Structure-based Drug Design and Drug Discovery for G Protein-coupled Receptors

Tony Ngo, Angela Finch, Renate Griffith*

Department of Pharmacology, School of Medical Sciences, University of New South Wales, NSW 2052
**Corresponding author: r.griffith@unsw.edu.au*

Introduction

One of the major obstacles for G protein-coupled receptor (GPCR) drug discovery has been the lack of structural information available. Membrane proteins have always been difficult to crystallise and therefore the publication of a bovine rhodopsin crystal structure (PDB ID: 1F88) in 2000, the first ever for GPCRs, signified a turning point in GPCR drug discovery. Over the last decade, we have seen a dramatic improvement in crystallisation methods, particularly through protein engineering with the GPCR-T4-lysozyme fusion protein. This has lead to the publication of structures of thirteen GPCRs to date, including the β_2-adrenoceptor (β_2-AR), dopamine D_3 (D_3R), muscarinic M_2 and the serotonin 5-HT_{1B} receptors. Another challenge for Class A GPCRs in particular is the design of selective ligands. This is due to the highly conserved orthosteric binding pocket within the transmembrane regions, which leads to off-target side effects, as a result of binding to related receptors.

The availability of GPCR crystal structures opened up opportunities to use alternative methods for GPCR drug discovery, especially structure-based drug design (SBDD). SBDD approaches involve discovering novel compounds based on three dimensional (3D) protein structures using various computational methods. The impact of GPCR crystal structures on SBDD has been immediate and has led to the discovery of novel ligands for multiple GPCRs. The crystal structures have also provided opportunities for homology modelling. This review will present an overview of structure-based approaches for identifying novel leads for GPCR-targeting drugs.

Docking-based Virtual Screening

GPCR crystal structures have enabled virtual screening of large chemical libraries through docking, which is cost effective compared to typical high-throughput screening programs, and can rapidly identify potential ligands (1). Docking predicts the binding orientation of a small molecule and explores interactions it may form, with the quality of the protein-ligand interactions assessed by a variety of scoring functions (2). Following docking, the top ranked 'hits' based on docking score, are selected using a set of criteria including chemical diversity, drug-likeness and predicted interactions with key residues within the binding pocket. After selection, these compounds are purchased and tested, using different assays for G protein-mediated activity and/or binding at the target GPCR. This approach has been very successful in identifying novel classes of ligands for a range of GPCRs implicated in various diseases (**Table 1**).

Even though docking-based virtual screening has brought success in drug development, there are a number of points to consider. A recent evaluation of the performance of the four most commonly used docking programs (GOLD, Glide, LigandFit, Surflex) found that, although docking was able to predict the native ligand pose in the crystal structure, docking scores did not correlate with binding activities (3). This means that potential highly active compounds are not selected for testing, but this has not prevented high hit rates for GPCRs (**Table 1**). It has been suggested that this is due to the bias of the lead-like subset of the ZINC compound database towards GPCR ligands (4), and a relatively low hit rate was indeed obtained using the histamine H_4 receptor homology model and the complete ZINC database. However, using other databases achieved similarly high hit rates for the β_2-AR and the adenosine A_{2A} receptor (**Table 1**). It has also been shown that the quality of the database compounds with respect to input conformations and ionisation states can affect docking score and predicted interactions (3), and this quality varies between studies. Notwithstanding these limitations, it is evident that docking-based virtual screening approaches are viable for the identification of novel ligands for various GPCRs.

Although no conscious effort was made to address selectivity, two of the active ligands (compound 3 and 7) from the D_3R screen, showed at least a six-fold better affinity for the D_3R than the D_2R and all active compounds had no measurable affinity at the β_2-AR (5). Similar results were found for the adenosine $A_{2A}R$ screen, where two ligands (compounds 17 and 50) displayed at least ten-fold selectivity for the adenosine $A_{2A}R$ over the adenosine A_1R (6). While this is promising, ligand selectivity remains largely unaddressed in docking-based virtual screening studies. This could be achieved by docking the top 'hits' into different GPCRs (off-targets) as an additional measure for compound selection.

Structure-based Pharmacophores: a Novel Structure-based Approach

Virtual screening of chemical databases using structure-based pharmacophores offers an alternative approach to docking to identify novel ligands. Pharmacophores encode the spatial arrangement of features required of a ligand to interact with a biological target to produce a desired response. These features include hydrogen bond acceptors/donors, hydrophobic and charged groups. Structure-based pharmacophores utilise structural information about the binding site and allow targeting of previously unexplored spaces. This is different to ligand-based pharmacophores, which are 3D-quantitative structure-activity relationship (QSAR) models constructed

Table 1. GPCR drug discovery examples.

