

## CHARACTERIZATION OF G PROTEIN COUPLED RECEPTORS THROUGH THE USE OF BIO- AND CHEMO- INFORMATICS TOOLS

TRAN PHUOC DUY

*Faculty of Applied Science, Hochiminh city University of Technology, Vietnam National  
University of Hochiminh*

ALEJANDRO GIORGETTI, NGUYEN HA HUNG CHUONG, PAOLO CARLONI  
*Computational Biophysics, German Research School for Simulation Sciences,  
Forschungszentrum Juelich, 52425 Juelich, Germany*

HOANG ZUNG

*Department of Science and Technology, Vietnam National University of Hochiminh*

**Abstract.** *G-protein-coupled receptors (GPCRs) are the largest membrane-bound receptor family expressed by mammals (encompassing more than 1 percent of the genome). They are involved in an enormous variety of intra- and extracellular signaling, including detection of light, sense of smell, neurotransmission, inflammation, and cardiac and smooth muscle contractility [Kroeze et al., J Cell Sci 116,4867 (2003); Sakmar et al., Curr Opin Cell Biol 14, 189 (2002)]. Ligand (or photon) binding to GPCRs activates a cascade of events, producing an electrical signal as output. They are of utmost pharmaceutical relevance, being the targets of almost 30 percent of all marketed drugs [Landry and Gies, Fundam Clin Pharmacol 22, 1 (2008)]. The aim of this study is to perform a systematic and detailed analysis of sequence-structure relationships of known GPCR structures. A web-server for the automatic modeling and ligand docking of GPCRs at different activation states has been developed.*

### I. INTRODUCTION

G Protein Coupled Receptors (GPCRs) are the seven transmembrane receptors, which have seven transmembrane  $\alpha$ -helices connected by alternating intracellular and extracellular loops. GPCRs involved in an enormous variety of intra- and extracellular signaling [1, 2]. Nowadays, there are limited number of known crystal structure of GPCRs due to the facts that most GPCRs are expressed at low levels in native tissues, and overcoming the protein-stability problems [3]. On the other aspect, the tertiary structures of proteins have found to be better conserved during evolution than the amino acid sequences. As a result of that, homology modeling is used to predict the three-dimensional (3D) structure of unknown 3D-structure proteins using their amino acid sequences, and based on their homology to the known 3D-structure proteins [4, 5, 6]. The choice of which experimental GPCR structures to use for building a comparative model of a particular GPCR is unclear and without detailed structural and sequence analyses, could be arbitrary. Recently, Worth et al. have analysed in detail conserved and unique sequence motifs and structural features in experimentally-determined GPCR structures [7]. Deeper

insight into specific and important structural features of GPCRs as well as valuable information for template selection has been gained. They have also formulated a workflow for identifying the most appropriate template(s) for building homology models of GPCRs of unknown structure.

More than half GPCRs (about 900) are olfactory receptors (ORs) [8, 9]. This clearly underlies the crucial role of the sense of smell during evolution. ORs possess high affinity for thousands of volatile molecules associated with odour. With such a large number of different ORs, the olfactory system is capable to discriminate between 10,000 different odours: one odorant can activate numerous types of ORs, while a single OR can be activated by several different odorants [10]. Odorant receptors (ORs) in the olfactory sensory neurons of the nasal epithelium translate odorants into neural signals. Each OR is thought to be specialized to recognize physicochemical features, such as functional groups or molecular size, of odorant molecules; these features are then translated into a neural signal that in turn leads to an olfactory perception. The physicochemical features of an odorant, therefore, are a key determinant of the olfactory percept. The rules governing the translation of molecule into percept, however, remain largely unknown. To develop a theory that predicts olfactory percept from molecular structure, we must first be able to predict OR activation from molecular structure. The overall sequences of mammalian ORs are diverse, with amino acid sequence similarity between different ORs ranging from less than 40%. Deciphering olfactory encoding requires not only a thorough description of the ligands that activate each odorant receptor (OR) but also a description of the putative binding cavities of the latter. In mammalian systems, however, ligands are known for fewer than 90 of over 1400 human and mouse ORs, greatly limiting our understanding of olfactory coding. In 2009, Saito and coworkers, have performed high-throughput screening of 93 odorants against 464 ORs expressed in heterologous cells and identified agonists for 52 mouse and 10 human ORs [11]. They used the resulting interaction profiles to develop a predictive model relating physicochemical odorant properties, OR sequences, and their interactions. Recently we have developed a database, i.e. OlfactionDB [12], that includes all the known receptor/ligand pairs found in literature. The latter will be used as the starting point for the classification.

