2011 International Conference on Bioscience, Biochemistry and Bioinformatics
IPCBEE vol.5 (2011) © (2011) IACSIT Press, Singapore

In Silico 3D Structure Prediction of Activated GPCRs as a Drug Target

Yoko Ishino
Collaborative Research Center,
Hiroshima University
Higashi-Hiroshima, Japan
y8ishino@hiroshima-u.ac.jp

Takanori Harada
Graduate School of Biomedical Sciences,
Hiroshima University
Hiroshima, Japan
tharada@hiroshima-u.ac.jp

Abstract— **This paper describes a novel method to predict the biologically activated structures of G-protein-coupled receptors (GPCRs) with high accuracy so that the structures predicted can be used in in silico virtual screening. We propose a new method for modeling GPCR thermal fluctuations, where conformation changes of the proteins are modeled by combining fluctuations on different time scales. The core idea of the method is that a molecular dynamics 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 the evolutionary computation 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. The method was validated using human leukotriene B4 receptor BLT1 as a sample GPCR. Our study demonstrated that the established method was able to provide the appropriate 3D structure of the activated GPCR to dock with its agonists.**

Keywords- GPCR; protein structure prediction; evolutionary computation; receptor-ligand docking simulation; molecular dynamics simulation

I. INTRODUCTION

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 [1]. All GPCRs possess seven transmembrane helices and are localized to the cell membrane. GPCRs are involved in a variety of biological and pathological processes such as development and proliferation [2], neurological disorders [3], angiogenesis [4], and metabolic disorders [5]. 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 [6].

In general, substances that possess affinity for receptors are called ligands, which are further classified into agonists and antagonists depending on the effects that they induce upon receptor-ligand binding. The agonist-bound GPCR form is the biologically active one, whereas antagonist-bound GPCRs are biologically inactivated. Structural studies of GPCR-ligand interactions are still challenging due to the difficulty of functional GPCR production and receptor stability. To date, only a limited number of inactivated GPCRs, such as bovine rhodopsin and human beta-adrenergic receptor, have been resolved at high resolution by X-ray crystallography [7-11].

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 [12]. Nevertheless, great cautions are needed when utilizing the homology-based models for detail GPCR structural and functional annotations since helix kinks are often different in different receptors. Modeling the subtle distinctions, which is essential for ligand docking and screening, remains a major challenge [13]. Additionally, since all known GPCR structures are the inactive type, the homology modeling has practically no predictive power for the activated structure of GPCRs.

It has been reported that applying an appropriate molecular dynamics (MD) simulation after conducting homology modeling could predict activated GPCR structures in view of the induced-fit mechanism [14]. 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 activated structure of GPCRs.

This paper describes a novel method to predict the activated GPCR structures with high accuracy so that the structures predicted can be used in in silico virtual screening. First, we have developed a new method for modeling thermal fluctuations in GPCRs by using combinations of fluctuations with different time scales. Based on this model, we propose a novel computational method to search for one of the best 3D structures of an activated GPCR.

II. PROPOSED MODEL AND FRAMEWORK

Proteins undergo thermal fluctuations of considerable magnitude. The time scale of thermal fluctuations in proteins ranges from femtoseconds to minutes, or even higher-order units [15]. In terms of spatial considerations, various levels of fluctuations exist, which include rapid motions within

microspace (such as thermal oscillations between atoms and rotations of amino acid side chains) and relatively slow but large structural changes (such as local unfolding). Not all these motions can be simulated because of the physical limits of currently available MD calculations. We therefore propose a new method for modeling GPCR thermal fluctuations, in which fluctuations are combined on different time scales. In this approach, we first measure the oscillations and fluctuations of various atoms at the picosecond and nanosecond levels to obtain native structure within a short time span. We then simulate the thermal fluctuations of GPCRs in terms of their characteristic rotational motions around helical axes, which represent fluctuations of greater time scales. This approach is based on the previous findings that GPCRs are restricted in their motions within the cell membrane lipid, and that intracellular structural changes seem to occur by rotational motions around the axis of each transmembrane helix when a GPCR is biologically activated [16]. Although this method is not a rigorous chemical calculation, our approach has the clear advantage of reducing calculation load.

Based on this model, we propose a framework for structural search which comprises three stages, namely: (i) acquisition of initial structures in homology modeling by using existing inactivated GPCR structures as a template; (ii) acquisition of average structures 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 structure of an activated GPCR by using the evolutionary computing technique including modularized docking simulations between GPCRs and their 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.

III. Methodology for Prediction of GPCR 3D Structure

The detailed procedure of the proposed structural search method is described as follows.

