ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
http://web.icf.ro/rrch/

Rev. Roum. Chim.,
2014, *59*(3-4), 193-199

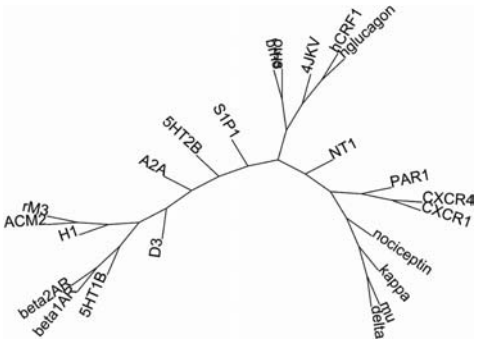
TEMPLATE STRUCTURE SELECTION FOR COMPARATIVE MODELING OF ORPHAN G-PROTEIN COUPLED RECEPTORS

Liliana OSTOPOVICI-HALIP*

Roumanian Academy – Institute of Chemistry, 24 Mihai Viteazu, 300223-Timişoara, Roumania

Received September 12, 2013

The G-protein coupled receptors (GPCRs) are the largest family of membrane receptors having major contributions in the transduction and regulation of signaling pathways in cardiovascular, respiratory, gastrointestinal, immune and central nervous systems. GPCRs are targets for almost half of the marketed drugs, and the increasing number of experimentally solved structures for diverse GPCRs opens new paths for the structure-based drug design in this family. However, for orphan GPCRs relevant structural data is almost impossible to be determined experimentally in the absence of a fully characterized biological profile of the receptor and therefore the comparative modeling techniques represent the only alternative to generate a three dimensional structure. Here we present an inventory of the most appropriate template structures for orphan GPCRs, resulted from a systematic investigation of a set of 128 GPCR sequences by using a combination of bioinformatics and molecular modeling approaches.



INTRODUCTION

Comparative modeling or homology modeling is one the most frequent computational method used to build a three dimensional (3D) structure of a protein when no experimental data is available. The method relies on the observation that between homologue proteins 3D structure is more conserved than the sequence. Therefore the structure's generation is done based on a sequence alignment between the protein of interest (target) and the template structure which represents the three-dimensional model for the target protein and which is determined by X-ray crystallography or nuclear magnetic resonance (NMR). As well, for accuracy and reliability of the final model, a certain sequence similarity (preferable over 40%) must exist between the target and template protein. A homology model is very helpful for the structure-based drug design process, since it can guide the future experiments (*e.g.* mutagenesis analysis) or it can provide valuable information to help the elucidation of an activation mechanism or the understanding of structure-function relationships.

G-protein coupled receptors (GPCRs) are the largest family of membrane receptors with relevance in the transduction and regulation of signaling pathways in cardiovascular, respiratory, gastrointestinal, immune and central nervous systems. The structural information of GPCRs was unknown until 2000[1] when crystal structure of bovine rhodopsin was experimentally elucidated. Despite the atypical profile for drug design process (the ligand is bound covalently and the receptor activation is triggered under the action of the light photons[1]), rhodopsin was intensively used for modeling of other GPCR members with relevance for the drug design process. In 2007, thanks to the innovative procedures accomplished in the protein expression and purification area[2] and also to the new developments brought in the X-ray crystallography, the crystal structure of a first

* Corresponding author: lili.ostopovici@acad-icht.tm.edu.ro

GPCR with pharmacological relevance became available.[3] Since then, the number of solved GPCR structures has been continuously growing, and in the present structures for twenty three GPCRs[1, 3-23] were deposited in the Protein Data Bank,[24] from which two belong to class B and one to class F GPCR. Even though the number of solved GPCR structures is expected to grow continuously in the next years, there are around 100 GPCRs, so-called orphans, for which the knowledge of structural information is still unknown since these receptors do not have a known natural ligand or biological function. However, the availability of a 3D structure for an orphan – even though it is a theoretical model – will help the discovery of active compounds and the full characterization of their pharmacological importance and relevance. Consequently, homology modeling of orphan GPCRs is the only technique (at least for the moment) which would allow conducting the structure-based experiments, and to analyze and study ligand-receptor interactions and mutant proposals.

