

KO-18 KP-32

GPCR structure modeling and ligand docking: GPCRDock2008

Kazuhiko Kanou^{*1}, Daisuke Takaya², Genki Terashi¹,
Mayuko Takeda-Shitaka^{1,2} and Hideaki Umeyama^{1,2}

1) School of Pharmacy, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan

*2) RIKEN Systems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama
230-0045, Japan*

Introduction

Molecular modeling has an important role in rational drug design. Reliable three-dimensional (3D) models can provide valuable insights into basic principles of molecular recognition and aid in structure-based approaches to lead discovery and optimization. G protein-coupled receptors (GPCRs) are membrane proteins involved in signal transduction pathways and are important therapeutic targets for numerous diseases. As such, significant structure prediction efforts using methods ranging from de novo to homology-based approaches have been applied to members of the GPCR family. Until recently, most GPCR homology modeling efforts have been based on the templates of bovine rhodopsin and bacteriorhodopsin, with refinement of the models achieved through molecular dynamics simulations, ligand docking and incorporation of additional biochemical and biophysical data. The refinement step is necessary in building accurate models especially around the ligand-binding site, owing to the expected structural differences among members of the family. These differences result from the generally low sequence identity and the large diversity of ligands accommodated within the family, and from the various conformational states that are associated with different levels of ligand efficacy. The most recently solved GPCR structure is the 2.6 Å crystal structure of the human adenosine A2A receptor bound to an antagonist. Adenosine receptors belong to the class A rhodopsin-like GPCR family and represent promising therapeutic targets in a wide range of conditions, including cerebral and cardiac ischemic disease, sleep disorders, immune and inflammatory disorders, and cancer. To evaluate current progress in GPCR structure prediction and the docking of potential ligands, a community-wide, blind prediction assessment — GPCR Dock 2008[1] — was held in August 2008, and we participated. GPCR Dock 2008 was organized in a similar manner to the previous CASP (Critical Assessment of methods for Protein Structure) and CAPRI (Critical Assessment of PRediction of Interactions) studies. In August 2008, before the publication of the human adenosine A2A receptor structure in October 2008 and the public release of the three-dimensional coordinates, participants were asked to predict and submit up to ten ranked models of the human A2A receptor in complex with the ligand ZM241385, starting from the amino acid sequence of the receptor and a two-dimensional structure of the ligand. After publication of the crystal structure of the human adenosine A2A receptor, the organizers assessed the models constructed by GPCR Dock 2008 participants. We describe our modeling/docking algorithm and the results for the GPCR Dock 2008.

Methods

1. Modeling of the human adenosine A2A receptor

The X-ray structure of the turkey β_1 adrenergic receptor with bound cyanopindolol (PDB code: 2VT4) was used as a template structure for homology modeling. The sequence alignment between target protein and the template was generated by our in-house programs FAMSD [2,3], and SKE-CHIMERA[4]. Sequence identity was 32.6%. Using FAMS Ligand&Complex[5], we constructed ten 3-dimensional (3D) structures of the target protein with two ligand molecules, cyanopindolol and (2S)-1-(9H-Carbazol-4-yloxy)-3-(isopropylamino)-propan-2-ol, obtained from PDB code: 2VT4 and PDB code: 2RH1, respectively. The coordinates of the latter ligand were obtained by superimposing 2RH1 onto 2VT4 based on structural fitting between two 3D structures.

2. Algorithm for ligand docking

We generated the initial coordinates of the target ligand so that the coordinates obtained by the docking would interact with the significant amino acid residues, ASN 181, ASN 253 and TYR 271. TYR 271 seemed to form a hydrogen-bond from the location of the binding site, referring to 2VT4 and 2RH1. ASN 181 and ASN 253 seemed to pinch a functional group, based on the positions of the two amino acid residues. Indeed, searches for the proteins that bind the ligand with adenine-like rings in the PDB database showed that the two asparagine residues can pinch various adenine-like rings. Next, we executed the docking calculations, including the optimization process for protein-ligand interaction, using our original potential function, which is called GENIUS. This potential function includes the potential energy of hydrogen bonds, hydrophobic interactions, stacking, collisions, internal energy and protein-ligand interaction information such as atom-atom distances that seem functionally important. The docking process was performed for each of the protein models. In this way we obtained many ligand docked models.

