Cell
PRESS

Structure-based drug screening for G-protein-coupled receptors

Brian K. Shoichet[1] and Brian K. Kobilka[2]

[1] Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, USA
[2] Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA, USA

G-protein-coupled receptors (GPCRs) represent a large family of signaling proteins that includes many therapeutic targets; however, progress in identifying new small molecule drugs has been disappointing. The past 4 years have seen remarkable progress in the structural biology of GPCRs, raising the possibility of applying structure-based approaches to GPCR drug discovery efforts. Of the various structure-based approaches that have been applied to soluble protein targets, such as proteases and kinases, _in silico_ docking is among the most ready applicable to GPCRs. Early studies suggest that GPCR binding pockets are well suited to docking, and docking screens have identified potent and novel compounds for these targets. This review will focus on the current state of _in silico_ docking for GPCRs.

Introduction

GPCRs mediate cellular responses to the majority of hormones and neurotransmitters, and are therefore attractive targets for drug discovery. In the decade following the cloning of the first genes and cDNAs for GPCRs [1,2], there was great hope that these discoveries would rapidly translate into new and more effective therapeutics. Cloning and later mining the human genome sequence led to the identification of new GPCR subtypes [3] and the establishment of cell lines that could be used for high-throughput screening (HTS) of large compound libraries. The identification of polymorphisms for specific GPCRs suggested the potential for individualized medicines. Unfortunately, the promise of new drugs for new GPCR targets or safer and more effective drugs for previously identified targets has largely gone unfulfilled [4].

Several reasons may explain the slow pace of drug discovery in the face of more targets and screening modalities. If the advent of the molecular era gave us unprecedented tools and abundant targets, it also disrupted the integrated, tissue-based pharmacology of the classical era of drug discovery [5,6]; the underlying biology was more complicated than anticipated by the reductionist, molecular view. Many GPCRs signal through multiple pathways, often in a ligand-specific manner. For example, the β_2 adrenergic receptor (β_2AR) activates specific cellular signaling pathways through G_s, the stimulatory G protein for adenylyl cyclase, and independently through arrestin. Thus, the drug carvedilol is an inverse agonist for β_2AR activation of G_s, but a partial agonist for activation of

arrestin [7]. Consequently, high-throughput assays that only monitor a single readout, for instance cAMP accumulation or calcium flux, may not reflect the physiologically relevant signaling pathway [8]. Not only do we need to identify the correct GPCR target and signaling pathway, we must find a drug with the appropriate efficacy profile: agonist, partial agonist, neutral antagonist and inverse agonist. Drugs that satisfy these criteria must then pass through a gauntlet of assays to assess toxicology and pharmacokinetics. For this and other reasons, the cost of drug development has escalated whereas revenue from new drugs has slipped [9]. Consequently, some pharmaceutical companies are abandoning small molecule development programs in favor of biologics [10] and the cost of the few new drugs that make it to the market will further escalate the cost of healthcare.

In 2007, we entered the new era of GPCR structural biology. Since the initial crystal structures of the β_2AR [11] and the β_1AR [12], the number of published GPCRs which have yielded to crystallography has grown to ten and includes the adenosine A2A receptor [13], the D3 dopamine receptor [14], the CXCR4 receptor [15], the histamine H1 receptor, [16], the sphingosine 1 phosphate receptor [17], the M2 and M3 muscarinic receptors [18,19], and the mu opioid receptor [20], with at least two new structures anticipated in 2012. This is largely attributable to the application of high-throughput methods for lipidic cubic phase (LCP) crystallography [21] and protein engineering with GPCR–T4 lysozyme [11,22] and thermostabilization [23] methods being generally applicable to structurally diverse GPCRs. Although structural biology is not a panacea for the challenges described above, there is reason to hope that GPCR crystal structures can facilitate drug discovery based on success with soluble protein targets such as kinases and proteases. In this review, we will discuss the application of structure-based screens of large compound libraries to GPCR drug discovery.