Target	Chemical database	Hit rate (actives/tested)	Potential therapeutic areas
β_2-AR (19)	Combination	20% (30/150)	Pulmonary disorders
β_2-AR (4)	ZINC: lead-like subset	24% (6/25)	Pulmonary disorders
Adenosine A_{2A} (6)	MolSoft ScreenPub	41% (23/56)	Neurodegenerative and cardiovascular disorders
CXCR4 (10)	ZINC: lead-like subset	17% (4/23) (crystal structure) 4% (1/24) (homology model)	HIV and various cancers
Dopamine D_3 (5)	ZINC: lead-like subset	20% (5/20) (crystal structure) 23% (6/26) (homology model)	Parkinson's disease and schizophrenia
Histamine H_1 (20)	ZINC: curated fragment subset	73% (19/26)	Allergens
Histamine H_4 (21)	ZINC: total, some curation	6% (16/255) (homology model)	Allergens and inflammation

from a training set of known ligands, and complex-based pharmacophores, developed from a protein-ligand complex structure or docking pose. These pharmacophores have the disadvantage that the ligand can heavily bias the features and their arrangement in space.

To develop structure-based pharmacophores, a map of potential protein-ligand interactions is generated from the proposed binding pocket(s) of the GPCR target, based on grid-based interaction energies of small probes, binding of small chemical fragments, and/or a combination of these approaches (7,8). These interaction maps can then be translated into pharmacophore features. One issue that arises is the large number of features generated in the binding pocket, which needs to be reduced to allow real, small molecules to fit them. Various approaches can be used to reduce features including the clustering of closely situated features, the removal of inaccessible features and comparison to docked poses of known ligands and to site-directed mutagenesis results. After refinement, the pharmacophores can be used for virtual screening with top 'hits' selected, as described above, and tested for biological activity.

The use of structure-based pharmacophores for virtual screening is well established for various proteases and kinases (reviewed in (9)). However, there is limited literature for the use of this method for GPCRs. This has been due to the lack of GPCR structural information, but the recent GPCR crystal structures may direct GPCR drug discovery research towards structure-based pharmacophores.

Crystal Structures versus Homology Models

The most important basis of SBDD is the availability of the target GPCR structure. Although there has been success in obtaining several different GPCR crystal structures, this still only represents a small percentage of all GPCRs. Thus, many drug discovery efforts need to be based on GPCR homology models.

Contradictions have been encountered when comparing docking-based virtual screening results using a crystal structure versus a homology model. For the crystal structure and homology model of the D_3R, hit rates of 20% and 23% were achieved (5), while for the CXCR4 receptor hit rates of 17% (crystal structure) and 4% (homology model) were found (10). This could be explained by the accuracy of the models, which depends on the similarity between the crystal structure (template) and model sequences. However, it should be noted that, although hit rates were similar, unique hit lists were obtained when screening with the D_3R crystal structure and homology model, respectively. A recent evaluation of docking accuracy between GPCR crystal structures and homology models confirmed that the reliability of the models was dependent on sequence similarity and suggested a minimum of 30% sequence identity between the template and model is required (11).

Another concern for GPCR homology models is the structure of the extracellular loops (ECLs), particularly ECL2. Depending on the template used, different loop conformations will be obtained because of the differences seen in flexibility and structure in current GPCR crystal structures (**Fig. 1**). For example, the D_3R has a long flexible ECL2 (blue in **Fig. 1**) which protrudes into the binding pocket, compared to the helical structure found in the β_2-AR (green). The structure and placement of the ECLs is important as they play a significant role in ligand recognition (12) and can therefore affect docking results. This is particularly important for the SBDD of allosteric modulators (described below) where the ECLs act as a ligand binding site. It was shown that removal of the ECLs, in cases where their position cannot be predicted from the template, improved docking accuracy, whereas docking quality diminished when accurately modelled ECLs from close homologues were removed (11). This illustrates the importance of proper loop placement for docking accuracy in homology models and the careful selection of a structural template.

The crystal structures published so far are all Class A GPCRs. This limits the reliability of homology models for other classes of GPCRs. Other classes of GPCRs are difficult to crystallise due to the large N-termini, which act as the orthosteric binding pocket for Class B and Class C GPCRs. However, since the N-termini are large, they can fold into independent structures and retain ligand binding ability, which has allowed the structure determination of several extracellular domains of Class B GPCRs on their own, including corticotropin-releasing factor type 1 (CRF-1) receptor (PDB ID: 3EHT) and parathyroid hormone 1 receptor (PTH1R, PDB ID: 3C4M). Application of structure-based approaches to identify novel ligands using these extracellular domain structures is limited and the inability to solve crystal structures of the whole receptor remains a challenge in GPCR structural biology and limits opportunities to use SBDD techniques.