In the present work we carried out a strategy to create an inventory of the most appropriate template structures for orphan GPCRs, developed based on systematic investigation and analysis of a set of 105 human orphan GPCRs and 23 sequences of GPCRs with known 3D structure.[1, 3-23]

METHODS

Sequence alignment

The 128 GPCR sequences were first downloaded from Uniprot database[25] and then automatically aligned following a multistep alignment procedure using the T-Coffee package.[26] First the GPCR sequences having a known 3D structure were aligned according to their class affiliation (A or B) and the resulted alignment were further refined manually in order to preserve the highly conserved residues specific for each class.[27] Next, based on the structural information available, the alignments of class A and B GPCR were merged together with the sequence of human smoothen homologue (class F) to create a intermediate alignment (IA) which was further used to guide the combination of the final alignments of each class containing also orphan GPCRs. Orphans were aligned separately with the final alignment of the class they belong. Class C GPCR for which no crystal structure is available yet has been aligned with all the high-resolution protein structures in order to estimate the most appropriate template structure. In the final step the alignments specific for each class were merged together according to the IA alignment.

Criteria for selecting the best template structure

The criteria used to select the most suitable template structure refer to the sequence similarity and the smallest distance from the sequences space between each pair of orphan and GPCR with available 3D data. For class A orphans, the binding site similarity was considered and was weighted to represent a quarter from the final sequence similarity. Binding site similarity was estimated by calculating the sequence similarity of fragments containing the amino acids found at 5Å around each crystallized ligand.

Distance calculations

The minimum distances from the sequences space between each pair of orphan and GPCR with available 3D structure were identified by calculating a distance matrix (M) using the Neighborhood Joining algorithm.[28]

Phylogenetic trees were generated using iTOL software.[29,30]

RESULTS AND DISCUSSIONS

The sequencing of human genome has identified around 400 GPCRs with relevance in pharmacology but out of these more than 100 receptors do not have a known natural ligand or function. These receptors are so-called orphans and finding their closest homologues it is the first step to gain new insights about their biological significance and to impart new directions for the drug discovery research. Template selection for each GPCR orphan has been done based on the degree of relatedness between orphans and characterized GPCR with known 3D structure: sequence and binding similarity and the minimum distance between each pair of orphan and characterized GPCR.

A number of 89 structures of 23 GPCR receptors have been elucidated so far and deposited in the Protein Data Bank (Table 1). Each one of these structures constitutes a potential GPCR template structure for the comparative modeling of the other GPCR.