Structure-based screens for new ligands

Structure-based design has been pivotal in the development of over ten marketed drugs, including recent successes against renin with aliskiren [24], and against hepatitis C virus protease with telapravir [25], and has contributed to the development of multiple others, since the technique came into widespread use in the 1990s. Although this is far fewer than initially promised by advocates of the technique, it is probably larger than the number of drugs whose origins can be traced directly to

Table 1. Docking hit rates against GPCRs

Target	Hit rate (actives/tested)	New scaffolds	Best affinity
β_2AR [32]	24% (6/25)	2	9 nM
A2a Aden [36]	41% (23/56)	4	32 nM
A2a Aden [31]	35% (7/20)	2	200 nM
D3 X-ray [31]	20% (5/25)	2	300 nM
D3 homology model [35]	23% (6/26)	2	200 nM
H1 histamine [45]	73% (19/26)	10+	6 nM

HTS [6,26], the dominant technique for new ligand discovery in pharmaceutical research, and has contributed especially to drugs for new targets. Protein structures have contributed in two ways to drug development: guiding the optimization of lead candidates and enabling the discovery of new chemical series, the latter using molecular docking and related techniques. Whereas structure has arguably had the greater impact on lead optimization in pharmaceutical research, it remains too early to evaluate the impact the new GPCR structures have had on this area, because most of these efforts remain closely held. Conversely, the impact of the new GPCR structures on docking screens has been immediate, with active molecules not only returned with high hit rates but also characterized by substantial novelty, as reflected by the new chemical scaffolds discovered, and potency against each of the four GPCRs targeted thus far in the literature (Table 1, Figure 1). In the case of the new β_2AR ligands, not only were several of the new hits potent but also the most active of them is among the most effective inverse agonist ever found for the receptor, depressing basal activity even more than the standard in the field (ICI118551).

Structure-based docking begins with a protein structure, either determined experimentally or modeled by homology, a large library of available small organic molecules, and a docking program that fits the two together. The insistence on screening available, typically purchasable compounds, ensures rapid cycles of modeling and experimental testing. For many years the molecular databases and docking programs remained the domain of specialists, but for the past several years databases of over 10 million, commercially available small molecules, such as ZINC (zinc.docking.org) [27], have been freely available,

Box 1. Key steps in structure-based docking screens against GPCRs

- Widely used docking programs include AutoDock [47], ICM [48], FlexX [49], Glide [50], and DOCK [51,52]. A web-based, automated version of the last program is available at http://blaster.docking.org
- A library of small molecules to dock against the GPCR. Over 10 million commercially available molecules, with structures, charges, molecular properties, and links to suppliers is available for free at http://zinc.docking.org
- Defining the docking site. Optimization of the structure for residue conformations, hydrogen positions, inclusion or exclusion of ordered water molecules. For physics-based programs, calculation of molecular potential functions for the target site.
- Control calculations to establish that known ligands rank higher than decoy molecules [53]. Such controls are easy to undertake and guide the optimization of parameters for new compounds. Known ligands may often be found from ChEMBL (www.ebi.ac.uk/chembl/).
- Screen the library against the binding site. We often choose a subset of compounds with 'lead-like' properties [54] to dock; these molecules are smaller and often better behaved than the larger 'drug-like' molecules, and allow for optimization. Over 3 million are commercially available.
- From the top 0.01% to 0.1% of the docking ranked list, pick 25 to 50 to test. This is done by eye, selecting for or against features that the docking scoring function might under- or overweight. This includes complementation of ligand polar groups by the protein, picking diverse chemotypes, balancing between polar and non-polar complementarity.
- Careful assays to ensure that any 'hits' that emerge are well behaved, e.g. are reversible, with good dose–response curves, compete with a classical ligand, if available, and do not form colloidal aggregates [55,56].

as have web-based docking applications (blaster.docking.org) (Box 1), lowering the barriers to entry.