1. Determination of an initial structure: Rhodopsin or beta-adrenergic receptor is used as a template. After alignment of the primary amino acid sequences of the template and target GPCR, the inactivated structure of the target GPCR is generated by regular homology modeling, which is an initial structure.
2. Average molecular structures with respect to fluctuations on a short time scale: An appropriate amount of water molecules is added to the surroundings of the target GPCR. The fluctuations of GPCR molecules are then calculated by MD simulations at the level of several hundred picoseconds. The average structures are defined as those for which the energy values stabilize.
3. Molecular structural changes with respect to rotational motions of helices: Hypothetical rotational motions around the axes of helices constituting the target GPCR (as fluctuations of greater time scale) are generated. These are fed into real-coded genetic algorithms (real-coded GA) [17] including simulations of binding with known ligands to determine optimal values. Details of the steps are described as follows. In addition, the schemes for generation alternation and offspring generation are shown in Fig 1.

3-1. Generation of initial populations: Individuals are assigned to real-number vectors representing rotational angle changes relative to the average structures of various helices. These real-number vectors are treated as values within the space adjoining the two ends of the search space (i.e., toroidal space). In real-coded GA, the number of individuals in population pools is taken to be m, for which the initial populations are randomly generated.

3-2. Generation of offspring for each generation: Parents (two individuals) for random crossover of population pools are selected to match with a third parent, so as to generate offspring by the unimodal normal distribution crossover (UNDX) method [18]. The third parent is used to determine the standard deviation of normal distribution in UNDX. This procedure is repeated n times to provide n offspring individuals. In total, the number of individuals in a generation is n+2 (offspring and two parents).

3-3. Selection in each generation: To evaluate each individual, receptor-ligand docking simulations are conducted using a known agonist and a known antagonist. For each individual, genes are translated into their corresponding 3D structures, while their side chains are optimized before the binding simulations. An evaluation function has been established to represent how good the binding is. For a single individual, ligand-GPCR docking simulations are conducted up to k times, and each evaluation score is recorded. Upon obtaining the top score for each individual, the scores are then used to rank all individuals in a generation. Two individuals high in ranking are selected (elite strategy) to replace the two parents in the population.

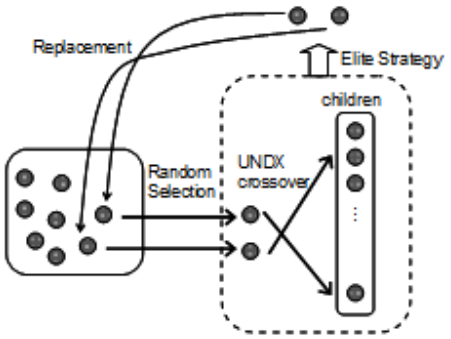


Figure 1. Schemes for generation alternation

3-4. Reiteration: Steps 3-2 and 3-3 are repeated until some stop criteria are met.

In our proposed search method, UNDX is used as a crossover method in real-coded GA, and MGG (minimal generation gap) is used for generation transition. MGG has good compatibility with UNDX and maintains the diversity in the population [19].

IV. Case Study

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

A. Leukotriene Receptor

As an example to illustrate our method, leukotriene (LT) receptor (one of the many GPCRs) was chosen for calculations. Leukotriene 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. The amino acid sequence of one of these receptors, the high-affinity human LTB4 receptor BLT1 (GPCRDB entry ID: LT4R1_HUMAN; Swiss-Prot entry ID: Q15722), was used in our experiments.

B. Experimental Procedure

We first used MODBASE [20] to obtain the LT receptor structure by running homology modeling with rhodopsin (Protein Data Bank ID: 1L9H) as a template for the initial 3D structure of LT4R1_HUMAN. MODBASE is a protein 3D structure database for homology modeling maintained by the University of California at San Francisco.

After annealing for atomic relaxation, MD simulations using TINKER (ver. 4.2) [21] were run on conditions based on structural restriction criteria proposed by Gouldson et al. [14]. Other simulation conditions used were force field = AMBER99, temperature = 310 K, pressure = 1 atm, and time step = 1.0 fs. The mean atomic positions were determined from continuous simulations upon achieving a stable state where there is no drop in energy in the GPCR molecules.