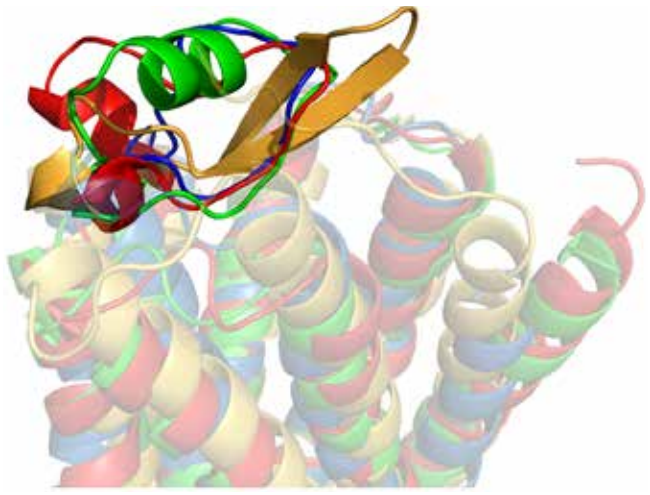


Fig. 1. Superimposition of selected GPCR crystal structures.
The different structures of the second extracellular loop are highlighted, using four GPCR structures: β_2-AR, green (PDB ID: 3NY9); D_3R, blue (PDB ID: 3PBL); CXCR4 receptor, orange (PDB ID: 3ODU); muscarinic M_2 receptor, red (PDB ID: 3UON). The structures were superimposed using the most conserved amino acid in each transmembrane helix as tethers.

Orphan receptors and allosteric modulators

The SBDD techniques described here provide exciting prospects for tackling orphan GPCRs, which are receptors whose endogenous ligands are yet to be identified. Currently, there are 133 known orphan GPCRs, including 92 Class A GPCRs. Orphan receptors have the potential to be novel drug targets for therapeutics. For example, GPR37L1 has been implicated in cardiovascular disease (13), however there is limited knowledge of its pharmacology, since there is no known ligand. With the lack of ligand information available for orphan GPCRs, structure-based techniques have great potential to identify potential ligands, which then can be used to probe function and propose an endogenous ligand.

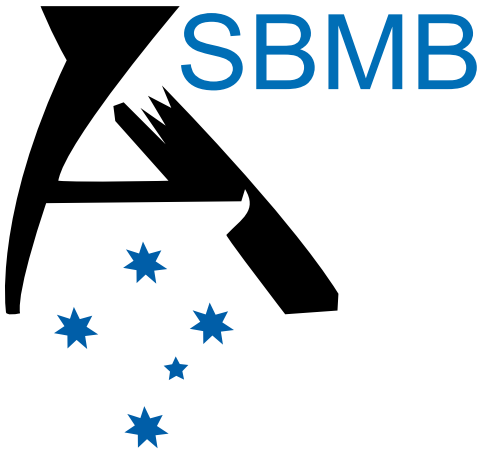
There is also currently a growing focus in GPCR research on the discovery of allosteric modulators over typical orthosteric agonists and antagonists because of the potential therapeutic benefits, including selectivity (14). Several studies have highlighted the critical role of ECL residues, where there is less conservation between subtypes, in mediating allosteric effects at the muscarinic M_2 and M_4 receptors (15,16) and the existence of an extracellular vestibule acting as a secondary binding site in Class A GPCRs (17,18). Allosteric modulators are known only for a limited number of GPCRs, but with the proven success of docking-based virtual screening and the potential of structure-based pharmacophores, this is likely to improve in the near future.

The advent of GPCR structures has unlocked several doors for drug discovery research. The ability to utilise well established structure-based techniques has had an immediate impact in identifying novel ligands for GPCRs and has provided an efficient alternative to high-throughput screening. This signifies a change in approach to GPCR drug discovery. Applications of structure-based methods show great promise for the discovery of allosteric modulators and ligands for orphan receptors and, with the continued progress in obtaining crystal structures, SBDD remains a key component of GPCR drug discovery.