Docking results against the new GPCR structures are given context by comparison to those against soluble proteins. In unbiased screens for new ligands against such soluble protein targets, docking hit rates have been approximately 5–10% of the molecules tested, with a hit typically defined as a molecule with micromolar to mid-micromolar affinity. Although two to three logs better than typical HTS hit rates [28–30], the docking hit rates and the affinity of the hits against soluble proteins are two to three log orders worse again than those against GPCRs to date. Two factors may contribute to this unusual success rate
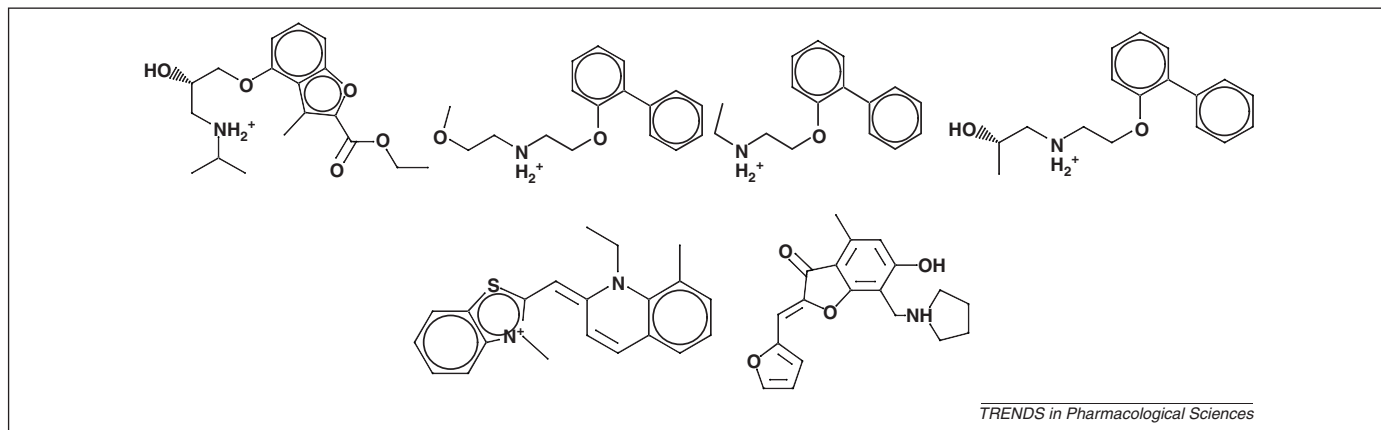

Figure 1. New inverse agonists against the β_2AR structure discovered by molecular docking. Top row: four relatively potent ligands that resemble known chemotypes, with K_i values from 9 nM to 2 μM. Bottom row: two novel chemotypes, with K_i values of 1.1 and 3.3 μM.

against GPCRs. First, even supposedly unbiased libraries, such as ZINC, in fact contain a disproportionate number of molecules that resemble GPCR ligands [31], partly owing to the longstanding interest in these receptors among medicinal chemists. Second, the well-buried GPCR orthosteric sites can almost entirely sequester and complement a small organic molecule, allowing them to be recognized with high ligand efficiency (potency/size). By contrast, for many soluble proteins, there is little certainty that the 'right' molecules exist to be found in our libraries [31,32] and their binding sites are often larger, flatter, and more exposed to solvent than the orthosteric sites of those GPCRs determined to date.

Expanding structural coverage with homology models

As of this writing, we are aware of 12 GPCRs whose structures have been determined, including published and unpublished structures. If this is a great expansion from only 4 years ago, it still represents only 4% of the pharmacologically relevant GPCRs, considering that there are approximately 300 nonolfactory GPCRs in the human genome [3]. Even with the current pace of structure determination, for the foreseeable future there are likely to be many fewer structures than good targets for structure-based discovery. For many, the target of particular interest is likely to lack an experimental structure, and there will be a temptation to model a structure based on homology to a structurally determined template. Will a model-built structure lead to a successful screen for new ligands?

