PREDICTION OF THE STRUCTURE OF G-PROTEIN COUPLED RECEPTOR WITH APPLICATION FOR DRUG DESIGN

YOUYONG LI, WILLAM A. GODDARD III[†]

Materials and Process Simulation Center (MC 139-74),
California Institute of Technology, Pasadena CA 91125

G protein-coupled receptors (GPCRs) mediate our sense of vision, smell, taste, and pain. They are also involved in cell recognition and communication processes, and hence have emerged as a prominent superfamily for drug targets. Unfortunately, the atomic-level structure is only available for rhodopsin family. We have recently developed first principles methods MembStruk and HierDock for predicting structures and functions of GPCRs. Here we report the application of those methods to prostanoid receptors and urotensin receptors. We obtain the binding sites consistent with available mutation data and optimize antagonists which have been confirmed experimentally.

1. Introduction

G protein-coupled receptors (GPCRs) mediate senses such as odor, taste, vision, and pain in mammals.[1] In addition, important cell recognition and communication processes often involve GPCRs. The GPCR superfamily is diverse, and sequencing of the human genome has revealed over 850 genes that encode them. The diversity of the GPCRs is matched by the variety of ligands that activate them, including odorants, taste ligands, light, metals, biogenic amines, fatty acids, amino acids, peptides, proteins, nucleotides, lipids, Krebs-cycle intermediates and steroids.

Indeed, many diseases involve malfunction of GPCRs,[2] making them important targets for drug development. Greater than 30 % of all marketed therapeutics act on those receptors. Prostanoids (prostaglandins (PG) and thromboxanes (TX), both metabolites of arachidonic acid)[3] play important physiological roles in the cardiovascular and immune systems and in pain sensation in peripheral systems. They exert a variety of actions in the body through binding to specific cell surface prostanoid receptors and mediate many processes known to be inhibited by non-steroidal anti-inflammatory drugs (NSAIDs). Urotensin II (U-II) is the most potent vasoconstrictor known and plays an important role in cardiovascular regulation upon the urotensin II receptor (UT2R). U-II is also a neuropeptide and may play a role in tumor development.

2

The development of subtype specific agonists and antagonists has been hampered by the lack of 3D structures for GPCR receptors. Here we report the application of MembStruk[4-12] and HierDock to study 3D structure and function of prostanoid receptors and urotensin receptors. We obtain the binding sites consistent with available mutation data, which validates our predicted structures. Meanwhile, we optimize antagonists collaborating with pharmaceutical companies, which have been confirmed experimentally.

2. Methods

We predicted the three dimensional structures of prostanoid receptors and urotensin receptors using MembStruk4.1 computational method summarized here and in Figure 1.

2.1 Prediction of the TM regions and hydrophobic centers: The TM regions were predicted using TM2ndS method described in reference[8]. We retrieved various sequences similar to the target sequence. Multiple sequence alignment of those sequences was performed using clustalW. Using the multiple sequence alignment as input, the TM regions were predicted using TM2ndS procedure.[8] The hydrophobic maximum was chosen as the central residue (referred to as the *centroid*) for each helix that divides the area under the hydrophobicity curve equally. The centroid for each helix is positioned to be in the same xy plane (the midpoint of the lipid).

2.2 Prediction of the 3D structure: Based on the predicted TM regions and the TM centroids, the MembStruk program was used to build and optimize the 3-D structure. The steps of MembStruk and the predicted structure are described below.

1. *Helix packing*: First, canonical α-helices were built for each TM domain. These α-helix structures were then bundled together as follows. The predicted helix centroid is placed on the xy plane using x,y coordinates based on the low resolution (7.5 Å) electron density map of frog rhodopsin. The orientation of each helix about its z axis (the χ angle) is chosen so that its helical face with the maximum hydrophobic moment points outwards to contact the lipid. In this analysis, we calculate the hydrophobic moment over the full helix but including only the half of the residues that would face outward. Then each helix is tilted about the point at which the central axis intersects the xy plane to match the tilt angles (θ,φ) from frog rhodopsin.

