accuracy of screening.

3 3D Structure of Target Proteins in Drug Discovery

Many pharmaceutical companies have been dealing with low-molecular-weight compounds as main objective drugs, where the search phase of R&D activity contains the following three processes. The first step (S1) is random screening to find hit compounds that interact with a target protein using high-throughput screening (HTS). HTS allows us to quickly select hit compounds out of millions of various compounds in the compound library previously provided, using robotics, data processing and control software, and sensitive detectors. In general, tens or even hundreds of thousands of compounds are subjected to HTS, with the probability of finding a hit compound being approximately 0.01%. In the second step (S2), a lead compound having a basic structure for the desired pharmacological activity is identified using the structural information of hit compounds found in S1. This process is extremely important because the quality of the lead compound determines the speed of the subsequent research and development. The third step (S3) is the lead compound optimization, in which the chemical structure of a lead compound is used as a starting point for chemical modifications in order to improve potency, pharmacokinetic and physicochemical parameters, safety, and so on.

In recent years, the computer-assisted drug design (CADD) has become widely used in the search phase of R&D activity. CADD uses computational chemistry like molecular mechanics or molecular dynamics to predict whether a given molecule will bind to a target protein and if so how strongly. In theory, CADD would help streamline the processes in each step of the search phase (S1-S3). In the case of S1, CADD can be used for selecting a suitable variation of chemical compounds in the compound library used in HTS. In S2, CADD would help identify a lead compound by investigating how hit compounds bind to a target protein and how they interact with each other and then by extracting pharmacophore information that directly ties to the expression of pharmacological activity. In S3, CADD would help optimize a lead compound by predicting how effective the chemical modification of a lead compound is for the interaction with a target protein.

However, at present, it is not always possible to use CADD in the search phase. When a candidate compound is computationally docked to a target protein, highly accurate values of interaction energy

between them are needed to conduct CADD successfully. This requires high-accuracy 3D structure data of a target protein. However, many proteins remain poorly understood due to technical difficulties in protein crystallization, and consequently there is an unavailability of structural information of proteins. That is the reason why the use of CADD does not guarantee favorable results. It has long been a hope that the unknown 3D structure of the drug target protein could be accurately predicted and the predicted structure information could be used for the screening process. It should accelerate the research activity in the search phase.

The present study focuses on G-protein coupled receptors (GPCRs) known as highly promising drug target proteins. To date, the 3D structures of many of GPCRs have not been resolved yet. Therefore, we propose a new prediction method for GPCR active state conformations, which are difficult to predict with high accuracy by existing prediction methods.

4 New Method for Predicting 3D Structure of G-protein-coupled Receptors

4.1 G-protein-coupled receptors (GPCRs)

G-protein-coupled receptors (GPCRs), the largest family of membrane proteins (around 800 in humans), are activated by a wide range of stimuli, including hormones, neurotransmitters, ions, odorants, and photons of light[3]. All GPCRs possess seven transmembrane helices and are localized to the cell membrane. GPCRs are involved in a variety of biological and pathological processes[4-7]. Consequently, they are one of the largest classes of drug targets. Approximately half of all current pharmaceuticals target GPCRs. However, only 10% of the GPCRs, excluding olfactory receptors are targeted by marketed drugs, indicating the possibility that the remaining 90% of GPCRs are available as targets for the treatment of human disease[8].

To date, only a limited number of GPCRs, such as bovine rhodopsin and human beta-adrenergic receptor, have been resolved at high resolution by X-ray crystallography[9-13]. For predicting 3D protein structure, homology modeling is the most commonly employed method that allows the prediction of target GPCR 3D structures by using the crystal structures as a template[14]. Nevertheless, great cautions are needed when utilizing the homology-based models for GPCRs since helix kinks are often different in different receptors. Modeling the subtle distinctions, which is essential for docking calculation in CADD, remains a major challenge[15]. Additionally, since all known GPCR structures are in inactive state where the protein binds to its antagonists, the homology modeling has practically no predictive power for GPCRs in active state where the protein binds to its agonists.

It has been reported that applying an appropriate molecular dynamics (MD) simulation after conducting homology modeling could predict GPCR structures in active state in view of the induced-fit mechanism[16]. However, this approach needs to impose appropriate structural restrictions prior to the MD simulation, which requires both skills and experience of a MD specialist. Moreover, the MD method can normally simulate the movement of atoms on the microsecond time scale as a limit; although structural changes in induced fit require calculations for fluctuations greater than milliseconds. Due to its time-consuming nature, it is virtually impossible to simulate all processes involved in structural changes using only MD on common computers. For this reason, currently no high-accuracy methods are available for predicting the GPCR structure in active state.