Next, we made use of a previous finding from amino acid residue substitution analysis and spectroscopic experiments that "helices no. 3, 5, 6 (denoted as TM3, TM5, TM6) of bioactive LT receptor are particularly important in contributing to ligand-receptor binding [22, 23]," and for only these 3 helices we defined 3D structures capable of free rotational motions as the individuals in real-coded GA. In other words, these individuals, which provide the coding information, were defined as real number vectors representing the rotational angles of TM3, TM5 and TM6. In this search method combining real-coded GA and binding simulations, the following operations were repeated for each generation: (i) generation of individual populations, (ii) reconstruction of 3D structures and their structural optimization for each individual, (iii) binding simulations using GOLD (ver. 3.1) [24] for each individual, and (iv) evaluation and selection. In terms of computational parameters, calculations involved 50 individuals from initial populations, 8 offspring for each generation, and reiteration up to the 200th generation. For evaluation of individuals, separate binding simulations were conducted with 2 compounds (LT agonist 12-keto-LTB4 and LT antagonist pranlukast), from which the maximum GOLD scores were assigned as x_1, x_2, which were defined as follows.

When $x_2 > 0$, evaluation function value = $x_1 - x_2$; and for all cases other than this condition, evaluation function value = x_1.

C. Results and Discussion

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. Fig 2 shows the trends of scores in evolutionary computation (the higher the scores, the better). The optimal solutions for rotational angles after 200 generations were: TM3=11°, TM5=14° and TM6=255°.

Fig 3 shows the state of binding between the LT receptor and its agonist in one of the best 3D structure. The 3D structures of binding are shown in Fig 3(A) with a view from above the cell plasma membrane, and in Fig 3(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 Fig 3, it is clear that the LT agonist binds to a recessed region (pocket) formed by TM3, TM5, and TM6. Upon further analysis regarding the binding state of LT receptor and its ligand,

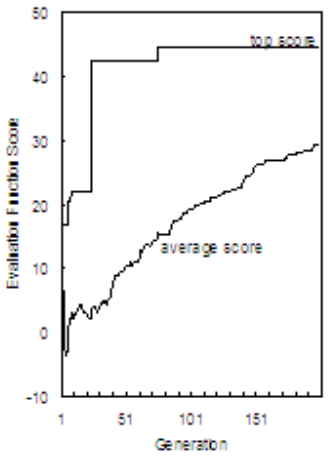
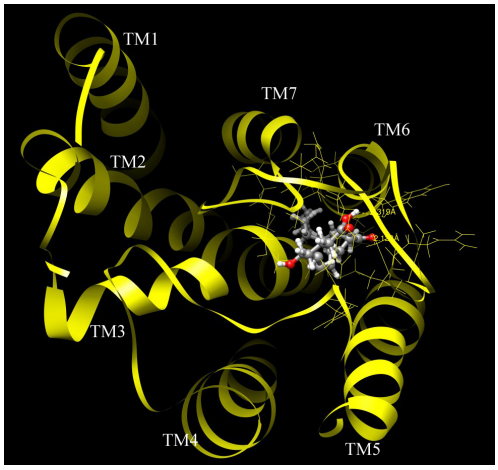


Figure 2. Evaluation function score in the evolutionary computation process

(A)



(B)

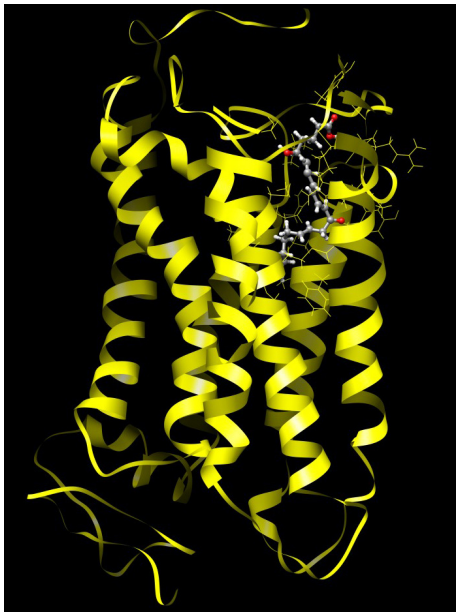


Figure 3. The best 3D structure of BLT1 active form

these helices responsible for interacting with the LT ligand are composed mostly of hydrophobic amino acid residues, suggesting that the LT ligand is drawn into a strongly hydrophobic environment. Since the LT agonist has many hydrophobic groups, it is likely that ligand-receptor binding is contributed mainly by hydrophobic interactions, though several hydrogen bonding sites contributes the binding. The following interactions are enumerated in detail.