2. *Helix bending*: Next, molecular dynamics (MD) simulations were performed (200 picoseconds) for each individual helix, allowing the helix to attain its equilibrium structure (in some cases it bends or kinks). We found that 200ps is enough to equilibrate individual helix. Then we chose the structure with the lowest potential energy for each helix and assembled it back into the bundle so that the average axis coincides with the original axis. The side chains were then optimized using SCWRL[13, 14] and the total energy minimized (conjugate gradients).

3. *RotMin*. This initial packed structure was minimized and then we allowed the individual packing interactions to optimize as follows. Each helix was independently rotated (χ) by +5º and -5º, the side chains repositioned using SCWRL, and then all atoms of the bundle optimized. If either new angle was lower, it was selected.

4. *Lipid Insertion*: At this point we inserted the 7 helix bundle into a lipid framework ending up with 48 lipids molecules arranged as a bilayer. These lipid molecules were optimized using rigid body dynamics.

5. *RotScan*. Starting from the final RotMin structure, we performed a full 360-degree rotational scan (χ) on each of the helices in 5° increments. For each angle, the side chains were re-assigned with SCWRL and full bundle re-minimized. Multiple minima based on energy and interhelical hydrogen bonds were chosen for each helix. Combination of multiple minima for each helix leads to an ensemble of conformations which were then sorted by the number of interhelical hydrogen bonds and then by total energy.

2.3 Prediction of the Extra-Celluar (EC) and Intra-Cellular (IC) loop structure

We took the best structure from the previous step and added the three EC and IC loops. We expect the three EC and three IC loops of GPCRs to be quite flexible and strongly affected by the solvent, which is treated only implicitly in MembStruk. Thus to provide initial loop structures for our MD studies, we used the alignment of our predicted structure with bovine rhodopsin and then homology threaded the loops to the crystal structure of bovine rhodopsin (1L9H.pdb). Then we carried out minimization and dynamics on the loops with fixed helix bundle atoms.

In the crystal structure of bovine rhodopsin the ECII loop (connecting TM4 and TM5) is closed over the 7-TM barrel, contributing to the binding of 11-cis-retinal. This ECII loop has a disulfide bond to TM3 (C105-C183), which is highly conserved among the rhodopsin super family of GPCRs. Thus, we include this disulfide bond in our loop structures. It is generally believed that the disulfide bond plays critical role in the folding of 7 helices and in the closing of the ECII loop over the 7-TM barrel.[15] Since the rhodopsin in the crystal study is in the inactive form, it is possible that substantial changes occur in ECII and in other loops upon activation.

2.4 Molecular Dynamics Simulation

Since the description of lipid and water in MembStruk is implicit with a layer of lipid bilayer, we performed molecular dynamics (MD) simulations of the predicted structure with and without ligand for 1ns in explicit lipid bilayer and water. We carried out MD simulations using NAMD including explicit water and a periodically infinite lipid to determine the interactions of the protein with lipid and water.[12, 16]

We started with the predicted structure, stripped away the lipid molecules, and inserted it in a periodic structure of (POPC: 1-palmytoil-2-oleoyl-sn-glycero-3 - Phosphatidylcholine). In this process, we eliminated lipid molecules within 5 Å of the protein. Then we inserted this in a box of water molecules and eliminated waters within 5 Å of the lipid and protein. Then keeping the protein fixed we allowed the lipid and water to relax using minimization. Then we minimized the whole system before doing dynamics. MD simulation includes 100 lipid molecules, 6617 water molecules. There are ~ 33000 atoms per periodic cell.