4.2 New method for prediction of 3D structure of GPCRs

We propose a new prediction method for conformations of active-state GPCRs, which are difficult to predict accurately by existing prediction methods. The framework of the method comprises three stages, namely: (i) gaining of an initial structure by homology modeling in which known inactive-state GPCR structures are used as a template; (ii) calculation of an average structure of GPCR fluctuations at the levels of several hundred picoseconds to nanoseconds, as calculated from MD simulations; and (iii) search for one of the best 3D structures of an active-state GPCR by using the evolutionary computing technique including modularized docking simulations between a GPCR and its known agonists. The core idea of the framework is that a MD simulation is used to calculate an average 3D coordinates of all atoms of a GPCR protein against heat fluctuation on the picosecond or nanosecond time scale, and then real-coded genetic algorithm including receptor-ligand docking simulations functions to determine the rotation angle of each helix of a GPCR protein as a movement on wider time scale.

5 Case Study Using Leukotriene Receptor

5.1 Leukotriene teceptor

The method was validated using human leukotriene B4 receptor BLT1 as a sample GPCR.

Leukotriene (LT) is a bioactive lipid that serves as an important mediator of host defense, though it is also known to be implicated in bronchial asthma as a pathogenetic or precipitating factor. So far, four types of LT receptors have been cloned. BLT1 is one of them.

5.2 Experimental procedure

We first selected LT4R1_HUMAN from MODBASE[17], a database of protein 3D structures predicted by homology modeling, as the BLT1 structure after homology modeling in which bovine rhodopsin was used as a template. After annealing for atomic relaxation, MD simulations were run using TINKER (ver. 4.2)[18]. Simulation conditions used were as follows: force field = AMBER99, temperature = 310 K, pressure = 1 atm, and time step = 1.0 fs. Next, based on the previous finding that "the rotational motion of helices no. 3, 5, 6 (denoted as TM3, TM5, TM6) of LT receptor are particularly important in contributing to agonist-receptor binding in active state[19, 20]," we allowed just these three helices to revolve freely in the subsequent evolutionary search combining real-coded GA and receptor-ligand docking simulations. In other words, each candidate individual for solution in the evolutionary search has real number vectors representing the rotational angles of TM3, TM5 and TM6. In the evolutionary search, the following operations are repeated until the search finished: (i) generation of individual populations as candidates for solution, (ii) reconstruction of 3D structures and their structural optimization for each individual, (iii) binding simulations using GOLD (ver. 3.1)[21] for each individual, and (iv) evaluation and selection.

5.3 Results

A Dell computer with an Intel Xeon CPU 3.6 GHz (dual processor) was used for computation. For MD processing of 219.3-ps simulations, 524 hours were required. We investigated the changes in GPCR molecular energy values (kcal/mol) for every 0.1 ps time elapsed in MD. Low energy values suggest that the structures are stable. At the level of about 200 ps, energy values cease to fall, which suggests that a region of stability has been reached. As our study focused on MD simulations for structural changes on a short time scale (meaning that the objectives were to simulate relaxation), we considered 200 ps to be a sufficient level, and therefore stopped at 219.3 ps.

For subsequent evolutionary computation, 899 hours were required for calculations for 200 generations. Figure 1 shows the state of binding between the LT receptor and its agonist in one of the best 3D structures. The 3D structures of binding are shown in Figure 1(A) with a view from above the cell plasma membrane, and in Figure 1(B) with a cross-section of the membrane (the upper part being the extracellular space). In order to highlight receptor-ligand interactions, the main chains (backbones) of the helices are shown as ribbons, while the LT agonist 12-keto-LTB4 is drawn as a ball-and-stick structure. From Figure 1, it is clear that the LT agonist binds to a recessed region (pocket) formed by TM3, TM5, and TM6.

