Customizing G Protein-Coupled Receptor Models for Structure-Based Virtual Screening

Chris de Graaf[1,2,]* and Didier Rognan[1]

[1]*Structural Chemogenomics group, Laboratory of Therapeutic Innovation, UMR7200 CNRS-UdS (Université de Strasbourg), F-67400 Illkirch, France;* [2]*Current address: Division of Medicinal Chemistry, Faculty of Sciences, LACDR, Vrije Universiteit Amsterdam, The Netherlands*

Abstract: This review will focus on the construction, refinement, and validation of G Protein-coupled receptor models for the purpose of structure-based virtual screening. Practical tips and tricks derived from concrete modeling and virtual screening exercises to overcome the problems and pitfalls associated with the different steps of the receptor modeling workflow will be presented. These examples will not only include rhodopsin-like (class A), but also secretine-like (class B), and glutamate-like (class C) receptors. In addition, the review will present a careful comparative analysis of current crystal structures and their implication on homology modeling. The following themes will be discussed: i) the use of experimental anchors in guiding the modeling procedure; ii) amino acid sequence alignments; iii) ligand binding mode accommodation and binding cavity expansion; iv) proline-induced kinks in transmembrane helices; v) binding mode prediction and virtual screening by receptor-ligand interaction fingerprint scoring; vi) extracellular loop modeling; vii) virtual filtering schemes. Finally, an overview of several successful structure-based screening shows that receptor models, despite structural inaccuracies, can be efficiently used to find novel ligands.

Keywords: GPCR, modeling, virtual screening, drug design.

1. THE USE OF RECEPTOR MODELS IN STRUCTURE-BASED VIRTUAL SCREENING FOR GPCR LIGANDS

G-protein-coupled receptors (GPCRs) constitute a large family of heterogeneous membrane receptors characterized by a typical heptahelical membrane-spanning fold usually described as a seven-transmembrane (TM) domain [1, 2]. A striking feature of this protein family is the tremendous chemical diversity of possible ligands including light, small molecular-weight ions (e.g. glutamate, Ca^{2+}), biogenic amines (e.g. dopamine, serotonin), nucleosides and nucleotides (e.g. adenosine, adenosine triphosphate), peptide and protein hormones (e.g. chemokines, glucagon), lipids and eicosanoids (e.g. sphingolipids, prostaglandins) [3]. Activation of GPCRs upon binding of the above-mentioned ligands induces a conformational change of the receptor, thereby triggering a specific interaction with intracellular G proteins and subsequent activation/ inhibition of secondary messengers [4]. Because of the ubiquitous distribution of GPCRs at the surface of many cells, these receptors are regulating a wide array of physiological and pathological processes. As a consequence, GPCRs are particularly attractive targets for therapeutical intervention. Hence, about 30% of top-selling drugs modulate the activity of this family of receptors [3]. Up to now, few GPCRs (ca. 40) have been targeted by existing drugs. Analyzing human genomic sequences suggests the existence of about 400 nonolfactory GPCRs [5] and opens a new avenue for drug discovery, especially with respect to the 100 orphan receptors for which even the endogenous ligand still has not been characterized [6].

Knowledge of the three-dimensional structure of GPCRs can provide important insights into receptor function and receptor-ligand interactions, and can be used for the discovery of new drugs. So far structural modeling of GPCRs has been limited to either *ab initio* models [7, 8] or bovine rhodopsin (bRho)-based [9] homology models. Recently, crystal structures of squid rhodopsin, the beta adrenergic receptors type 1 (ADRB1) and 2 (ADRB2), the A2A adenosine receptor (AA2AR), and the ligand-free opsin (Ops*) were solved [10-16]. Although most of these crystal structures are relatively homologous in the trans-membrane (TM) binding cavity, the differences might however be large enough to influence the outcome of structure-based virtual screening [17]. In fact, the binding mode of the antagonist co-crystallized with AA2AR [12] deviates dramatically from previous computer models [18, 19], while ligand binding modes in earlier ADRB2 models (e.g., [20, 21]) are generally in agreement with the inverse agonist and antagonist-bound crystal structures of this receptor [10, 11]. Furthermore the structural divergence between the GPCR crystal structures appears to be much larger in the intracellular regions as well as the extra-cellular loop regions of the receptors [10, 12]. Comparing the inactive bRho to the active Ops* structure suggests ligand-induced conformational changes [13, 14], while comparison of different ligand-bound ADRB1 and ADRB2 structures show small subtle rotamer changes to optimize receptor-ligand interactions [11, 16]. Careful and critical receptor modelling strategies, guided and validated by experimental data, are

*Address correspondence to this author at the Structural Chemogenomics group, Laboratory of Therapeutic Innovation, UMR7200 CNRS-UdS (Université de Strasbourg), F-67400 Illkirch, France;
E-mail: ?????????????

therefore still needed to construct receptor models which can be used for structure-based virtual screening.