Table 1

GPCRs with known three-dimensional structures, deposited in PDB

CLASS A GPCR					
GPCR receptor[a]	**Receptor state[b]**	**Crystallized ligand[b,c]**	**X-Ray Res**	**R-factor**	**R-free factor**
Bovine rhodopsin	Ground state	Retinal (10)	2.2-4.1	0.175-0.384	0.212-0.418
		Ligand free	2.90	0.231	0.266
	Active	Retinal, β-ionone	2.60	0.224	0.262
		Retinal	4.15	0.378	0.382
		Ligand free	3.2	0.215	0.248
	Intermediate	retinal	2.95	0.243	0.289
	Bathorhodopsin		2.6	0.178	0.181
	Lumirhodopsin		2.80	0.218	0.238
	Metarhodopsin II (3+1)		2.85-3.3	0.217-0.218	0.245-0.262
			NMR		
	Isorhodopsin		3.4	0.29	0.330
	Dark-adapted		NMR		
Beta 2 adrenergic receptor	Inactive	Carazol (3)	2.4-3.4	0.198-0.226	0.232-0.28
		Timolol	2.80	0.230	0.273
		ICI118551	2.84	0.229	0.291
		AC1LIXHF	2.84	0.222	0.278
		Alprenolol	3.16	0.235	0.290
		ERC[d]	3.50	0.241	0.283
		Ligand free	3.4	0.242	0.28
	Active (2)	P0G	3.2-3.5	0.228-0.243	0.277-0.308
Squid rhodopsin	Ground state (2)	retinal	2.5-3.7	0.188-0.302	0.206-0.330
	Bathorhodopsin		2.8	0.295	0.345
	Isorhodopsin		2.7	0.18	0.2
Turkey beta1 adrenergic	Inactive	**cyanopindolol** (3)	2.7-3.25	0.215-0.274	0.268-0.325
		I-cyanopindolol	3.65	0.25	0.27
		Carazolol	3.00	0.25	0.295
		3WC	2.7	0.226	0.266
		XF5	2.80	0.223	0.274
		Dobutamine (2)	2.50-2.6	0.228	0.266
		Salbutamol	3.05	0.225	0.255
		Carmoterol	2.60	0.231	0.268
		Isoprenaline	2.85	0.222	0.254
		Bucindolol[e]	3.2	2.44	0.279
		Carvedilol[e]	2.30	0.204	0.240
		Ligand free	3.5	0.312	0.355
Adenosine A2A	Inactive	**ZM241385 (5)**	1.8-3.3	0.176-0.278	0.213-0.316
		XAC	3.31	0.3	0.319
		Caffeine	3.60	0.298	0.311
		T4G	3.27	0.295	0.335
		T4E	3.34	0.282	0.29
		Adenosine	3.00	0.246	0.269
		NECA	2.60	0.234	0.258
	Active	UK432097	2.71	0.220	0.273
Dopamine D3R	Inactive	**Eticlopride**	2.89	0.245	0.272
Chemokine CXCR4	Inactive	**IT1t (4)**	2.5-3.2	0.240	0.282
		CVX15	2.90	0.217	0.267
Histamine H1R	Inactive	**Doxepin**	3.1	0.217	0.249
S1P1R	Inactive (2)	**ML5**	2.8-3.35	0.228-0.231	0.272-0.281
Muscarinic M2R	Inactive	**QNB**	3.00	0.227	0.276
Rat muscarinic M3	Inactive	OHK	3.4	0.254	0.303
κ-type opioid	Inactive	**JDTic**	2.90	0.228	0.265
Mu-type opioid	Inactive	**Morphinan**	2.8	0.235	0.275
Delta-type opioid	Inactive	**Naltrindole**	3.4	0.255	0.282
N/OFQ opioid	Inactive	**C-24**	3.01	0.25	0.288

Table 1 (continued)

PAR1	Inactive	**Vorapaxar**	2.2	0.219	0.235
Chemokine CXCR1	Inactive		RMN		
Rat Neurotensin NTS1	Inactive	Neurotensin	2.8	0.228	0.282
5HT1B	Inactive	dihydroergotamine	2.8	0.237	0.257
	Inactive	Ergotamine	2.7	0.225	0.261
5HT2B	Inactive	Ergotamine	2.7	0.227	0.266
CLASS F GPCR					
Human smoothened	Inactive	**LY2940680**	2.45	0.201	0.231
CLASS B GPCR					
Glucagon receptor	Inactive	No ligand	3.3	0.286	0.338
CRF1	Inactive	**CP-376395**	2.98	0.243	0.265

[a] all receptors are human, otherwise is specified
[b] When multiple structures are possible, the number is give between brackets
[c] Ligand type is presented as follows: bold letters-antagonist, italic letters partial agonist, underlined letters-inverse agonist, normal letters – agonist
[d] irreversible agonist
[e] biased agonist

Table 2

Selected template structures for building comparative models of class A orphan GPCRs