- Hydrophobic interactions between TM3 and the alkyl side chain near the central carbonyl (-C=O) group of the LT agonist;
- Hydrophobic interactions between TM5 and the alkyl side chain near the central carbonyl group of the LT agonist;
- Hydrogen bonding between the -OH group in the carboxyl (-COOH) group of the LT ligand and the carbonyl group of Asn241 on TM6 (C=O···H−O intermolecular distance = 2.32 Å.);

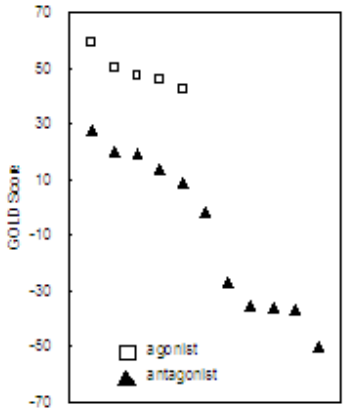


Figure 4. GOLD Score when docking with other agonists and antagonists

- Hydrogen bonding between the central carbonyl group of the LT agonist and the main chain (-N−H peptide bond) of Gly246 on the loop joining TM6 and TM7 (C=O···H−O intermolecular distance 2.13 Å.);
- Hydrophobic interactions between TM6 and all alkyl side chains of the LT agonist.

In addition to TM3, TM5, and TM6, it is speculated that TM7 is also able to contribute to hydrophobic interactions with the alkyl side chains of the LT agonist.

In order to check the validity of LT receptor structures obtained by our method, we further used 5 agonists and 11 antagonists to perform binding simulations with BLT1 that had the optimal solutions, from which the corresponding maximum GOLD scores were determined. The results are as shown in Fig 4. In Fig 4, the antagonists are plotted in order of decreasing scores. The agonist group produced scores that were more visibly than those of the antagonist group.

We then calculated the root mean square deviation (RMSD) of various atoms. Firstly, we obtained initial structures from homology modeling as structure I, MD simulated structures as structure II, and structures generated from post-MD evolutionary computation as structure III to calculate RMSD of 1,316 atoms forming the backbones. The resultant RMSD was 4.49 Å. for structures I and II, 1.55 Å. for structures II and III, 4.63 Å. for structures I and III. Based on these values, it can be deduced that the transition from structure I to structure II involved a large change.

In view of the above, we have developed a search method capable of obtaining the 3D structures of biologically activated GPCRs, which is broadly consistent with known facts of the target proteins and is reasonably persuasive as a model.

V. Conclusions

In this study, we proposed a method for finding one of the best structures of an activated GPCR at a level of accuracy acceptable to virtual screening. The method was

validated using human leukotriene B4 receptor BLT1 as a sample GPCR. Our study demonstrated that the established method including homology modeling, MD simulations, and the evolutionary computing was able to provide the appropriate 3D structure of the activated leukotriene receptor to dock with its agonists.

For future investigation, we plan to use GPCRs other than leukotriene receptor to test the applicability and robustness of our search method. For such investigation, it is important to carefully select the appropriate study targets. Based on protein sequence similarity, GPCRs are commonly divided in three distinct families, A, B and C [25]. In general, the class A or rhodopsin-like GPCRs (to which the leukotriene receptor belongs) are unlikely to pose problems. For other classes of GPCRs, however, the corresponding GPCR template may contain substantially different amino acid sequences, thus calling for extra processing at the preliminary stage of homology modeling.

In addition, as a further avenue of research, we plan to develop new evaluation methods suitable for assessing predicted 3D structures of GPCRs.

ACKNOWLEDGMENT

This work was in part supported by the aid fund from Tateishi Science and Technology Foundation, and also by the Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

REFERENCES

[1] K. Lundstrom, "Structural genomics and drug discovery," J. Cell. Mol. Med., vol. 11, pp. 224–238, 2007.

[2] V.L. Lowes, N.Y. Ip and Y.H. Wong, "Integration of signals from receptor tyrosine kinases and G protein-coupled receptors," Neurosignals, vol. 11, pp. 5–19, 2002.

[3] K.J. Miller, B.J. Murphy and M.A. Pelleymounter, "Central G-protein coupled receptors (GPCR)s as molecular targets for the treatment of obesity: assets, liabilities and development status," Curr. Drug Targets CNS Neurol. Disord., vol. 3, pp. 357–377, 2004.

[4] S.L. Parker, M.S. Parker, R. Sah and F. Sallee, "Angiogenesis and rhodopsin-like receptors: a role for N-terminal acidic residues?," Biochem. Biophys. Res. Commun., vol. 335, pp. 983–992, 2005.

[5] G.V. Rayasam, V.K. Tulasi, J.A. Davis and V.S. Bansal, "Fatty acid receptors as new therapeutic targets for diabetes," Expert Opin. Ther. Targets, vol. 11, pp. 661–671, 2007.