This question was taken up in a recent prospective comparison of the dopamine D3 receptor. The community had been challenged to predict the D3 dopamine receptor/eticlopride complex, as well as complexes involving the CXCR4 receptor, before their X-ray structure was released [33,34]. This competition provided an opportunity to prospectively compare docking against a homology model to that against the X-ray structure of the same receptor, once released. As in earlier docking screens against the β2 adrenergic and the A2a adenosine receptors, the screen against the D3 receptor had a high hit rate, 23%, with affinities as good as 200 nM. To our surprise, the docking screen against the homology model, whose predictions were tested before the X-ray structure was released, had hit rates, affinities, and novel structures as good as those against the experimental structure (Table 1) [35]. This suggests that targets that have sequence identities of 35% or greater in the trans-membrane region to GPCRs of known structure may be well enough modeled to allow reliable docking screens, thereby increasing the impact of experimentally determined structures.

Efficacy

State-specific structures

All of the initial GPCR structures captured the receptors in inactive states, and *in silico* screens against the inactive state structure of the β2AR [32], the A2a adenosine [31,36], and the D3 dopamine receptor [35] yielded only antagonists and inverse agonists. Whereas retrospective modeling suggested that modest manipulation of the inactive states led to recognition of agonists in docking studies [37,38], prospective screens – where new molecules were

tested – returned no agonists whatsoever. Thus, there has been great interest in determining active state structures, but these have been more challenging because, for many GPCRs, agonists alone do not fully stabilize the active state [39].

A common property of many GPCRs is that G protein coupling enhances agonist binding affinity through an allosteric mechanism. For example, the affinity of the β2AR for the agonist isoproterenol is 100-fold greater when coupled to the G protein Gs [40,41]. As such, stabilization of the cytoplasmic domains by a G protein (or a G protein mimetic binding protein) may be necessary for obtaining the structure of an active state binding pocket. This was the case for the β2AR, where even a covalent agonist failed to fully stabilize an active conformation [39]. The first active state structure of the β2AR required stabilization with a G protein mimetic camelid single chain antibody fragment, also known as a nanobody (Nb80) [42]. Similar to the G protein Gs, Nb80 bound to the β2AR in an agonist-dependent manner and stabilized the same high affinity state for the agonist isoproterenol. The structure of the binding pocket of the β2AR–Nb80 complex is indistinguishable from that observed for the recent β2AR–Gs crystal structure [43]. A comparison of active and inactive state binding pockets for the β2AR reveals relatively small changes when individual amino acids are considered; however, the overall volume of the active state binding pocket is reduced relative to the inactive state (Figure 2). It will be fascinating to learn whether docking hits against this activated state will be dominated by agonists as docking against the inactive state was dominated by antagonists. This would suggest that the state of the receptor structure strongly influences the efficacies of the ligands that emerge from the docking screens, with antagonists dominating screens against the inactivated states. Preliminary studies from our laboratories hint that this may be the case.
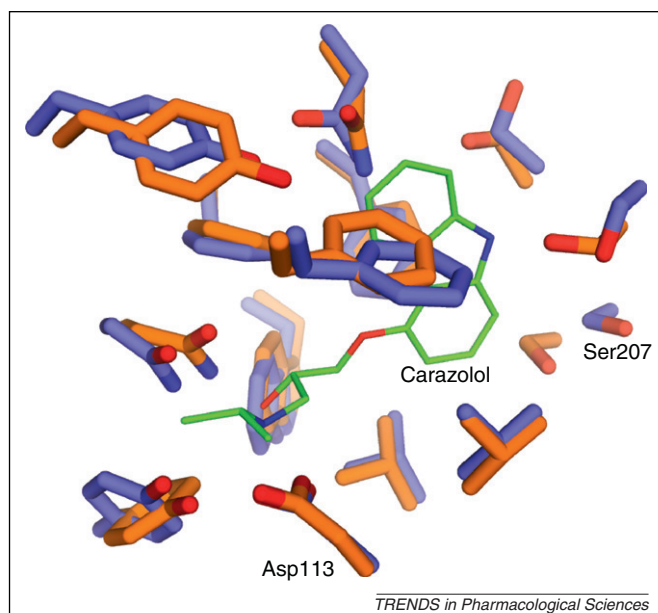