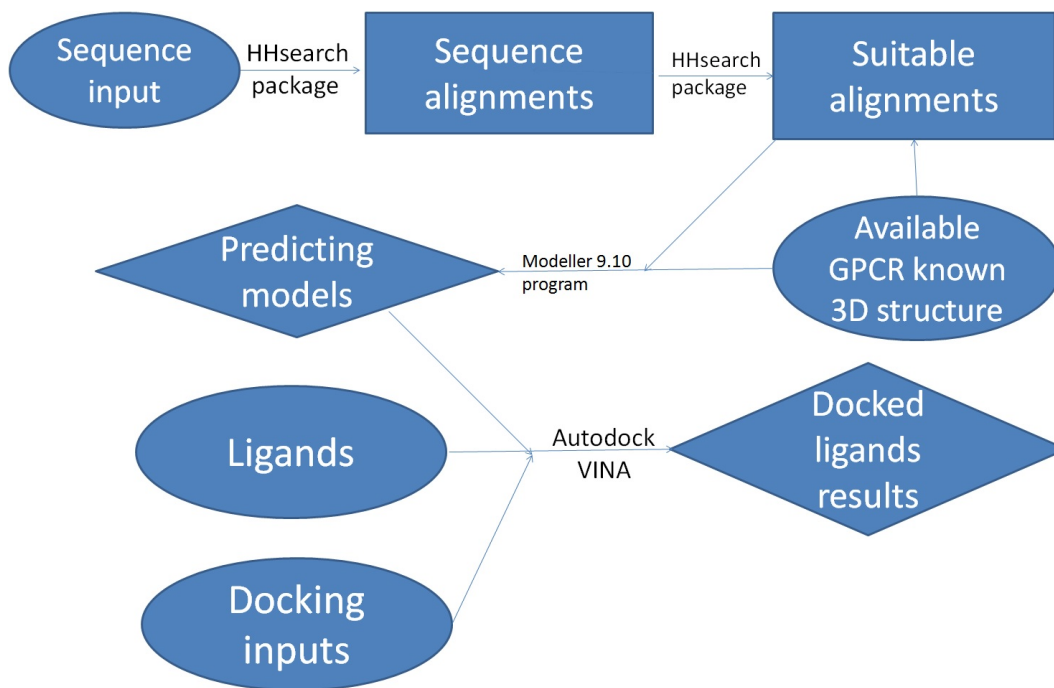
In this paper, we report the workflow of homology modeling and the built web-server automatically modeling of GPCRs at different activation states automatic modeling of GPCRs at different activation states, and docking odorant ligands.

## II. WEB-BASED SERVER FOR AUTOMATICALLY HOMOMOLOGY MODELING AND LIGAND DOCKING

Homology modeling is process consisting of 4 main steps: template search, target-template sequence alignment, model construction and model quality assessment respectively [13]. The server will play the role doing the first three steps. The template search and target-template sequence alignment steps often come together due to the common sequence alignment of these two steps. It has been shown that profile-profile alignment methods include the most sensitive and accurate alignment protocols to date and are the method of choice for identifying and aligning templates for homology modeling [14]. These

methods are used by many of the best protein structure prediction servers in CASP competition [15]. We choose the HHsearch package [16] from HHpred for performing these steps in our program. Note that after doing the template search and target-template sequence alignment, only the GPCR templates which have the experimental structures listed in the BLANCO databases [17] are used to perform the next steps.

Three major classes of model generation methods have been proposed: (i) Modeling by assembly of rigid bodies; (ii) Modeling by segment matching or coordinate reconstruction; (iii) Modeling by satisfaction of spatial restraints. The form of these restraints was obtained from a statistical analysis of the relationships between many pairs of homologous structures. These relationships were expressed as conditional probability density functions and can be used directly as spatial restraints. An important feature of the method is that the spatial restraints are obtained empirically, from a database of protein structure alignments. Next, the spatial restraints and CHARMM energy terms enforcing proper stereochemistry [18] are combined into an objective function. Finally, the model is obtained by optimizing the objective function in Cartesian space. One of the commonly used program for this step is MODELLER [19].