[6] W.K. Kroeze, D.J. Sheffler and B.L. Roth, "G-protein-coupled receptors at a glance," J. Cell Sci., vol. 116, pp. 4867–4869, 2003.

[7] V. Cherezov, D.M. Rosenbaum, M.A. Hanson, S.G. Rasmussen, F.S. Thian, and et al., "High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor," Science, vol. 318, pp. 1258–1265, 2007.

[8] T. Warne, M.J. Serrano-Vega, J.G. Baker, R. Moukhametzianov, P.C. Edwards, and et al., "Structure of a beta1-adrenergic G-protein-coupled receptor," Nature, vol. 454, pp. 486–491, 2008.

[9] V.P. Jaakola, M.T. Griffith, M.A. Hanson, V. Cherezov, E.Y. Chien, and et al., "The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist," Science, vol. 322, pp. 1211–1217, 2008.

[10] T. Shimamura, K. Hiraki, N. Takahashi, T. Hori, H. Ago, and et al., "Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region," J. Biol. Chem., vol. 283, pp. 17753–17756, 2008.

[11] K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, and et al., "Crystal structure of rhodopsin: a G protein-coupled receptor," Science, vol. 289, pp. 739–745, 2000.

[12] J. Moult, K. Fidelis, A. Kryshtafovych, B. Rost, and A. Tramontano, "Critical assessment of methods of protein structure prediction-Round VIII," Proteins, vol. 77, pp. 1-4, 2009.

[13] M. Michino, E. Abola; GPCR Dock 2008 participants, C.L. Brooks 3rd, J.S. Dixon, J. Moult, R.C. Stevens, "Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008," Nat. Rev. Drug Discov., vol. 8, pp. 455-463, 2009.

[14] P.R. Gouldson, N.J. Kidley, R.P. Bywater, G. Psaroudakis, H.D. Brooks, and et al., "Toward the active conformations of rhodopsin and the beta2-adrenergic receptor," Proteins, vol. 56, pp. 67-84, 2004.

[15] H. Saito, Y. Kawase, A. Kira, K. Yamamoto, M. Tanio, and et al., "Surface and dynamic structures of bacteriorhodopsin in a 2D crystal, a distorted or disrupted lattice, as revealed by site-directed solid-state 13C NMR," Photochem Photobiol., vol. 83, pp. 253-262, 2007.

[16] Z.L. Lu, and E.C. Hulme, "A network of conserved intramolecular contacts defines the off-state of the transmembrane switch mechanism in a seven-transmembrane receptor," J Biol Chem., vol. 275, pp. 5682-5686, 2000.

[17] L.J. Eshelman, and J.D. Schaffer, "Real-coded genetic algorithms and interval-schemata," Foundations of Genetic Algorithms, vol. 2, pp. 187-202, 1993.

[18] I. Ono, H. Sato, and S. Kobayashi, "A real-coded genetic algorithm for function optimization using the unimodal normal distribution crossover," J. of Japanese Society for Artificial Intelligence, vol.14, pp. 1146-1155, 1999.

[19] H. Sato, I. Ono, and S. Kobayashi, "A new generation alternation model of genetic algorithms and its assessment," J. of Japanese Society for Artificial Intelligence, vol. 12, pp. 734-744, 1997.

[20] U. Pieper, N. Eswar, B.M. Webb, D. Eramian, L. Kelly and et al., "MODBASE: a database of annotated comparative protein structure models and associated resources," Nucleic Acids Res., vol. 37, D347-354, 2009.

[21] J.W. Ponder, Tinker Software Tools for Molecular Design. Version 4.2., 2004. http://dasher.wustl.edu/tinker

[22] C. Altenbach, A.K. Kusnetzow, O.P. Ernst, K.P. Hofmann, and W.L. Hubbell, "High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation," Proc Natl Acad Sci USA, vol. 105, pp. 7439-7444, 2008.

[23] A. Sabirsh, R.P. Bywater, J. Bristulf, C. Owman, and J.Z. Haeggström, "Residues from transmembrane helices 3 and 5 participate in leukotriene B4 binding to BLT1," Biochemistry, vol. 45, pp. 5733-5744, 2006.

[24] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, and R.D. Taylor, "Improved protein-ligand docking using GOLD," Proteins, vol. 52, pp. 609-623, 2003.

[25] K.L. Pierce, R.T. Premont and R.J. Lefkowitz, "Seven-transmembrane receptors," Nat. Rev. Mol. Cell Biol., vol. 3, pp. 639–650, 2002.