GPCR template	**All sequences**	**Class A orphans_consensus**
5HT1B	GPR62; GP161; GP149	GPR45; GPR63; GP151; GP161; GP149; GP160
5HT2B	-	GP173; ETBR2
A2A	GP162	GP119; MRGRF
ACM2	-	LGR5
beta1AR	GPR45; GPR63; GPR85; GPR88; GP101; GP119; GP135; GP151	GPR84; GPR85; GPR88; GP101
Beta2AR	GP160	-
brho	GPR84; GP153; OPN3	GP153; OPN3
CXCR1	GPR15; GPR25; GPR27; GPR31; GPR35; GPR61; GPR75; GPR78; GP146; GP152; GP173; MAS1L; MRGRD;MRGRE; MRGRF; MRGX1; MRGX2; MRGX3; MRGX4; O3FA1; PSYR; SUCR1; LGR4	GPR15; GPR25; GPR27; GPR31; GPR35; GPR61; GPR75; GPR78; GP146; GP152; GP176; GP182; MAS1L; MRGRD; MRGRE; MRGX1; MRGX2; MRGX3; MRGX4; O3FA1; PSYR; LGR4
CXCR4	GPR4; GPR18; GPR33; GPR87; GP171; CCRL2; OGR1; OXGR1	GPR33; CCRL2; OGR1; SUCR1
D3	GPR19; GP148; MTR1L; LGR6	GPR19; GPR26; GPR62; GP148; MTR1L; LGR6
delta	GPR32; GPR82; CML1	GPR32; GPR82; CML1
H1	GPR22; MAS; LGR5	GPR22; GP135
kappa	GPR37; GP182; ETBR2; P2Y10	-
mu	-	GPR1; GPR18; GPR37; GPR87
nociceptin	GPR1; GPR83; GP141; GP176; MRGRG	GPR83; GP141
NT1	GPR39; GP139; GP142; GP150	GPR39; GP139; GP142; GP150
PAR1	GPR17; GPR20; GPR42; GPR55; GP132; GP174; GP183; P2RY8	GPR4; GPR17; GPR20; GPR34; GPR42; GPR55; GP132; GP171; GP174; GP183; MAS; OXGR1; P2Y10; P2RY8
rM3	GPR21; GPR26; GPR52	GPR21; GPR52; GP162
S1P1	GPR3; GPR6; GPR12	GPR3; GPR6; GPR12; MRGRG
srho	OPN5	OPN5

As expected, the first choice for orphans from every GPCR class is a crystal structure of a receptor with the same class affiliation. For example, the class B or F templates are not in any case the closest relatives for class A orphans (Table 2). In the same manner, the recommended template for class B orphans is always one of the two class B receptor structures: glucagons receptor or corticotrophin releasing factor receptor.

The results for the class A orphans are presented in Table 2. Given the high number of 3D structures available for class A GPCRs crystallized with active agonists or antagonists (Table 1), binding site similarity was considered in the definition of the closest homologue for orphans and characterized GPCR. Thus, the amino acids found at 5Å around crystallized ligands have been labeled in the final alignment containing the potential templates and the fragments were used together with the overall sequence similarity to approximate the degree of relatedness between each orphan and characterized GPCRs. As results from Table 2, the binding site similarity has influenced the minimum distance between 27 pairs of orphans and characterized GPCR.

Except for five cases (hCELSR2, hCELSR3, hEMR1, GP124 and GP133), the corticotrophin-releasing factor 1 receptor (CRF1) is the general template structure for class B orphans. However, given the high sequence similarity (more than 50%) between CRF1 and glucagon receptor, the distances in the protein space between the two templates and each class B orphan are insignificant (Fig. 1) therefore the best template selection is recommended to be performed based on crystal structures quality parameters (Table 1) or based on particular needs for each study. For class B GPCR, binding site similarity was not considered for the calculation of the smallest distance between orphans and characterized GPCRs, because there is only one crystal structure with a bound ligand compared to the big number of orphans (12). In class A GPCR the ratio between the number of crystal structures with a bound ligand and the number of class A ratio is 1:1.03.

We have conducted a similar study for class C orphans, and the closest relatives from the available crystal structure were estimated. The results are presented in Table 3.