**Fig. 1.** Workflow of the modeling server

After built the models, one can used the built-in Autodock VINA [20] to dock the ligands. AutoDock Vina is a new open-source program for drug discovery, molecular docking and virtual screening, offering multi-core capability, high performance and enhanced accuracy and ease of use. The server automatically docked the ligands in the calculated binding

cavity which is derived from experimental structure of the templates. The procedure of the webserver is described in the Fig. 1.

### III. APPLICATION FOR HOMOLOGY MODELING AND ODOR LIGAND DOCKING OF THE MOUSE ODOR RECEPTOR MOR174-9

The mouse odor receptor sequence is obtained from the UNIPROT databases (accession code: Q920P2). After inserting the sequence as input, the built server choose 11 GPCR templates as listed in Fig. 2. The sequence conservation between MOR174-9 and the template sequences is low liked that found in [21]. The conservative positions of these protein are mostly located in 8 helices. These receptor residues, among them key GPCR signature residues, are conserved mainly to maintain a fine-tunable receptor-signaling network for the existance in two conformational states: signaling active and inactive [21]. Fig. 3 lists the results of template search and sequence alignment step carried out by HHsearch package. One can see that the best E-value and P-value (the lowest one) is Human A2A Adenosine Receptor (PDB code 3eml). Although 3eml has the highest E-value and P-value, but its sequence identity is lower than 3uon (the human M2 muscarinic acetylcholine receptor).

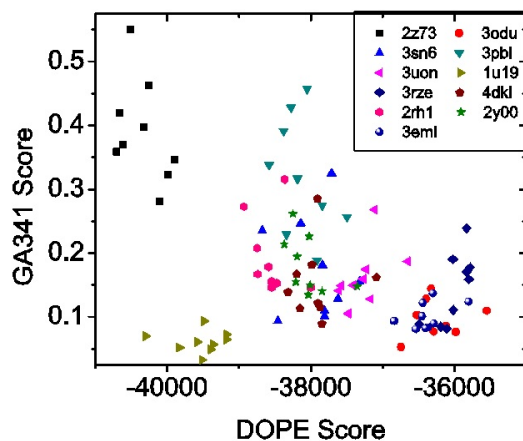


**Fig. 2.** GPCR templates and olfactory receptor multiple sequence alignment. Blue columns mark for the highly conservative amino acids

These sequence alignments are used to build the 3D models of MOR174-9 using MODELLER by the server. Fig 3 shows the DOPE scores vs the GA341 scores of the built models. DOPE score is Discrete Optimized Protein Energy, which means the lowest it is, the better model is. The GA341 scores always range from 0.0 (worst) to 1.0 (native-like).

No	Hit	Prob	E-value	P-value	Score	SS	Cols	Query HHM	Template HHM
1	3eml_A	100	1.20E-34	4.40E-39	228.1	10.6	301	5-313	1-472
2	3uon_A	100	6.20E-33	2.30E-37	217.7	13.1	299	3-313	1-463
3	3v2y_A	100	7.90E-33	2.90E-37	217.1	9.9	294	1-313	13-498
4	3rze_A	100	5.90E-32	2.20E-36	211.9	11.4	282	18-313	1-448
5	2rh1_A	100	4.90E-32	1.80E-36	212.3	10.8	302	1-313	8-482
6	2ks9_A	100	8.50E-31	3.10E-35	204.9	11.5	300	3-313	1-327
7	3sn6_R	99.9	8.00E-32	2.90E-36	211.1	1.2	304	1-313	147-496
8	3odu_A	99.9	4.40E-31	1.60E-35	206.6	3.3	299	2-313	1-498
9	2y00_A	99.9	2.20E-30	8.20E-35	202.4	5.7	290	15-313	1-306
10	2ydv_A	99.9	2.20E-29	8.20E-34	196.4	8.6	286	18-313	1-310
11	2z73_A	99.9	7.90E-29	2.90E-33	193	9.8	301	1-313	1-337
12	3pbl_A	99.9	1.10E-28	4.20E-33	192.1	8.7	291	1-313	16-477
13	4dkl_A	99.9	1.80E-28	6.70E-33	190.9	6.6	295	4-313	1-461
14	1u19_A	99.9	6.00E-28	2.20E-32	187.7	7	304	2-313	1-329

**Fig. 3.** Alignment results by the HHsearch package. Hit denotes the PDB code and chain of the template. Prob denotes for the Probability. SS denotes for Sequence Similarity. HHM denotes for Hidden Markov Model



**Fig. 4.** GA341 score vs DOPE score of built models based on chosen templates

Note that GA341 score need to be larger than 0.6. Due to the low sequence identities, the GA341 of the models is lower than 0.6. This suggests that they need to be refined. Overall, it is suggested the models built by template 2z73 is the best models.