It has been stated before that the ground state of ADRB2 and bRho crystal structures are only suitable for discovering inverse agonists and antagonists [20, 22]. Recent virtual screening studies have shown however, that *agonists* can be retrieved by using receptor models based on inactive crystal structure templates [23-27], and, vice versa, *antagonists* have been found using *agonist*-biased receptor models [24, 28, 29]. Moreover, a recent study showed that it is even feasible to model early intermediate states in agonist binding, based on the inverse agonist bound crystal structure of ADRB2 [30]. Thus the appearance of more GPCR crystal structures can help to push computer-aided drug design studies towards more challenging predictions.

Although insights into the molecular *mechanisms* of ligand-induced receptor activation derived from computational studies [31-33] can be valuable for the design of specific ligand types (i.e. agonist vs. antagonists), the current paper will focus on the use of three-dimensional receptor models for predicting ligand binding, and enabling structure-based virtual screening studies in particular.

2. A MODELING WORKFLOW FOR TAILOR-MADE GPCR MODELS

Considering the apparent flexibility of GPCRs and the diversity in ligand types, binding pockets, and binding modes, GPCR models should be tailor-made, using all possible experimental data to guide their construction and validation along different modelling steps. Furthermore it should be acknowledged that the quality and applicability of GPCR models strongly depends on the amount of experimental data available to construct and validate them, more than on the exact modelling programs and techniques *via* which they are constructed. In Fig. (**1**) a typical GPCR modeling workflow is presented to construct GPCR receptor models.

In short, the GPCR modeling workflow follows seven steps: 1) Construction of the initial receptor model, consisting of the 7 transmembrane (TM) helices only, after identification of the TM helices (template-independent de novo modeling), amino acid sequence alignment between target and template receptors (homology modeling), and by rotation/de novo modeling of certain TM helices with putative alternative helical kinks; 2) Construction of a preliminary TM-ligand complex; 3) Energy minimization (EM) of the TM-ligand complex; 4) Molecular dynamics (MD) simulation refinement of the receptor-ligand complex; 5) Modeling the loops connecting the TM helices; 6) Selection and refinement of the full receptor-ligand complex; 7) Validation of the full receptor-ligand complex.

Depending on the availability of experimental data and the purpose of the receptor model, alternative workflows can be followed. For example, if there are no experimental data available to guide the construction of extracellular loops, it is
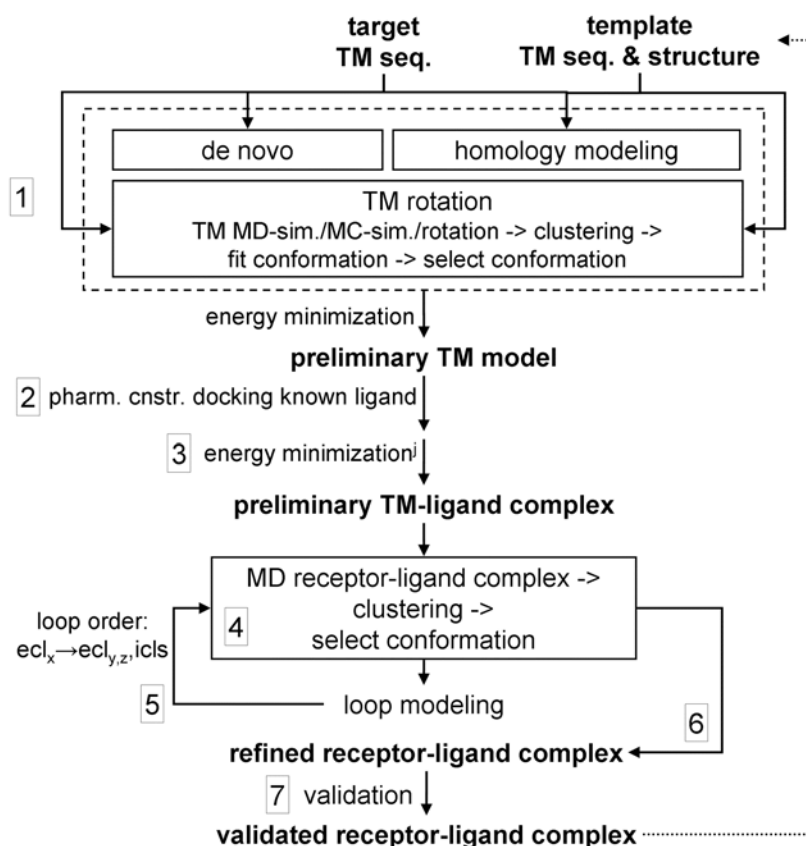

Fig. (1). GPCR modeling workflow along the different steps (boxed numbers) described in the text.

advised to omit step **5** (see section **4.5**). At the moment that a receptor model has been thoroughly evaluated and validated, it can serve as modeling template for constructing homologous targets [34]. For example, a validated cotico-tropin-releasing factor type 1 receptor (CRFR1) model is likely to be a better template to model future secretin-like (class B) GPCR models than less homologous bRho, AA2AR, or ADRB2 structure tem-plates. The selection of a proper template at the start of the modeling process (step **1** in Fig (**1**)) can be crucial and should be done with care, based on a preliminary analysis of the ligand binding pocket topography and knowledge of the structural differences between the different available modeling templates. Therefore a thorough analysis of the different GPCR crystal structures will be presented in the following section **3**. The different GPCR modeling steps will then be described in more details in section **4**.