2.5 HierDock Method: *Scan the entire receptor for binding sites*

We used HierDock approach to predict the binding mode of ligand to the receptor. The first step is to scan all void regions in the entire receptor structure to locate putative binding regions for the ligand. The void region in the entire receptor structure was partitioned into 27 regions and the HierDock method was used to dock the ligand in each box. Here we examined the best binding sites that have at least 80% buried surface area. Subsequently we docked the ligand in this putative binding region using the HierDock2.0 method.
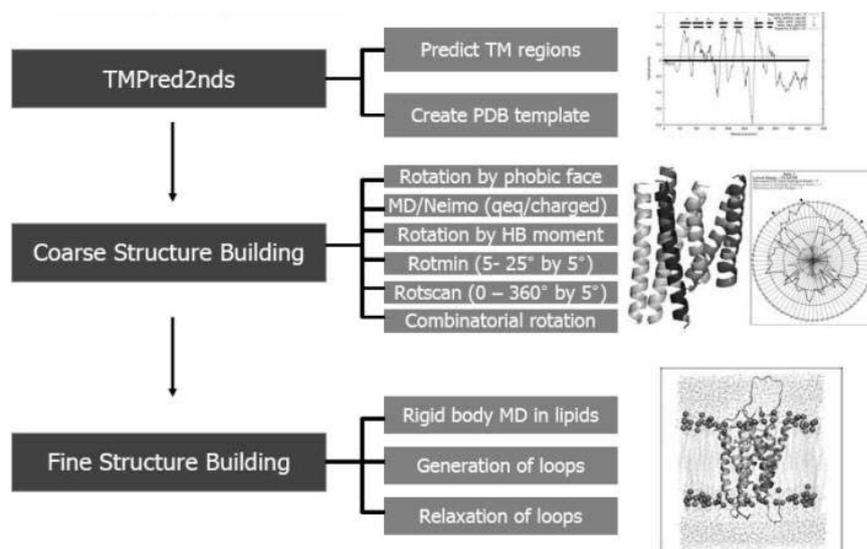
4

Figure 1. Whole MembStruk procedures contains 1)TMprediction, 2) Coarse structure building, and 3) fine structure building in explicit membrane and water

3.0 Results and discussion

3.1Predicted human DP structure and binding modes with PGD2 and antagonists.

The DP receptor of prostanoid receptors lacks some of the well-conserved motifs present in class A GPCRs. For example, the DRY motif on TM3 is ECW, the well-conserved Trp on TM4 becomes Leu, the WXP motif on TM6 becomes SXP and the NPXXY motif on TM7 is a DPWXF in the DP receptor. Thus, we can expect that the DP receptor might have a different set of stabilizing interhelical hydrogen bonds from rhodopsin. The predicted 3D structure of human apo-DP receptor is shown in Figure 2.

We find an interhelical hydrogen bond between N34(1) and D72(2). N34(1) and D72(2) are conserved in the rhodopsin family A including DP, but the conserved Asn of the NPXXY motif in TM7 is a DPWXF motif in the DP receptor. S316(7), which is not a conserved residue, makes a hydrogen bond with the N34(1) and D72(2). D319(7) makes a hydrogen bond with S119(3). D72(2) also forms a strong salt bridge with the K76(2) on the same helix. K76(2) is a conservative replacement in other prostaglandin receptors except for thromboxane receptors. We also find a hydrogen bond between R310(7) and Y87(2), where R310(7) is conserved across all prostaglandin receptors while Y87 is present only in DP receptors.
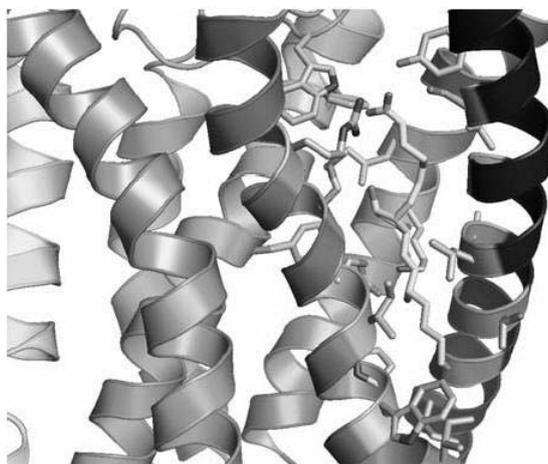
Figure 2. Predicted human DP receptor and binding mode with PGD2.