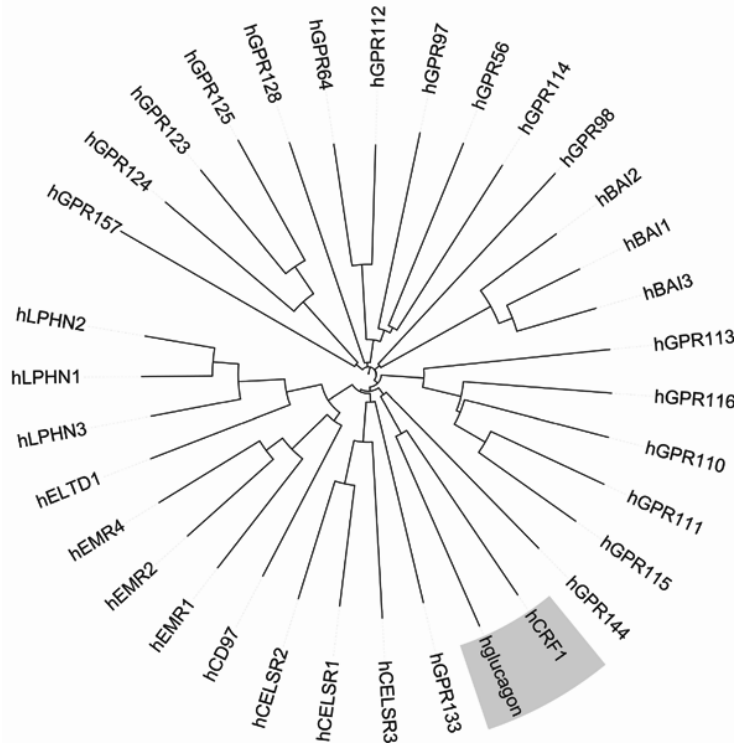


Fig. 1 – Phylogenetic tree of the orphan and the two characterized class B GPCRs (gray background).

Table 3

Selected template structures for building comparative models of class C orphan GPCRs

Class C orphans	Selected template
GPRC5A	S1P1
GPRC5B	A2A
GPRC5C	S1P1
GPRC5D	S1P1
GPR156	bRHO
GPR158	CXCR1
GPR179	glucagon

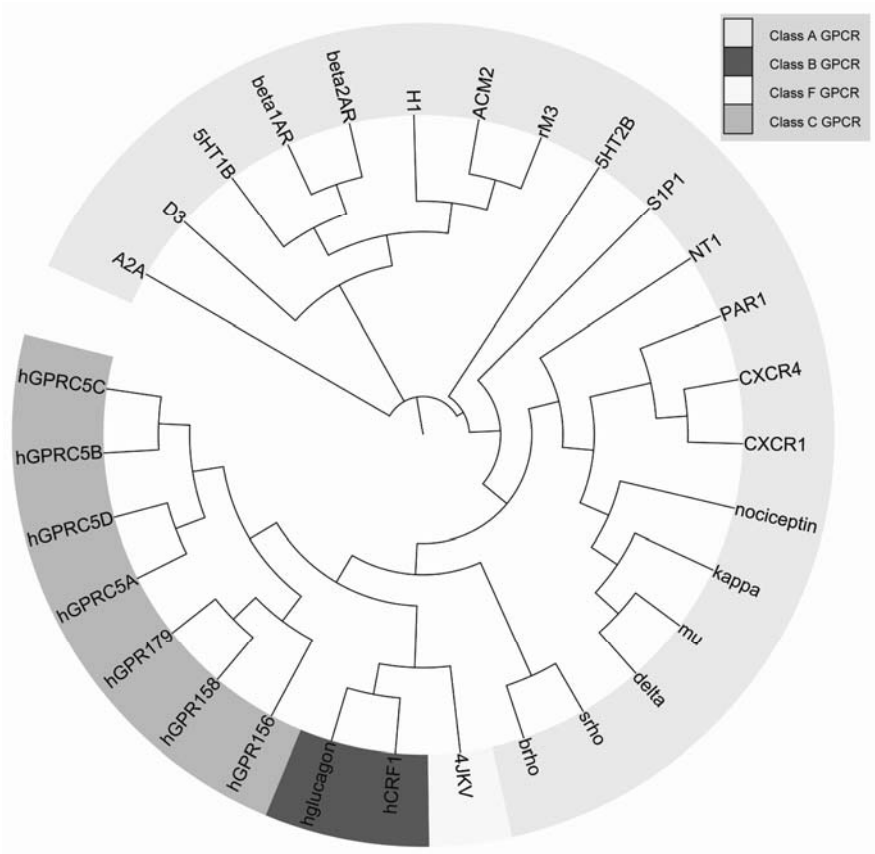


Fig. 2 – Phylogenetic trees of class C orphans and characterized GPCRs with known crystal structure.