3. Implications of GPCR Crystal Structures on GPCR Homology Modeling

With the exception of *de novo* GPCR models [7, 8] which are template-independent, most GPCR models have been derived from an experimental structural template (or from another receptor model initially derived from an experimental template). For many years, the crystal structure of dark-state retinal-bound bovine rhodopsin (bRho) [9] has served as a template to model GPCRs. In the past years, other ligand-bound bRho structures of higher resolution[35, 36] as well as different photointermediates [37, 38] of generally lower resolution have appeared. These structures only showed significant structural changes in the intracellular loops compared to the first bRho structure. The recently solved structure of the squid rhodopsin, on the contrary shows significant changes in the TM backbone compared to bRho, especially in TM2, as the result of a $G^{2.59}XP^{2.61}$ induced kink [15] (residues are numbered according to the Ballesteros-Weinstein nomenclature [39]). This motif is not present, however, in other non-olfactive human GPCRs. The large structural differences between the inactive bovine rhodopsin structures and the recently published ligand-free bovine opsin structures are primarily located in the intracellular regions, but are linked to structural rearrangements in the extracellular TM binding pocket as well [13, 14]. Around the retinal binding pocket, inward movements (towards the retinal binding pocket) of TM helices 2, 6, and 7 and an outward movement of TM3 (away from the binding pocket) are observed in opsin compared to rhodopsin [13, 14]. Only recently were the first human GPCRs crystal-lized with a diffusible ligand (ADRB1[16], ADRB2[10, 11], AA2A [12]). Although the relative orientations of some TM helices differ from bRho (Fig. (**2A**)), the main structural differences around the ligand binding pocket stem from the extracellular loops (ecls), especially ecl2 (Fig. (**2B,C**)). Furthermore, the exact location of the ligand binding pocket varies between the bRho, ADRB1/ADRB2, and AA2AR crystal structures (Fig. (**3**)). Site-directed mutagenesis mapping of the ligand binding cavities of many other receptors also suggest a wide variety of binding site locations among GPCRs, and often also depending on the ligand itself (i.e. agonists vs. antagonists),

although there seems to be at least a partial overlap between these sites (Fig. (**3**)).

Fig. (**2**) shows the structural divergence between GPCR crystal structures. In Fig (**2A**), the crystal structures are aligned on the C alpha atoms of 33 residues proposed to represent a consensus GPCR binding cavity [40], but for the bRho-ADRB2 alignment the positions of TM1 are not taken into account because of the large divergence in that region [10]. Compared to bRho, TM1 of ADRB2 and TM2 and 3 of AA2AR are highly divergent, while TM 6 and 7 of ADRB2 and TM6 of AA2AR are similar. TM2 and 4 of ADRB2 and TM1, 4, and 7 of AA2AR are moderately constant. TM 3 and 5 of ADRB2 and TM 5 of AA2AR are moderately divergent. The TM helices in bRho and all other GPCR crystal structures are far from ideal helices. Most of these deformations are induced by unusual proline-induced kinks stabilized by adjacent glycine and/or threonine/serine residues, resulting in some cases in local openings or closures of specific helical turns [41-43].

Thus, while the overall fold of most helices in the published GPCR crystal structures is relatively conserved (suggesting structural mimicry among GPCRs [43]) there are striking differences in some helical bends (described above) and relative orientations of some helices (Fig. (**2**)). These differences can often be explained by specific sequence motifs (P/G/T/S) and/or TM-TM interactions (e.g., the conserved H-bond between $N^{1.50}$ side chain and the backbone carbonyl oxygen of residue 7.46) inducing and/or stabilizing kinks in the TM helices. Interestingly, different sequence motifs can induce the same kink (e.g., $G^{2.56}G^{2.57}XT^{2.59}T^{2.60}$ in bRho and $P^{2.59}XG^{2.61}$ in ADRB2) and in some cases, the orientation of TM helices seems to be affected by the conformation of extracellular loops as well. The top of TM3 in AA2AR, for example, bends further towards TM4 (Fig. (**2A**)) by the two disulfide bridges between TM3 and ecl2 (Fig. (**2B**)). In other cases, the fold around the TM cavity (i.e., the positions facing the ligand binding pocket) is often conserved despite differences in helical kinks (e.g. TM7). Although this makes the prediction of TM kinks based only on sequence motifs difficult [41], it should be stressed that the selection of the proper modeling template and/or modeling of alternative TM kinks can also be guided by experimental data supporting and/or excluding the involvement of specific residues at the top of the helix.