3. Criteria for prediction analysis and ranking

We performed clustering of docked ligand conformations based on ligand RMSD. The biggest cluster was selected, and the top ten conformations were determined from the score values.

Results and Discussion

GPCR Dock 2008 organizers defined the numerical measure of accuracy for the ligand binding mode, which was based on two metrics, ligand RMSD and the number of correct receptor-ligand contacts. The ligand RMSD between model and crystal structure is calculated as the coordinate root

kanouk@pharm.kitasato-u.ac.jp

mean-square deviation (RMSD) for the 25 non-hydrogen atoms of ZM241385 after superimposing the protein C α atoms of the model and the crystal structure. In addition, the ligand RMSD is also calculated excluding the phenoxy group that has high B-factor values. The number of correct contacts is counted as the number of correctly predicted native contacts observed between protein atoms and the ligand. A native contact was defined as any inter-atomic distance within 4 Å of the ligand in the crystal structure. Neither metric alone was sufficient to capture the accuracy of prediction around the ligand-binding site; hence, both were used and combined into a Z-score scoring scheme to rank the models. The models were ranked by assigning a combined Z-score to each model as shown in Table 1. The ligand RMSD, the number of correct contacts and the combined Z-score of our model ('kanou' in Table 1) were 5.5Å, 8 and 0.91, respectively, and our model was ranked 7th in relation to the combined Z-score.

Our model captured the hydrogen bonding interaction between the N253^{6.55}, the superscript of which means the Ballesteros numbering, side-chain and the exocyclic amino group (N15 atom) of the ligand, however the ligand makes no interaction with residues in ECL2 (extracellular loop 2). The inaccuracy in the ligand position was most likely due to errors in the side-chain positions of the ligand-binding residue, F168^{5.29}, in ECL2 as shown in Fig. 1.

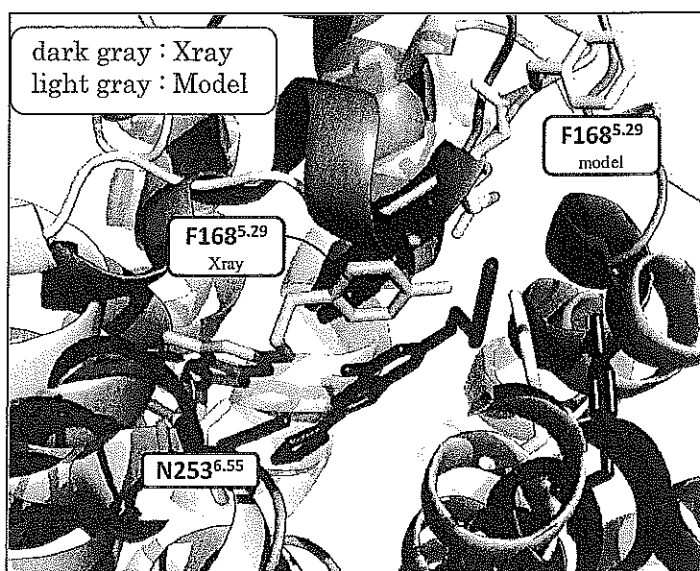


Fig. 1 Overlaid structure between native structure and our model. The superscript of the residue number means the Ballesteros numbering. Residues in the TM domain are assigned two numbers, N1 and N2. N1 is TM number. N2 is a number relative to the most conserved residue in the TM, which is assigned 50. N2 decreases towards N-terminus and increases towards C-terminus.

Conclusion

We had participated in the GPCR Dock 2008, using our original methods such as FAMSD, CKE-CHIMERA and GENIUS. As the result, our model ranked 7th in the assessment of the combined Z-score. The GPCR Dock 2008 organizers reported in reference (1) that our model was slightly less accurate but show similar trends to the top six models in their ability to accurately predict the ligand binding mode. The ligand is modeled in a native-like extended conformation and the hydrogen bonding interaction between the N253^{6.55} side-chain and the ligand was modeled accurately; whereas, the aromatic stacking interaction between the F168^{5.29} side-chain and the bicyclic ring of the ligand was modeled inaccurately due to errors in the side-chain positions in ECL2. Thus, further improvement for accurate modeling of the structurally divergent region such as the extracellular loops is required. Furthermore, the prediction of the helical shifts in the TM region, the prediction of the ligand binding residues and the ranking of docked models remain the difficult problems.