However, a class A structure is not the best choice for building of a class C homology model due to the lower sequence similarity between classes, as results also from Fig. 2, where the separation of receptors according to their class affiliation can be noticed. For the special cases when a 3D structure is needed, supplementary corrections and refinements of the model should be performed with the purpose of adjusting the model according to available experimental data like mutagenesis analysis, or ligand binding.

CONCLUSIONS

Latest progresses recorded in the structural information field of GPCR offer new opportunities for the structure-based drug design for this family in general, and for orphan receptors in special. The most appropriate template structure for each orphan GPCR was identified by exploring the sequence and binding similarity. Except for class C GPCR, the selected templates have the same class affiliation with the given orphan receptor.

Acknowlegments: This work was supported by CNCSIS-UEFISCU project number PN-II-RU-PD-500, agreement no. 119/5.08.2010.

REFERENCES

1. K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E Stenkamp, M. Yamamoto, M. Miyano, *Science*, **2000**, *289,*739-745.
2. C.G. Tate, *Trends Biochem. Sci.,* **2012**, *37*, 343-352.
3. S.G. Rasmussen, H.J. Choi, D.M. Rosenbaum, T.S. Kobilka, F.S. Thian, P.C. Edwards, M. Burghammer, V.R. Ratnala, R. Sanishvili, R.F. Fischetti, G.F.X. Schertler, W.I. Weis and B.K. Kobilka, *Nature*, **2007**, *450*, 383-387.
4. T. Shimamura, K. Hiraki, N. Takahashi, T. Hori, H. Ago, K. Masuda, K. Takio, M. Ishiguro and M. Miyano *J. Biol. Chem*., **2008**, *283*, 17753-17756.
5. T. Warne, M.J. Serrano-Vega, J.G. Baker, R. Moukhametzianov, P.C. Edwards, R. Henderson, A.G. Leslie, C.G. Tate CG, and Schertler GF, *Nature*, **2008**, *454*, 486-491
6. V.P. Jaakola, M.T. Griffith, M.A. Hanson, V. Cherezov, E.Y. Chien, J.R. Lane, A.P. Ijzerman, and R.C. Stevens, *Science*, **2008**, *322*, 1211-1217.
7. B. Wu, E.Y. Chien, C.D. Mol ,G. Fenalti, W. Liu, V. Katritch, R. Abagyan, A. Brooun, P. Wells, F.C. Bi, D.J. Hamel, P. Kuhn, T.M. Handel, V. Cherezov, R.C. Stevens, *Science*, **2010**, *330*, 1066-1071.
8. E.Y. Chien, W. Liu, Q. Zhao, V. Katritch, G.W. Han, M.A. Hanson, L. Shi, A.H. Newman, J.A. Javitch, V. Cherezov, R.C. Stevens, *Science*, **2010**, *330*, 1091-1095.
9. T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W.