Figure 2. Comparison of the binding pocket of the β2AR in the inactive (carbons in blue) and active (carbons in orange) conformations. The inverse agonist carazolol is shown with carbons in green. The largest difference is observed around Ser207 in transmembrane segment (TM) 5.
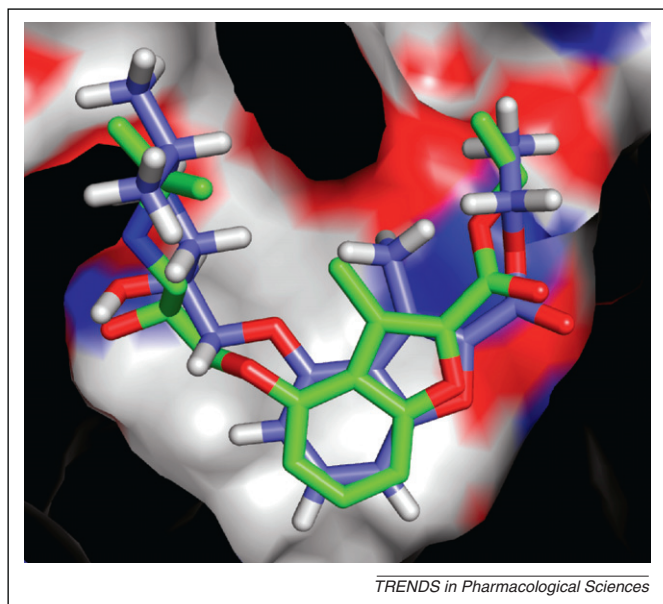
Figure 3. A 9 nM inverse agonist of β_2AR discovered by docking: docked orientation (carbons in blue) superposed on the subsequent X-ray crystallographic result (carbons in green) [46].

As noted above, there is evidence that GPCRs have more than one active state. Several GPCRs couple to more than one G protein and signal through G protein-independent pathways in a ligand-specific manner. It is probable that the differences in the orthosteric site responsible for this pathway selective signaling are more subtle than what is observed between inactive and actives states of the β_2AR. These small structural differences may not be captured by crystallography. Although this may limit the application of *in silico* screening, it is also possible that the evolution of initial hits through iterative rounds of modifications to improve selectivity and specificity may produce pathway selective ligands. Finally, it is conceivable that determination of the structure of a GPCR in complex with a non-G protein partner, such as β-arrestin, may illuminate still other conformations of the receptor that may be directly exploited for pathway-specific ligand discovery.

Concluding remarks

When the field began docking against the new GPCR structures, it was uncertain whether the technique would find potent and novel chemical matter at all. What we have learned over the past 3 years is that GPCRs are unusually well suited to docking screens, returning molecules whose affinities and hit rates are several logs better than we have come to expect in docking against soluble proteins. The campaigns to date have focused on technical features of hit rates, affinities, compound novelty, and the use of models, rather than the discovery of chemical probes and leads as ends in themselves. As the field matures, discovering such chemical matter will become more important. The new GPCR structures open avenues to molecules biased to particular signaling pathways and to molecules specific for particular receptors; such bias and specificity in turn enable the exploration of new biology and new therapies. Exploiting this opportunity will demand not only the collaborations between structural biologists, computational