Furthermore, the helical bundle of GPCRs is believed to be rather flexible, and it has been proposed that receptor activation is accompanied by rotation and/or translation of certain TM helices (notably TM6) to facilitate binding of the G-protein [32, 44]. In very recently solved crystal structures of opsin in the activated state (Ops*) [13, 14], of which one is bound to a peptide segment of the Gα-protein [14], several conformational changes in the TM helices are obser-ved compared to the bRho ground-state crystal structure [9]. The largest structural changes are located at the intracellular side where the G-protein binds, breaks the ionic lock [9] between $R^{3.50}$ and $E^{6.30}$ and relocate the intracellular regions of especially TM5 and TM6, as well as TM3 and TM7. This movement is however linked to structural rearrangements in the TM cavity at the extracellular side: i) TM5 and TM6 move down (towards the intracellular side) and towards each
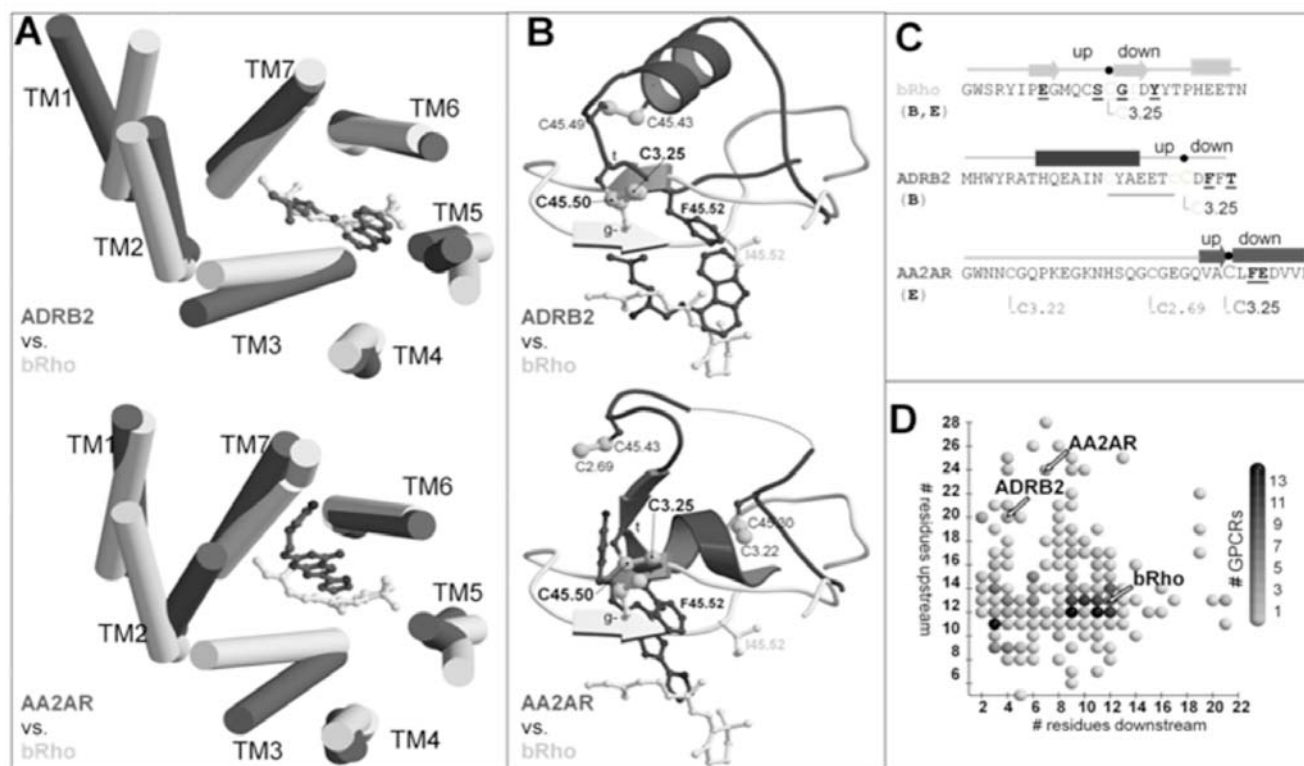
Fig. (2). Structural divergence between GPCR crystal structures. A) Comparison of the 7TM-bundle of bRho (light grey cylinders) [36] with that of ADRB2 (top, dark grey cylinders) [10] and AA2AR (down, dark grey cylinders) [12] . Ligands (retinal in bRho (light grey carbon atoms), carazolol in ADRB2 (dark grey), and ZM241385 in AA2AR (dark grey) are depicted by ball-and-sticks. B) Comparison of the extracellular loops 2 (ecl2) of bRho (light grey), ADRB2 (top, dark grey), and AA2AR (down, dark grey). Ligands are presented as in panel A. Disulfide forming cysteine residues as well as residues at position 45.52 are depicted by ball-and-sticks. C) Comparison of the ecl2 amino acid sequences of bRho, ADRB2, and AA2AR, along with their secondary structure (beta strands depicted by arrows, helical segments depicted by bars), disulfide bridges (grey lines), and residues involved in receptor-ligand contacts (underlined in bold). D) The number of ecl2 residues downstream from C45.50 plotted against the number of residues upstream from C45.50 for 325 GPCRs containing the conserved C3.25-C45.50 disulfide link [46].