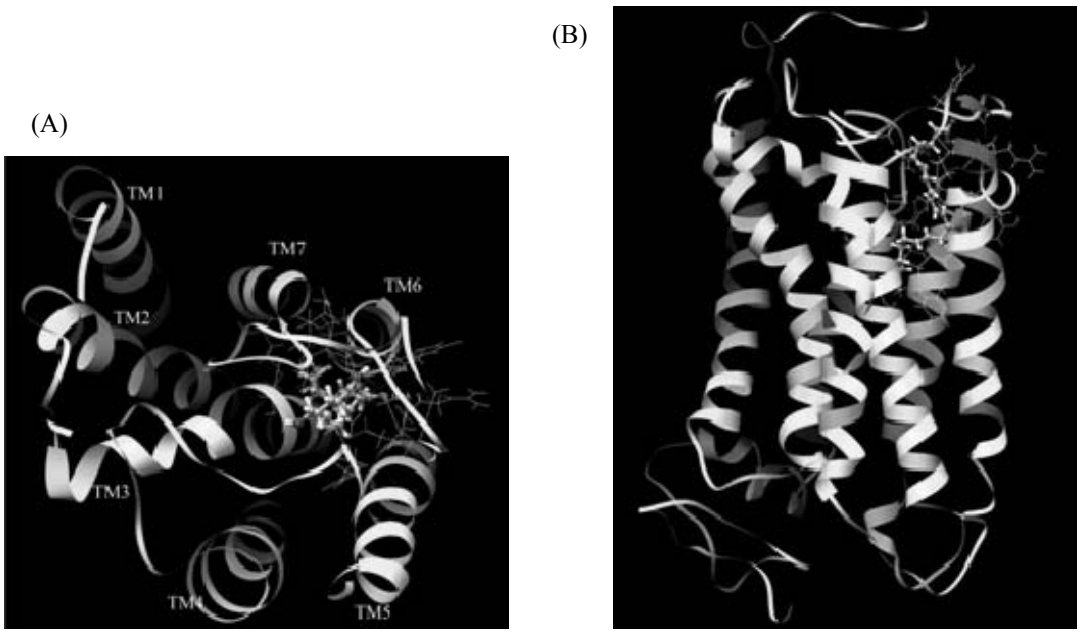

Figure 1 The Best 3D Structure of Active-State BLT1

In order to validate LT receptor structures obtained by our method, we further used 5 agonists and 11 antagonists to respectively perform binding simulations with BLT1 where BLT1 was in the optimal conformation. The maximum docking scores for each ligand are shown in Figure 2. The agonist group produced visibly higher scores than the antagonist group. The proposed method predicted the BLT1 conformation accurately, to the extent of discriminating agonists from antagonists. This suggests that the GPCR conformation obtained by the proposed method can be used to find hit compounds out of many compounds.

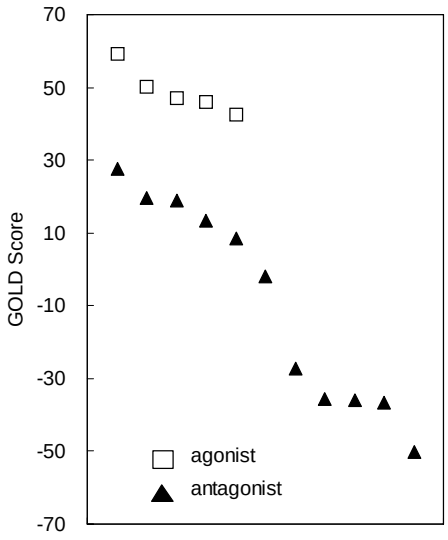

Figure 2 GOLD Score When Docking with Other Agonists and Antagonists

6 Discussion and Conclusions

This study proposed a new method for finding one of the best 3D structures of active-state GPCRs at a level of accuracy acceptable to CADD. The method was validated using human leukotriene B4 receptor BLT1 as a sample GPCR.

The search phase of the pharmaceutical R&D activity contains the three steps (S1-S3) stated in Section 3. The 3D structure of a target GPCR obtained using the proposed method is expected to be exploited in all steps from S1 to S3: in the process of constructing the compound library in S1, in the docking study in S2 for extracting pharmacophore information or for identifying a lead compound, and in the process of optimizing the lead compound in S3. In the search phase of the pharmaceutical R&D activity, the process of finding hit compounds and identifying a lead compound is the key to lead the R&D activity to successful conclusion. The proposed method supports the key process by being used in the foregoing processes.

Eventually, using the obtained 3D structure of a target GPCR in the search phase would have a beneficial impact on human activities in drug discovery and development. First, researchers would be able to use CADD systems to analyze experimental results, which provides us the possibility that the phenomena occurring in the experiments can be explained using a theory. This would facilitate making plans of following experiments in the search process, so that the efficiency in the search process would increase. The increased efficiency enables us to test much more chemical compounds for hit compounds within the same period of time, thereby increasing the probability of finding a lead compound. Moreover, greater ease in theoretically explaining experimental results would accelerate knowledge sharing among members of a research organization. It might facilitate managing research resources in the search phase of the R&D activity.

References

[1] Kuwashima K. Management of Uncertainty: Solution for Pharmaceutical R&D Management[M]. Tokyo: Nikkei Business Publications, Inc, 2006 (In Japanese)

[2] Gambardella A. Science and Innovation: The U.S. Pharmaceutical Industry during the 1980s[M]. Cambridge, England: Cambridge University Press, 1995.