chemists, and molecular pharmacologists but also animal pharmacologists, medicinal chemists, and signaling biologists. Such close engagement has perhaps been underemphasized in the era of molecular pharmacology, contributing to the lower than expected rate of drugs that exploit the target-based approach. With targets as rich in biology and disease as GPCRs, we can re-imagine a pharmacology that brings to bear the structural tools of the modern era with the integrated approach of the classical period [5]. For structural biology to have the greatest impact drug discovery, it will be necessary to have iterative cycles of docking, verification of ligand–receptor interactions by structure determination, and lead optimization (Figure 3). This is not currently practicable for most GPCR targets where the quality of the structures is often dependent on the affinity of the bound ligand, and the time required to acquire structures is often longer than needed for the highly competitive pace of most drug discovery efforts. Nevertheless, given the remarkable progress over the past several years and the flood of new structures, structural biology will probably become an integral part of many new drug discovery efforts [44].

References

1 Dixon, R.A. *et al.* (1986) Cloning of the gene and cDNA for mammalian beta-adrenergic receptor and homology with rhodopsin. *Nature* 321, 75–79
2 Kubo, T. *et al.* (1986) Cloning, sequencing and expression of complementary DNA encoding the muscarinic acetylcholine receptor. *Nature* 323, 411–416
3 Fredriksson, R. *et al.* (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* 63, 1256–1272
4 Gribbon, P. and Sewing, A. (2005) High-throughput drug discovery: what can we expect from HTS? *Drug Discov. Today* 10, 17–22
5 Keiser, M.J. *et al.* (2010) The chemical basis of pharmacology. *Biochemistry* 49, 10267–10276
6 Gleeson, M.P. *et al.* (2011) Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat. Rev. Drug Discov.* 10, 197–208
7 Rajagopal, S. *et al.* (2010) Teaching old receptors new tricks: biasing seven-transmembrane receptors. *Nat. Rev. Drug Discov.* 9, 373–386
8 Kenakin, T. (2003) Predicting therapeutic value in the lead optimization phase of drug discovery. *Nat. Rev. Drug Discov.* 2, 429–438
9 Paul, S.M. *et al.* (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* 9, 203–214
10 Arrowsmith, J. (2012) A decade of change. *Nat. Rev. Drug Discov.* 11, 17–18
11 Rasmussen, S.G. *et al.* (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 450, 383–387
12 Warne, T. *et al.* (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* 454, 486–491
13 Jaakola, V.P. *et al.* (2008) The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322, 1211–1217
14 Chien, E.Y. *et al.* (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330, 1091–1095
15 Wu, B. *et al.* (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330, 1066–1071
16 Shimamura, T. *et al.* (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475, 65–70

17 Hanson, M.A. *et al.* (2012) Crystal structure of a lipid G protein-coupled receptor. *Science* 335, 851–855

18 Haga, K. *et al.* (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482, 547–551

19 Kruse, A.C. *et al.* (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482, 552–556

20 Manglik, A. *et al.* (2012) Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. *Nature*, doi:10.1038/nature10954

21 Cherezov, V. *et al.* (2004) A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr. D: Biol. Crystallogr.* 60, 1795–1807

22 Caffrey, M. (2009) Crystallizing membrane proteins for structure determination: use of lipidic mesophases. *Annu. Rev. Biophys.* 38, 29–51

23 Serrano-Vega, M.J. *et al.* (2008) Conformational thermostabilization of the beta 1-adrenergic receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. U.S.A.* 105, 877–882

24 Rahuel, J. *et al.* (2000) Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. *Chem. Biol.* 7, 493–504

25 Lin, C. *et al.* (2006) Discovery and development of VX-950, a novel, covalent, and reversible inhibitor of hepatitis C virus NS3.4A serine protease. *Infect. Disord. Drug Targets* 6, 3–16

26 Macarron, R. *et al.* (2011) Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* 10, 188–195

27 Irwin, J.J. and Shoichet, B.K. (2005) ZINC – a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182

28 Doman, T.N. *et al.* (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* 45, 2213–2221

29 Babaoglu, K. *et al.* (2008) Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase. *J. Med. Chem.* 51, 2502–2511

30 Ferreira, R.S. *et al.* (2010) Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. *J. Med. Chem.* 53, 4891–4905

31 Carlsson, J. *et al.* (2010) Structure-based discovery of A2A adenosine receptor ligands. *J. Med. Chem.* 53, 3748–3755