Next step, we choose from the database OlfactionDB 8 ligands with PubChem id: 62465 (EC60 > 300 $\mu$ mol), 7144 (EC60 > 300 $\mu$ mol), 8467 (EC60=290 $\mu$ mol), 3314 (EC60=46 $\mu$ mol), 7136 (EC60 > 300 $\mu$ mol), 1549045 (EC60 = 66 $\mu$ mol), 637796 (EC60 = 742 $\mu$ mol) and 1183 (EC60 = 36 $\mu$ mol). Using Autodock VINA, we dock all these ligands to the built models. Fig. 5 show the best docked configurations of the ligands. The best complexes mostly

Ligand name	Receptor's built by template:	VINA Affinity (kcal/mol)
1183	3uon	-7
3314	3rze	-7.6
7136	3rze	-8.2
7144	3rze	-7.2
8467	3uon	-7.3
62465	3rze	-7.4
637796	3rze	-8.6
1549045	3rze	-8.1

**Fig. 5.** The lowest Autodock VINA affinity of docked ligands

generated by the receptor built by template 3rze, although their DOPE and GA341 scores are not the best. All the ligand are well located in their binding pocket. For more information, the complexes should be checked by calculating the binding affinity using the steered Molecular dynamics or the root mean squared displacements of residues.

#### IV. CONCLUSIONS

We have built the application for characterization of GPCRs through the use of bio- and chemo- informatics tools. The server will play the role for doing homology modeling massively for the unknown 3D GPCRs based on the experimentally known 3D GPCRs structure. This will help to solve the problem of choosing the right GPCR templates due to their limited known 3D structures. Through the docking features added in the server, the predictive models of odorant receptors can easily to be built.

#### ACKNOWLEDGMENT

Financial support of National Foundation for Science and Technology Development (NAFOSTED, project No DFG.2011.01) is gratefully acknowledged.

#### REFERENCES

- [1] Kroeze et al., *J Cell Sci* **116** (2003) 4867.
- [2] Sakmar et al., *Curr Opin Cell Biol* **14** (2002) 189 .
- [3] Rosenbaum et al., *Nature* **459** (2009) 356
- [4] C Chothia, A M Lesk, *The embo Journal* **5** (1986) 823.
- [5] A Tramontano, A M Lesk, *Protein Structure Prediction: Concepts and Applications*, 1st edit., 2006 Wiley-VCH.
- [6] D Frishman, A Valencia, *Modern Genome Annotation: The Biosapiens Network*, 1st edit., 2008 Springer.
- [7] C L Worth, G Kleinau, G Krause, *PLoS One* **4** (2009) e7011.

- [8] L Buck, R Axel, *Cell* **65** (1991) 175.
- [9] S Takeda et al., *FEBS Lett.* **520** (2002) 97.
- [10] B Malnic et al., *Cell* **96** (1999) 713.
- [11] Saito et al., *Sci. Signal* **2** (2009) ra9.
- [12] <http://molsim.sci.univr.it/OlfactionDB>
- [13] L Bordoli et al., *Nature Protocol* **4** (2008) 1 .
- [14] M Remmert et al., *Nature Methods* **9** (2012) 173 .
- [15] CASP stands for the Critical Assessment of protein Structure Prediction, an experiment for protein structure prediction taking place every two years since 1994 (<http://predictioncenter.org/>)
- [16] J Soeding, *Bioinformatics* **21** (2005) 951
- [17] <https://blanco.biomol.uci.edu/mpstruc/listAll/list>
- [18] A D, MacKerell et al., *J. Phys. Chem. B* **102** (1998) 3586
- [19] M.A. Marti-Renom et al, *Annu. Rev. Biophys. Biomol. Struct.* **29** (2000) 291
- [20] O. Trott, *J. Comp. Chem.* **31** (2010) 455
- [21] K M Schlinkmann et al., *Prod. Nat. Acad. Sci.* **109** (2012) 9810

*Received 10-09-2012.*