[3] Lundstrom K. Structural Genomics and Drug Discovery[J]. J. Cell. Mol. Med., 2007, 11:224–238
[4] Lowes V.L., Ip N.Y. and Wong Y.H. Integration of Signals from Receptor Tyrosine Kinases and G Protein-Coupled Receptors[J]. Neurosignals, 2002, 11: 5–19
[5] Miller K.J., Murphy B.J. and Pelleymounter M.A. Central G-Protein Coupled Receptors (GPCR)s as Molecular Targets for the Treatment of Obesity: Assets, Liabilities and Development Status[J]. Curr. Drug Targets CNS Neurol. Disord., 2004, 3: 357–377
[6] Parker S.L., Parker M.S., Sah R. and Sallee F. Angiogenesis and Rhodopsin-Like Receptors: A Role for N-Terminal Acidic Residues?[J]. Biochem. Biophys. Res. Commun., 2005, 335: 983–992
[7] Rayasam G.V., Tulasi V. K., Davis J. A. and Bansal V. S. Fatty Acid Receptors as New Therapeutic Targets for Diabetes[J]. Expert Opin. Ther. Targets, 2007, 11:661–671
[8] Kroeze W.K., Sheffler D.J. and Roth B.L. G-Protein-Coupled Receptors at a Glance[J]. J. Cell Sci., 2003, 116:4867–4869
[9] Cherezov V., Rosenbaum D.M., Hanson M.A., Rasmussen S.G., Thian F.S. and et al. High-resolution Crystal Structure of an Engineered Human Beta2-Adrenergic G Protein-Coupled Receptor[J]. Science, 2007, 318:1258–1265
[10] Warne T., Serrano-Vega M.J., Baker J.G., Moukhametzianov R., Edwards P.C. and et al. Structure of a Beta1-Adrenergic G-Protein-Coupled Receptor[J]. Nature, 2008, 454: 486–491
[11] Jaakola V.P., Griffith M.T., Hanson M.A., Cherezov V., Chien E.Y. and et al. The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist[J]. Science, 2008, 322:1211–1217
[12] Shimamura T., Hiraki K., Takahashi N., Hori T., Ago H. et al. Crystal Structure of Squid Rhodopsin with Intracellularly Extended Cytoplasmic Region[J]. J. Biol. Chem., 2008, 283: 17753–17756
[13] Palczewski K., Kumasaka T., Hori T., Behnke C.A., Motoshima H. et al. Crystal Structure of Rhodopsin: a G Protein-Coupled Receptor[J]. Science, 2000, 289:739–745
[14] Moult J., Fidelis K., Kryshtafovych A., Rost B., Tramontano A. Critical Assessment of Methods of Protein Structure Prediction-Round VIII[J]. Proteins, 2009, 77:1-4
[15] Michino M., Abola E. GPCR Dock 2008 participants, Brooks, C.L. 3rd., Dixon, J.S., Moult, J., Stevens, R.C. Community-wide Assessment of GPCR Structure Modelling and Ligand Docking: GPCR Dock 2008[J]. Nat. Rev. Drug Discov., 2009, 8: 455-463
[16] Gouldson P.R., Kidley N.J., Bywater R.P., Psaroudakis G., Brooks, H.D. and et al., Toward the Active Conformations of Rhodopsin and the Beta2-Adrenergic Receptor[J]. Proteins, 2004, 56: 67-84
[17] Pieper U., Eswar N., Webb B.M., Eramian D., Kelly L. et al. MODBASE: A Database of Annotated Comparative Protein Structure Models and Associated Resources[J]. Nucleic Acids Res., 2009, 37: D347-354
[18] Ponder J.W. Tinker Software Tools for Molecular Design. Version 4.2., 2004. http://dasher.wustl.edu/tinker
[19] Altenbach C., Kusnetzow A.K., Ernst O.P., Hofmann K.P., Hubbell W.L. High-resolution Distance Mapping in Rhodopsin Reveals the Pattern of Helix Movement due to Activation[J]. Proc Natl Acad Sci USA, 2008, 105:7439-7444
[20] Sabirsh A., Bywater R.P., Bristulf J., Owman C. and Haeggström J.Z. Residues from Transmembrane Helices 3 and 5 Participate in Leukotriene B4 Binding to BLT1[J]. Biochemistry, 2006, 45:5733-5744
[21] Verdonk M.L., Cole J.C., Hartshorn M.J., Murray C.W., Taylor R.D. Improved Protein-Ligand Docking using GOLD[J]. Proteins, 2003, 52:609-623