References

1. Klebe, G. (2006) *Drug. Discov. Today* 11, 580-594
2. Waszkowycz, B., Clark, D.E., and Gancia, E. (2011) *WIREs Comput. Mol. Sci.* **1**, 229-259
3. Li, X., Li, Y., Cheng, T., Liu, Z., and Wang, R. (2010) *J. Comput. Chem.* **31**, 2109-2125
4. Kolb, P., Rosenbaum, D.M., Irwin, J.J., Fung, J.J., Kobilka, B.K., and Shoichet, B.K. (2009) *Proc. Natl. Acad. Sci. USA* **106**, 6843-6848
5. Carlsson, J., Coleman, R.G., Setola, V., Irwin, J.J., Fan, H., Schlessinger, A., Sali, A., Roth, B.L., and Shoichet, B.K. (2011) *Nat. Chem. Biol.* **7**, 769-778
6. Katritch, V., Jaakola, V.P., Lane, J.R., Lin, J., Ijzerman, A.P., Yeager, M., Kufareva, I., Stevens, R.C., and Abagyan, R. (2010) *J. Med. Chem.* **53**, 1799-1809
7. Barillari, C., Marcou, G., and Rognan, D. (2008) *J. Chem. Inf. Model.* **48**, 1396-1410
8. Sanders, M.P., Verhoeven, S., de Graaf, C., Roumen, L., Vroling, B., Nabuurs, S.B., de Vlieg, J., and Klomp, J.P. (2011) *J. Chem. Inf. Model.* **51**, 2277-2292
9. Sanders, M.P.A., McGuire, R., Roumen, L., de Esch, I.J.P., de Vlieg, J., Klomp, J.P.G., and de Graaf, C. (2012) *Med. Chem. Comm.* **3**, 28-38
10. Mysinger, M.M., Weiss, D.R., Ziarek, J.J., Gravel, S., Doak, A.K., Karpiak, J., Heveker, N., Shoichet, B.K., and Volkman, B.F. (2012) *Proc. Natl. Acad. Sci. USA* **109**, 5517-5522
11. Beuming, T., and Sherman, W. (2012) *J. Chem. Inf. Model.* **52**, 3263-3277
12. Peeters, M.C., van Westen, G.J., Li, Q., and IJzerman, A.P. (2011) *Trends Pharmacol. Sci.* **32**, 35-42
13. Min, K.D., Asakura, M., Liao, Y., Nakamaru, K., Okazaki, H., Takahashi, T., Fujimoto, K., Ito, S., Takahashi, A., Asanuma, H., Yamazaki, S., Minamino, T., Sanada, S., Seguchi, O., Nakano, A., Ando, Y., Otsuka, T., Furukawa, H., Isomura, T., Takashima, S., Mochizuki, N., and Kitakaze, M. (2010) *Biochem. Biophys. Res. Commun.* **393**, 55-60

References continued on page 12



References continued from page 6

14. Melancon, B.J., Hopkins, C.R., Wood, M.R., Emmitte, K.A., Niswender, C.M., Christopoulos, A., Conn, P.J., and Lindsley, C.W. (2012) *J. Med. Chem.* **55**, 1445-1464
15. Nawaratne, V., Leach, K., Felder, C.C., Sexton, P.M., and Christopoulos, A. (2010) *J. Biol. Chem.* **285**, 19012-19021
16. Bock, A., Merten, N., Schrage, R., Dallanoce, C., Batz, J., Klockner, J., Schmitz, J., Matera, C., Simon, K., Kebig, A., Peters, L., Muller, A., Schrobang-Ley, J., Trankle, C., Hoffmann, C., De Amici, M., Holzgrabe, U., Kostenis, E., and Mohr, K. (2012) *Nat. Commun.* **3**, 1044
17. Dror, R.O., Pan, A.C., Arlow, D.H., Borhani, D.W., Maragakis, P., Shan, Y., Xu, H., and Shaw, D.E. (2011) *Proc. Natl. Acad. Sci. USA* **108**, 13118-13123
18. Kruse, A.C., Hu, J., Pan, A.C., Arlow, D.H., Rosenbaum, D.M., Rosemond, E., Green, H.F., Liu, T., Chae, P.S., Dror, R.O., Shaw, D.E., Weis, W.I., Wess, J., and Kobilka, B.K. (2012) *Nature* **482**, 552-556
19. Sabio, M., Jones, K., and Topiol, S. (2008) *Bioorg. Med. Chem. Lett.* **18**, 5391-5395
20. de Graaf, C., Kooistra, A.J., Vischer, H.F., Katritch, V., Kuijer, M., Shiroishi, M., Iwata, S., Shimamura, T., Stevens, R.C., de Esch, I.J., and Leurs, R. (2011) *J. Med. Chem.* **54**, 8195-8206
21. Kiss, R., Kiss, B., Könczöl, A., Szalai, F., Jelinek, I., László, V., Noszál, B., Falus, A., and Keseru, G.M. (2008) *J. Med. Chem.* **51**, 3145-3153