other, away from TM3 ii) the highly kinked TM7 helix bends towards TM2 and the relocated TM6; iii) the top of TM2 kinks towards TM3 which moves away from TM7. It has to be investigated, however, whether the opsin structure is a more suitable template for modeling the agonist bound state of other receptors. The Cβ to Cβ distances between positions 3.32 and 5.43, and between 3.32 and 6.52, residues experimentally found to be involved in agonist binding in ADRB2 [45], are significantly longer in the Ops* than in the ADRB2 crystal structure (6 and 4 Å longer, respectively), making Ops* probably an unsuitable template to construct an agonist-bound ADRB2 model. Consideration of the relative orientation and helical kinks, determining the position and direction of residues into (or out of) the ligand binding pocket, in potential modeling templates is therefore a very important first step in the GPCR modeling process. And in cases where experimental data suggest that the TM conformation of the target is not captured by any of the available templates, it can be necessary to model this conformation (step 1 in the modeling scheme in Fig. (**1**)), as will be discussed in section **4.3**.

Fig. (**2B**) shows how the second extracellular loop of bRho is folded deep into the TM binding pocket forming a β

sheet between β strands upstream and downstream from $C^{45.50}$ (see [46] for sequence-independent numbering of ecl2 residues). The ecl2s of ADRB2 and AA2AR are located more towards the extracellular side of the helical bundle, containing a long helix upstream from C45.50, and a β strand (forming a β sheet with ecl1) upstream and a short helix downstream from C45.50, respectively. All three structures contain a conserved disulfide bridge between C45.50 in ecl2 and C3.25 in TM3. An additional disulfide bridge within ecl2 is present in ADRB2, while the ecl2 of AA2AR forms two additional disulfide bridges with ecl1. Analysis of the ecl2 of 325 GPCRs containing the conserved C3.25-C45.50 disulfide link shows that the upstream ecl2 loop length (showing an optimum at 12–13 residues) is less variable then the downstream ecl2 loop length (showing optimum between 3 and 12 residues) [46]. Compared to bRho, many GPCRs have approximately the same number of residues upstream and downstream, but other GPCRs have a somewhat shorter or longer (e.g. ADRB2, AA2AR) upstream ecl2 loop, and/or significantly shorter (ADRB2 and AA2AR) or longer downstream ecl2 loop. Long upstream combination with short downstream ecl2 loops, like ADRB2 and AA2R, are however relatively rare. Although bRho seems to be a relatively suitable modeling template for

modeling the upstream ecl2 segment, construction of ecl2 however should be done with care and guided by receptor-specific experimental data, rather than carried out in a high-throughput fashion and derived directly from the bRho crystal structure [46]. Moreover, loop-less TM models of GPCR receptors can be suitable targets for virtual screening as well ([8, 46-50]. The implication of extracellular loop modeling (step **5** in Fig. (**1**)) on structure-based virtual screening will be discussed in more detail in section **4.5**.

Experimental data such as site-directed mutagenesis (SDM) and ligand structure-activity relationships (SAR) data give insights into which residues in the receptor and which chemical groups in the ligand are involved in receptor-ligand binding. This information can be used to propose receptor-ligand binding modes and to construct *in silico* models of receptor-ligand complexes. In this way, the binding modes of inverse agonists and antagonists hypothesized in earlier ADRB2 homology models based on bRho generally resemble the binding orientations of carazolol and timolol in the ADRB2 crystal structures [20, 21, 51]. The binding mode of ZM241385 in the recently solved AA2AR crystal structure [12], however, deviates significantly from the antagonist binding poses proposed in molecular modeling studies [18, 19]. The apparent success in ADRB2 and failure in AA2AR modeling is mainly the result of differences in the character and location of experimental anchors to construct receptor-ligand interaction hypotheses.