32 Kolb, P. *et al.* (2009) Structure-based discovery of beta2-adrenergic receptor ligands. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6843–6848

33 Kufareva, I. *et al.* (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* 19, 1108–1126

34 Lam, A.R. *et al.* (2011) Importance of receptor flexibility in binding of cyclam compounds to the chemokine receptor CXCR4. *J. Chem. Inf. Model.* 51, 139–147

35 Carlsson, J. *et al.* (2011) Ligand discovery from a dopamine D3 receptor homology model and crystal structure. *Nat. Chem. Biol.* 7, 769–778

36 Katritch, V. *et al.* (2010) Structure-based discovery of novel chemotypes for adenosine A(2A) receptor antagonists. *J. Med. Chem.* 53, 1799–1809

37 Reynolds, K.A. *et al.* (2009) Identifying conformational changes of the beta(2) adrenoceptor that enable accurate prediction of ligand/receptor interactions and screening for GPCR modulators. *J. Comput. Aided Mol. Des.* 23, 273–288

38 de Graaf, C. and Rognan, D. (2008) Selective structure-based virtual screening for full and partial agonists of the beta2 adrenergic receptor. *J. Med. Chem.* 51, 4978–4985

39 Rosenbaum, D.M. *et al.* (2011) Structure and function of an irreversible agonist-beta(2) adrenoceptor complex. *Nature* 469, 236–240

40 Whorton, M.R. *et al.* (2007) A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7682–7687

41 De Lean, A. *et al.* (1980) A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled beta-adrenergic receptor. *J. Biol. Chem.* 255, 7108–7117

42 Rasmussen, S.G. *et al.* (2011) Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* 469, 175–180

43 Rasmussen, S.G. *et al.* (2011) Crystal structure of the beta2 adrenergic receptor–Gs protein complex. *Nature* 477, 549–555

44 Congreve, M. *et al.* (2011) The use of GPCR structures in drug design. *Adv. Pharmacol.* 62, 1–36

45 de Graaf, C. *et al.* (2011) Crystal structure-based virtual screening for fragment-like ligands of the human histamine H(1) receptor. *J. Med. Chem.* 54, 8195–8206

46 Wacker, D. *et al.* (2010) Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J. Am. Chem. Soc.* 132, 11443–11445

47 Morris, G.M. *et al.* (2008) Using AutoDock for ligand-receptor docking. *Curr. Protoc. Bioinform.* Chapter 8, Unit 8.14.1–40

48 Fernandez-Recio, J. *et al.* (2003) ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* 52, 113–117

49 Kamper, A. *et al.* (2006) Fully automated flexible docking of ligands into flexible synthetic receptors using forward and inverse docking strategies. *J. Chem. Inf. Model.* 46, 903–911

50 Repasky, M.P. *et al.* (2007) Flexible ligand docking with Glide. *Curr. Protoc. Bioinform.* Chapter 8, Unit 8.12.1–36

51 Mysinger, M.M. and Shoichet, B.K. (2010) Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* 50, 1561–1573

52 Irwin, J.J. *et al.* (2009) Automated docking screens: a feasibility study. *J. Med. Chem.* 52, 5712–5720

53 Huang, N. *et al.* (2006) Benchmarking sets for molecular docking. *J. Med. Chem.* 49, 6789–6801

54 Oprea, T.I. (2002) Current trends in lead discovery: are we looking for the appropriate properties? *Mol. Divers.* 5, 199–208

55 McGovern, S.L. *et al.* (2002) A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* 45, 1712–1722

56 Coan, K.E. *et al.* (2009) Promiscuous aggregate-based inhibitors promote enzyme unfolding. *J. Med. Chem.* 52, 2067–2075