Liu, G.W. Han, T. Kobayashi, R.C. Stevens, S. Iwata, *Nature*, **2011**, *475*, 65-70.
10. K. Haga, A.C. Kruse, H. Asada, T. Yurugi-Kobayashi, M. Shiroishi, C. Zhang, W.I. Weis, T. Okada, B.K. Kobilka, T. Haga T, T. Kobayashi, *Nature*, *2012*, *482*, 547-551.
11. A.C. Kruse,J. Hu, A.C. Pan, D.H. Arlow, D.M. Rosenbaum, E. Rosemond, H.F. Green, T. Liu, P.S. Chae, R.O. Dror, D.E. Shaw, W.I. Weis, J. Wess, B.K. Kobilka. *Nature*, **2012**, *482*, 552-556.
12. M.A. Hanson,C.B. Roth, E. Jo, M.T. Griffith, F.L. Scott, R. Reinhart, H. Desale, B. Clemons, S.M. Cahalan, S.C. Schuerer, M.G. Sanna, G.W. Han, P. Kuhn, H. Rosen, R.C. Stevens, *Science*, **2012**, *335*, 851-855.
13. H. Wu, D. Wacker, M. Mileni, V. Katritch, G.W. Han, E. Vardy, W. Liu, A.A. Thompson, X.P. Huang, F.I. Carroll, S.W. Mascarella, R.B. Westkaemper, P.D. Mosier, B.L. Roth, V. Cherezov, R. C.Stevens, *Nature*, **2012**, *485,* 327-332.
14. A. Manglik, A.C. Kruse, T.S. Kobilka, F.S. Thian, J.M. Mathiesen,R.K. Sunahara,L. Pardo, W.I. Weis, B.K. Kobilka, and S. Granier, *Nature*, **2012**, *485,* 321-326.
15. S. Granier, A. Manglik, A.C. Kruse, T.S. Kobilka, F.S. Thian, W.I. Weis, and B.K. Kobilka, *Nature*, **2012**, *485,* 400-404.
16. A.A. Thompson, W. Liu, E. Chun, V. Katritch, H. Wu, E. Vardy, X.P. Huang, C. Trapella, R. Guerrini, G. Calo, B.L. Roth, V. Cherezov, R.C. Stevens, RC, *Nature*, **2012**, *485,* 395-399.
17. C. Zhang, Srinivasan Y, Arlow DH, Fung JJ, Palmer D, Zheng Y, Green HF, Pandey A, Dror RO, Shaw DE, W.I. Weis, S.R. Coughlin, B.K. Kobilka, *Nature*, **2012**, *492,* 387-92.
18. S.H. Park, B.B. Das, F. Casagrande, Y. Tian, H.J. Nothnagel, M. Chu, H. Kiefer, K. Maier, A.A. De Angelis, F.M. Marassi,S.J. Opella, *Nature*, **2012**, *491,* 779-83.
19. J.F. White, N. Noinaj, Y. Shibata, J. Love, B. Kloss, F. Xu, J. Gvozdenovic-Jeremic, P. Shah, J. Shiloach, C.G. Tate, R. Grisshammer, *Nature*, **2012**, *491,*508-513.
20. C. Wang, Y. Jiang, J. Ma, H. Wu, D. Wacker, V. Katritch, G.W. Han, W. Liu, X.P. Huang, E. Vardy, J.D. McCorvy, X. Gao, X.E. Zhou, K. Melcher, C. Zhang, F. Bai, H. Yang, L. Yang, H. Jiang, B.L. Roth, V. Cherezov, R.C. Stevens, Xu HE, *Science*, **2013**, *340*, 610-614.
21. F.Y. Siu, M. He, C. de Graaf, G.W. Han, D. Yang, Z. Zhang, C. Zhou, Q. Xu, D. Wacker, J.S. Joseph, W. Liu, J. Lau, V. Cherezov, V. Katritch, M.W. Wang, R.C. Stevens, *Nature*, **2013**, *499,* 444-449.
22. K. Hollenstein, J. Kean, A. Bortolato, R.K. Cheng, A.S. Doré, A. Jazayeri, R.M. Cooke, M. Weir, F.H. Marshall., *Nature*, **2013**, *499,* 438-43.
23. C. Wang, H. Wu, V. Katritch, G.W. Han, X.P. Huang, W. Liu, F.Y. Siu , B.L. Roth, V. Cherezov, R.C. Stevens. *Nature*, **2013**, *497,* 338-343.
24. The Research Collaboratory for Structural Bioinformatics PDB, http://www.rcsb.org/pdb/
25. Consortium TU, *Nucleic Acids Res*., **2012**, *40*, D71–5.
26. C. Notredame, D.G. Higgins and J. Heringa, *J. Mol. Biol.*, **2000**, *30*2, 205-217.
27. J.M. Baldwin, G.F. Schertler,V.M. Unger, *J. Mol. Biol.*, **1997**, *272*, 144-164.
28. M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna,, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson and D.G. Higgins, *Bioinformatics,* **2007**, *23*, 2947-2948.
29. I. Letunic and P. Bork, *Bioinformatics*, **2006**, *23*, 127-128.
30. I. Letunic and P. Bork, *Nucleic Acids Res*., **2011**, doi: 10.1093/nar/gkr201.