ADRB2 ligands share an essential positively charged amine as well as an aromatic ring separated by ca. 5 (in most partial/full agonists) to 7 Å (in most inverse agonists/antagonists). This pharmacophore fits the binding pocket of the high-resolution ADRB2 crystal structure [10], in which both polar interactions (to $D^{3.32}$, $S^{5.42}$, $N^{7.39}$) and edge-to-face π stacking (mainly to $F^{6.52}$) contribute to strong directional constraints for receptor recognition. The binding mode of ADRB2-bound carazolol is in line with earlier site-directed mutagenesis studies, supporting the involvement of $D^{3.32}$ [52, 53], $S^{5.42}$ [54], and $N^{7.39}$ [55] in binding of both antagonist and agonists. Mutation of residues $S^{5.43}$, $S^{5.46}$, $N^{6.55}$, and $Y^{7.35}$ affects partial/full agonist binding [56-58]. In the case of ADRB2, three polar residues located in three different TM helices, namely $D^{3.32}$, $S^{5.42}$, and $N^{7.49}$, have been shown to be involved antagonist-ADRB2 interactions by SDM. The spatial distribution of these critical residues in a bRho-based ADRB2 homology model is quite close to that in the X-ray structure. Ligand-receptor binding modes are therefore relatively straightforward to predict by many docking tools (Fig. (**3**)).

Conversely, the recently described X-ray structure of AA2A in complex with the antagonist ZM241385 [12] was almost impossible to predict. Previous SDM studies had identified several polar anchoring residues, including $N^{6.55}$, to be important for agonists and xanthine antagonist bind-ing [59, 60]. Mutation of $Q^{3.37}$ had small but significant effects on binding of agonists, xanthine antagonists, and antagonists containing an adenine ring (like ZM241385) [60]. Further-more, mutation of ecl2 loop residues differently affected agonist and antagonist recognition [61]. Modeling the ecl2 of AA2AR based from bRho is however quite challenging because of the large difference in number of downstream and upstream residues (Fig. (**2C**)). In fact, only one of these negatively charged residues ($E^{45.53}$) interacts with ZM241385 in the AA2AR crystal structure. The plausible H-bond interaction hypo-thesis between the exocyclic amine and $N^{6.55}$ observed in the crystal structure is supported by the SDM data and was therefore proposed in many AA2AR receptor models as well [62]. Understandably, computational chemists have been tempted to match the other numerous polar H-bond acceptor and donor atoms in the ligand with H-bond interaction partners in the receptor, but surprisingly, three polar interactions with the ligand are mediated *via* water molecules. The possible role of conserved internal water molecules in structure and function of GPCRs has been hypothesized [63], but this is the first report of water-mediated receptor-ligand interactions. Furthermore, V3.32, F5.43, and H7.43, which are indicated by SDM data to be involved in antagonist binding [60, 64], are in fact not in close contact with ZM241385 in the AA2AR crystal struc-ture. The use of these experimental anchors to guide the construction of a ZM241385-AA2AR complex would therefore have been misleading. In conclusion, the lack of *directionality* in receptor-ligand interactions for guiding the modeling of the receptor-ligand complex explains the difficulty of modeling AA2AR-anta-gonist interactions compared to ADRB2-ligand interactions.

4. TIPS AND TRICKS TO CUSTOMIZE LIGAND-BIASED GPCR MODELS

One of the most important aspects of any GPCR modeling protocol is the incorporation of as much experi-mental data available to guide the construction of the receptor model at every step and as early in the modeling process as possible. Information on the involvement of certain residues in ligand recognition [65], the identifi-cation of essential chemical groups for ligand for binding, and the synthesis of these data into a preliminary ligand binding mode hypothesis can be used to: i) make rational decisions regarding the selection of the optimal modeling template and alignment of the target and template amino acid sequences (section **4.1**); ii) refine the receptor and shape the ligand binding pocket (section **4.2**); iii) model alternative helical kinks and orientations (section **4.3**); iv) select and score receptor-ligand docking poses (section **4.4**); v) model extracellular loops (section **4.5**); vi) set up smart virtual screening strategies (section **4.6**).