The predicted binding site of PGD2 is shown in Figure 2, where PGD2 is colored in yellow. PGD2 is located between the TM 1, 2, 3, 7 helices and is covered by the ECII loop.

We find favorable hydrophobic interactions of the α chain with L26(1) and F27(1). The α chain of PGD2 points up toward the EC region with the ω chain pointing down between TM1 and TM7. The critical elements of bonding are the carboxylic acid interacts with R310(7); the carbonyl on the cyclopentane ring of PGD2 has a hydrogen bond with K76(2); the hydroxyl on the ω chain interacts with S316(7) and K76(2); 9-OH forms a hydrogen bond with S313(7); a hydrophobic pocket surrounds the α chain with M22(1), G23(1), Y87(2), W182(ECII), L309(7), R310(7), L312(7), and S313(7) within 6 Å; a hydrophobic pocket surrounds the ω chain with L26(1), G30(1), I317(7), P320(7), and W321(7) within 6 Å.

The prediction that the carboxylic acid group of PGD2 interacts with R310(7) is confirmed strongly by various experiments. The carboxylic acid group and the hydroxyl group on the ω–chain are present in all the prostanoid compounds. R310(7) is 100% conserved among prostanoid receptor family and K76(2) is not. Other hydrophobic residues P320(7), W321(7) interacting with ω chain are also 100% conserved in prostanoid receptor family. Structure activity relationship studies of PGE2 show that the carboxylic acid group, both the ω chain itself and the hydroxyl group in the ω chain are critical for agonist activity.[17] The hydroxyl and carbonyl groups on the cyclopentane ring are *not* present in all prostanoid compounds and these groups offer receptor selectivity to the ligand as discussed next.

DP receptor binds to PGD2 and shows at least 2 orders of magnitude lower affinity to other prostanoid compounds. However the IP receptor binds to PGE1 and PGI analogs (iloprost), but it does not bind PGE2. Assuming that these other prostanoid compounds bind to the *h*DP receptor in similar binding mode as PGD2, we can explain the how the DP receptor prefers PGD2 to other prostanoid compounds like PGF2α, PGE2, and PGI2 as shown in Figure 3.

6

(a) PGD2

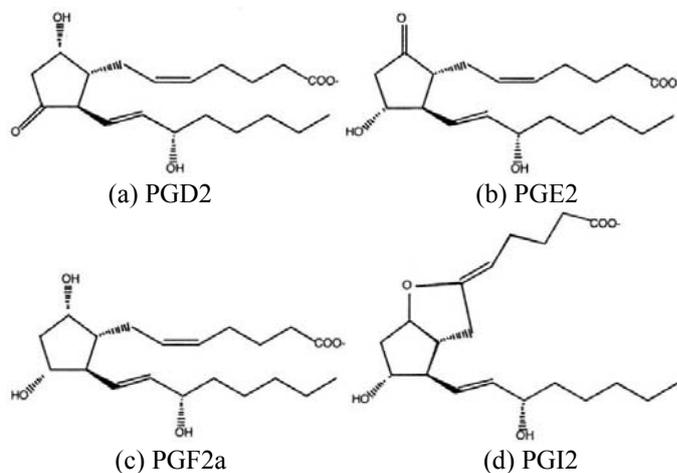(b) PGE2

(c) PGF2a

(d) PGI2

Figure 3. Prostanoid compounds

Based on our predicted structure of human DP, we collaborate with Sanofi-Aventis company to predict the binding of different antagonists and optimize them. One of the compounds is currently under pre-clinic trial.