Table 1. Summary of results for the best models from the top ranking groups.

Group Name	Ligand RMSD (Å)	Ligand RMSD w/o phenoxy group (Å)	Number of Correct Contacts	Binding site residues RMSD (Å)	Protein C α RMSD (Å)	TM I-VII C α RMSD (Å)	ECL2 C α RMSD (Å)	Combined z-score
Costanzi	2.8	2.7	34	3.4	3.0 (266)	2.5 (212)	3.8 (8)	3.02
Katritch / Abagyan	6.2	4	40	3.5	4.0 (283)	2.7 (214)	8.9 (23)	2.76
Lam / Abagyan	5.7	3.6	33	3.3	4.1 (283)	3.6 (214)	7.3 (23)	2.42
Davis / Barth / Baker	5.8	5.4	18	4	3.5 (283)	2.1 (214)	8.4 (23)	1.46
Maigret	2.6	2.1	5	7.3	5.1 (283)	4.1 (214)	9.1 (23)	1.23
Jurkowski / Elofsson	5.3	5.2	10	3.9	6.2 (283)	2.9 (214)	12.7 (23)	1.04
Kanou	5.4	5.5	8	6.9	3.5 (279)	2.8 (214)	7.1 (23)	0.91
Goddard	5	3.9	5	4.8	4.3 (284)	2.5 (214)	10.7 (23)	0.78
Bologa	6.7	2.8	9	3.9	3.4 (278)	2.5 (213)	7.2 (19)	0.72
Olson	4.8	4.7	3	5.8	3.5 (284)	2.3 (214)	7.5 (23)	0.69

The RMSD values are calculated for the heavy atoms of the ligand ZM241385 (all 25 atoms, and partially without the phenoxy group), heavy atoms of the binding site residues (F168^{5,29}, E169^{5,30}, M177^{5,38}, W246^{6,48}, L249^{6,51}, H250^{6,52}, N253^{6,55}, H264^{6,66}, M270^{7,35}), C α atoms of all residues, C α atoms of residues in the TM helices I to VII (helix I: 6-34; helix II: 40-67; helix III: 73-107; helix IV: 117-142; helix V: 173-205; helix VI: 222-258; helix VII: 266-291), and C α atoms of residues in ECL2 (143-172 excluding 149-155 that are missing in the crystal structure). All RMSD values are obtained after the models are superimposed to the crystal structure using the protein C α atoms in PyMOL (version 1.0r2, www.pymol.org). The assignment of residues in the ligand-binding site and the secondary structure elements is from the PDB header section (PDB ID: 3EML). The number of residues used in the RMSD calculation is in parenthesis. The combined z-score value for the best model among submitted models from each participant is shown. Kanou of Group Name indicates our results.

References

1. Michino M., Abola E., GPCR Dock 2008 participants*, Brooks C. L. 3rd, Dixon J. S., Moulton J., Stevens R. C., *Nat. Rev. Drug. Discov.*, **8(6)**, 455–463 (2009). *Kanou K., Takaya D., Terashi G., Takeda-Shitaka M., Umeyama H et al.
2. Kanou K., Iwadata M., Hirata T., Terashi G., Umeyama H., Takeda-Shitaka M., FAMSD: The combination of alignment methods, homology modeling, 3D structure estimation and molecular dynamics to develop a powerful protein modeling approach., *Chem. Pharm. Bull. (Tokyo)*, accepted.
3. Ogata K., Umeyama H., *J. Mol. Graph. Model.*, **18(3)**, 258–272, 305–306 (2000).
4. Takeda-Shitaka M., Terashi G., Takaya D., Kanou K., Iwadata M., Umeyama H., *Proteins*, **61 Suppl 7**, 122–127 (2005).
5. Takeda-Shitaka M., Terashi G., Chiba C., Takaya D., Umeyama H., *Med. Chem.*, **2**, 191–201 (2006).