4.1. Rationalizing Sequence Alignments

The first steps in GPCR modeling are the identification of the transmembrane helices and amino acid sequence alignment between target and crystal structure (or refined receptor model) template(s) (step **1** in Fig. (**1**)). Several algorithms have been developed to predict the rough location of the 7 TM helices [66]. While *de novo* GPCR modeling methods, like Predict [8] and Membstruck [7], do not require target to template amino acid sequence alignments as input, homology modeling does require such an alignment. Despite the relatively low sequence identity between the TM domains of rhodopsin-like (class A) GPCRs [67], there are several GPCR family-specific patterns and motifs [68] which facilitate their alignment [20]. The Ballesteros-Weinstein
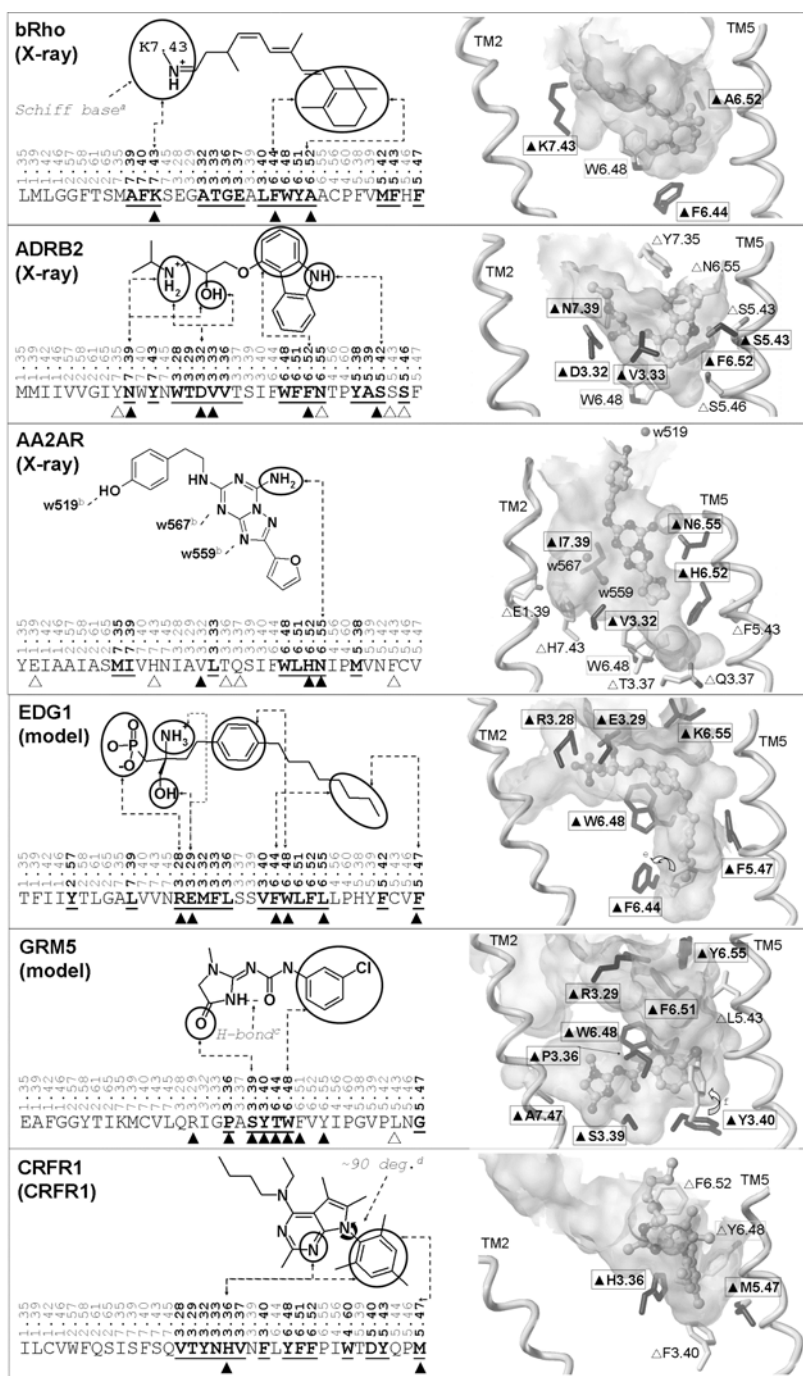
Fig. (3). Diversity of GPCR-ligand binding pocket and binding modes in six different GPCRs, (bRho, ADRB2, AA2AR, sphingosine 1-phosphate receptor 1 (S1PR1), metabotropic glutamate receptor 5 (GRM5), and corticotropin-releasing factor receptor 1 (CRFR1). Chemical groups in the ligands (retinal in bRho, S-carazolol in ADRB2, ZM241385 in AA2AR, S-FTY-720-P in S1PR1, fenobam in GRM5, and antalarmin in CRFR1) identified to be important for receptor binding by structure-activity relationship (SAR) studies (encircled) are linked to residues shown to be involved in receptor-ligand binding by site-directed mutagenesis (SDM) studies (indicated by a filled triangle if tested on the same chemotype, indicated by an empty triangle if tested on a different chemotype). Residues in contact with the ligand in the crystal structure (bRho, ADRB2, AA2AR) or receptor model (S1PR1, GRM5, CRFR1) are underlined in bold. The experimentally supported receptor-ligand binding modes can be used to guide the construction (steps 2-5 in Fig. (1)) as well as selection and validation of receptor models (steps 6-7) and are described in more details in the text. Notes: [a] A Schiff base is formed between K5.43 and retinal; [b] H-bond interactions between ZM241385 and water molecules in the binding pocket; [c] Internal H-bond is an essential feature of fenobam analogues; [d] orthogonal orientation of the substituted phenyl ring is an essential feature of CRFR1 negative allosteric modulators; [e] manual rotation of F6.44 rotamer in bRho-based EDG1 model; [f] different locations and rotameric states of Y3.40 in bRho-based (grey) vs. refined ADRB2-based (black) model.