Here we illustrate one example cyclopentanoindole (CPI) as shown in Figure 4. The binding mode of CPI in Figure 4b correlates very well with SAR data. CPI is predicted to be located among TM1237 region, which is similar to the endogenous ligand (PGD2) as discussed above, However, CPI with DP is shown to have different Molecular Dynamics (MD) motion as shown in PGD2 bound hDP (data not shown here). Those MD studies illustrate the difference of agonist bound receptor from antagonist bound receptor.
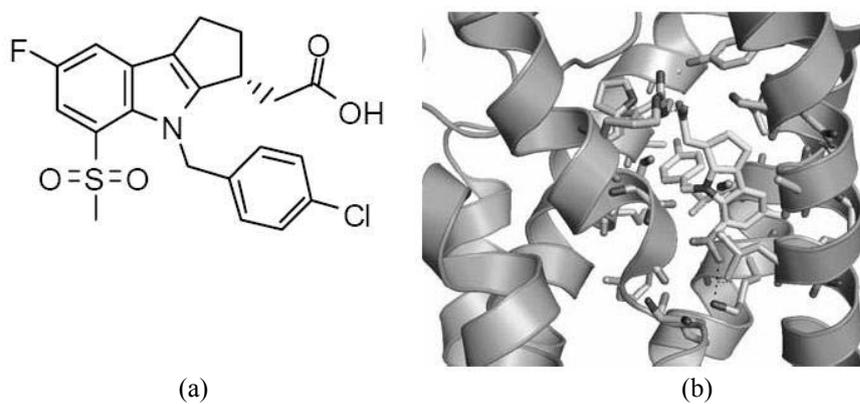
7

(a)   (b)

Figure 4. Cyclopentaoindole (CPI) compound and predicted binding mode

3.2Predicted human UT2R structure and binding mode with peptide CFWKYC

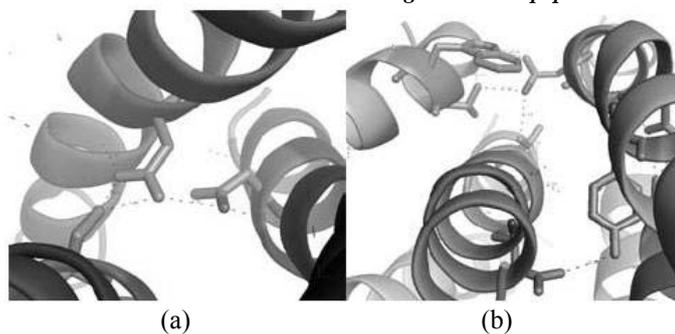(a)   (b)

Figure 5. Predicted human UT2R structure from MembStruk has conserved interhelical hydrogen bond network among (a) TM127; and (b) TM234.
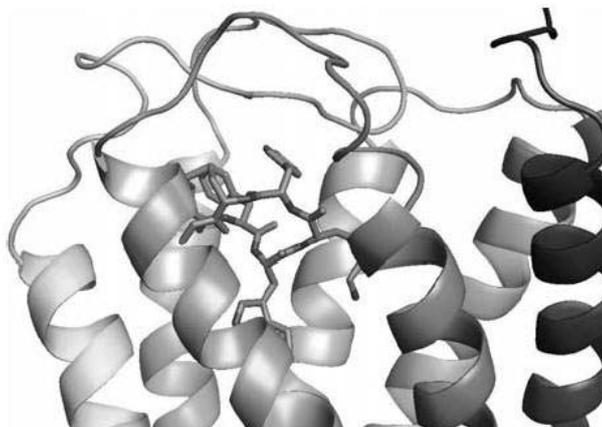
8

Figure 6. Predicted human UT2R structure and binding mode with agonist CFWKYC

We predict the 3D structure of human UT2R using Membstruk. As shown in Figure 5, the predicted structure has the strong interhelical interactions among TM127 (Figure 5a) and TM234 (Figure 5b) as in rhodopsin structure. Those important interhelical interactions have been recognized by MembStruk at RotScan step.

Urotensin II (UII), an urophysial peptide, is a cyclic dodecapeptide (AGTADCFWKYCV). The composition of UII ranges from 11 amino acids in humans to 14 amino acids in mice, always with the conserved cysteine-linked macrocycle CFWKYC, which is essential for the biological activity.[18] Indeed, CFWKYC itself has been shown to bind hU-II.[19] Thus we dock CFWKYC into predicted structure of hU-II using HierDock. The docked result is shown in Figure 6, where CFWKYC is located in TM3456 region and covered by ECII loop.

In the predicted binding mode, Lys in the middle of CFWKYC is forming salt bridge with D116(3). This is consistent with SAR data,[20] where K=>A will cause at least 1000 fold less active.

In the predicted binding mode, Tyr of CFWKYC is coupling with C109(3), F113(3), L184, C185, L186(EcII), which is a hydrophobic pocket. This explains that Y=>F keeps the peptide active.[20]

In the predicted binding mode, Trp of CFWKYC is located among F202(6), F257(7), F117(3), F113(3), an aromatic pocket. This is consistent with SAR data,[20] where W=>A or 2Nal, will dramatically lose the potency.

In the predicted binding mode, Phe of CFWKYC directly interacts with V170(4), which is consistent with photo-labeling experiment results.[21]

In summary, the binding conformation predicted with HierDock explains SAR data and photo-labeling experiment results. This validates both the MembStruk and HierDock protocols, suggesting that accuracy of ~ 3 Å for

9

structure prediction in the TM regions is adequate to identify binding site and structure.

Summary

We predict the structure and function of human DP receptor and human UT2R receptor by using Membstruk and Hierdock. Both receptors play important role in the human body, such as cardiovascular and immune systems.

The predicted structure and function for these two receptors are in good agreement with the experimental data currently available. The predicted binding position of PGD2 is located among TM127 region. It has important interactions with R310(7), S316(7), K76(2), and S313(7). These results suggest additional site-directed mutagenesis studies on these residues to test the predicted structure and function of GPCR.

The predicted binding mode of PGD2 provides the structure basis for understanding of the selectivity of prostanoid receptors. Based on our predicted structure of human DP, we collaborate with Sanofi-Aventis to predict the binding of different antagonists and optimize them. One of the compounds is currently under pre-clinic trial.

We also obtained high quality structure of human UT2R structure and binding structure with core peptide segment CFWKYC. The peptide is predicted to be located in TM3456 region covered by ECII and agrees with available SAR data.

1. Dong, X.; Han, S. K.; Zylka, M. J.; Simon, M. I.; Anderson, D. J., *Cell* **2001,** 106, 619.

2. Wilson, S.; Bergsma, D., *Pharmaceutical News* **2000,** 7, 3.

3. Narumiya, S.; Sugimoto, Y.; Ushikubi, F., prostanoid receptors: Structures, Properties, and Functions. *Physiological Reviews* **1999,** 79, (4), 1193.

4. Hall, S. E.; Floriano, W. B.; Vaidehi, N.; Goddard III, W. A., 3D Structures for mouse I7 and rat I7 olfactory receptors from theory and odor recognition profiles from theory and experiment. *Chem. Senses* **2004,** 29, 595.

5. Floriano, W. B.; Vaidehi, N.; Goddard, W. A., III, Making Sense Of Olfaction Through Predictions Of The 3d Structure And Function Of Olfactory Receptor. *Chemical Senses* **2004,** 29, 269.

6. Kalani, M. Y.; Vaidehi, N.; Hall, S. E.; Trabanino, R.; Freddolino, P.; Kalani, M. A.; Floriano, W. B.; Kam, V.; Goddard, W. A., III, Predicted 3D Structure Of The Human D2 Dopamine Receptor And The Binding Site And Binding Affinities For Agonists And Antagonists. *Proc. Natl. Acad. Sci.* **2004,** 101, 3815.

7. Freddolino, P.; Kalani, M. Y.; Vaidehi, N.; Floriano, W.; Hall, S. E.; Trabanino, R.; Kam, V. W. T.; Goddard, W. A., Predicted 3D structure for the human β2 adrenergic receptor and its binding site for agonists and antagonists. *Proc. Natl. Acad. Sci.* **2004,** 101, 2736.

10

8. Trabanino, R.; Hall, S. E.; Vaidehi, N.; Floriano, W.; Goddard, W. A., First Principles Prediction of the structure and Function of G protein-coupled receptors: validation for bovine rhodopsin. *Biophys. J.* **2004,** 86, 1904.

9. Hummel, P.; Vaidehi, N.; Floriano, W. B.; Hall, S. E.; Goddard, W. A., Test of the Binding Threshold Hypothesis for olfactory receptors: Explanation of the differential binding of ketones to the mouse and human orthologs of olfactory receptor. *Protein Science* **2005,** 14, 703.

10. Peng, J.; Vaidehi, N.; Hall, S.; Goddard III, W. A., The Predicted 3D Structures of the Human M1 Muscarinic Acetylcholine Receptor with Agonist or Antagonist Bound. *ChemMedChem* **2006,** 2006, in press.

11. Vaidehi, N.; Schlyer, S.; Trabanino, R.; Kochanny, M.; Abrol, R.; Koovakat, S.; Dunning, L.; Liang, M.; Sharma, S.; Fox, J. M.; Floriano, W. B.; Mendonça, F. L. d.; Pease, J. E.; III, W. A. G.; Horuk, R., Predictions of CCR1 chemokine receptor structure and BX 471 antagonist binding followed by experimental validation. *J Biol Chem* **2006,** 281, 27613.

12. Spijker, P.; Vaidehi, N.; Freddolino, P.; Hilbers, P.; Goddard, W., Dynamic behavior of fully solvated beta 2-adrenergic receptor, embedded in the membrane with bound agonist or antagonist. *Proc. Natl. Acad. Sci.* **2006,** 103, 4882.

13. Canutescu, A.; Shelenkov, A.; Dunbrack, R., A graph theory algorithm for protein side-chain prediction. *Protein Science* **2003,** 12, 2001.

14. Altschul, S.; al., e., Basic local alignment search tool. *Journal of Molecular Biology* **1990,** 215, 403.

15. Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C.; Motoshima, H.; Fox, B.; Trong, I.; Teller, D.; Okada, T.; Stenkamp, R.; Yamamoto, M.; Miyano, M., Crystal structure of rhodospin: A G Protein-Coupled Receptor. *science* **2000,** 289, 739.

16. Philips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Shulten, K., Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **2005,** 26, 1781.

17. Ungrin, M. K.; Carriere, M.-C.; Denis, D.; Lamontagne, S.; Sawyer, N.; Stocco, R.; Tremblay, N.; Metters, K. M.; Abramovitz, M., *Molecular Pharmacology* **2001,** 59, 1446.

18. Davenport, A. P.; Maguire, J. J., *Trends Pharmacol. Sci.* **2000,** 21, 80.

19. Foister, S.; Taylor, L. L.; Feng, J.-J.; Chen, W.-L.; Lin, A.; Cheng, F.-C.; Smith, A. B.; Hirschmann, R., *Organic Letters* **2006,** 8, 1799.

20. Kinney, W. A.; Almond, H. R.; Qi, J.; Smith, C. E.; Santulli, R. J.; de Garavilla, L.; Andrade-Gordon, P.; Cho, D. S.; Everson, A. M.; Feinstein, M. A.; Leung, P. A.; Maryanoff, B. E., *Angew. Chem. Int. Ed.* **2002,** 41, 2940.

21. Boucard, A. A.; Sauve, S. S.; Guillemette, G.; Escher, E.; Leduc, R., *BioChem. J.* **2003,** 370, 829.