numbering scheme [39] is based on the presence of several highly conserved residues among class A GPCRs: $N^{1.50}$ in TM 1 (1.30-1.59), $D^{2.50}$ in TM2 (2.38-2.67), $R^{3.50}$ in TM3 (3.22-3.54), $W^{4.50}$ in TM4 (4.40-4.62), $P^{5.50}$ in TM5 (5.35-5.60), $P^{6.50}$ in TM6 (6.30-6.55), and $P^{7.50}$ in TM7 (7.33-7.53). Previous sequence analysis and molecular modeling that incorporated experi-mental data such as creation of metal binding sites, disulfide cross-linking, and double revertant mutations, were successful in predicting helix ends and protein-protein interactions in bRho, which are also predicted for other class A GPCRs [43]. The GPCR crystal structures of bRho, ADRB2, and AA2AR share most of the positions facing the TM binding pocket [9, 10, 12], and most of these positions have also been identified to belong to the binding cavity by the substituted-cysteine accessibility method (SCAM) [43] or other SDM studies [40]. In fact, sequence similarity in TM cavity alignments can often be matched to conserved moieties in the ligands of these receptors [40, 69], and therefore can be used as a rational modeling template selection criterion (step **1** in Fig. (**1**)) instead of the overall TM sequence similarity.

While the alignment of most TM helices of class A GPCRs to the currently available GPCR crystal structure templates seems straightforward, there is still some debate on their alignment to class B and class C GPCRs. We have worked on class B as well as class C receptors and noticed that for some of the TM helices of these receptors alternative alignments to bRho have been reported [34, 70-78]. Our alignments and the experimental anchors used to support them are presented in Fig. (**4**). The alignment of class B GPCRs on the left in Fig (**4A**) is in agreement with inter-helical interactions displayed in the CRFR1 model on the right: i) a putative $R^{2.53}$-$Q^{7.45}$ H-bond proposed for the parathyroid type 1 receptor (PTHR1) based on correlated effects in SDM studies [79], and further supported by SDM studies of various other class B receptors [79, 80]; ii) a hydrogen bond network between TM2 ($R^{2.39}$, $H^{2.43}$), TM3 ($E^{3.46}$), and TM6 ($T^{6.37}$), which are all conserved among class B receptors shown to be important for either ligand binding [81] and/or functional activation [81-85]; iii) the relative locations of $H^{3.51}$, and the 6.30-6.33 helical segment of PTHR1 as derived from the alignment, agrees with Zn(II)-bridge linking studies [85]; iv) the formation of an H-bond between $W^{4.50}$ and a polar residue at position 2.45 (N/H/S) (both conserved residues in class A and B GPCRs [67]), which are observed in bRho, ADRB1, ADRB2, and AA2AR crystal structures [9, 10, 12, 16]. Mutation of $W^{4.50}$ in the growth hormone-releasing hor-mone receptor (GHRHR) greatly affects its stability [83]. Final arguments supporting the alignment of class B GPCRs to bRho concern the location of residues found to be involved in non-peptide binding as described in section **4.2** and illustrated in Fig. (**3**).

In Fig. (**4B**), residues in contact with ligands in the bRho [9] and ADRB2 [10, 11] (rhodopsin-like or class A GPCRs) crystal structures are marked and aligned with residues proposed to be involved in ligand binding based on SDM data in different class C GPCRs [72, 77, 84]). Site-directed mutagenesis studies suggest that the binding sites of Calhex 231 and NPS-2143 to the calcium-sensing receptor (CASR) overlap but are not identical, as predicted by the corres-ponding receptor model [72]. The class A to class C GPCR

amino acid sequence alignment in Fig. (**4B**) and the corres-ponding CASR model (as well the GRM5 model discussed later in section **4.2**) are not only supported by SDM studies probing the receptor-ligand binding pocket, but also agree with experimental data supporting inter-helical interactions in the cytoplasmic TM regions of class C GPCRs [82].

It should be noticed, however, that the same SDM data have often been used to propose and support different amino acid sequence alignments [72, 73] and/or ligand binding modes [86, 87]. A clear example of misinter-pretation of such experimental data is the deviation between the ligand binding mode in the recently published AA2AR crystal structure [12] and the earlier reported computational models [18, 19] as discussed in the previous section **3**. Furthermore, there are numerous studies in which the effects of site-directed mutations have been found to be ligand-dependent, discriminating not only between different ligand types (i.e. agonists vs. antagonists [56, 59]), but also between different chemotypes with the same pharmacological effect [72, 75, 88, 89]. Although such detailed information can be used to further refine and challenge receptor models, it should also be considered as a warning for over-interpretation of the receptor model by *retrospective* analysis.

The best one can do is to satisfy and use as many experimental constraints as possible to construct and validate the model (and not only those features which are line with the modelers preconceptions) at the earliest possible stage [90]. Moreover, receptor models should be applied and challenged by *prospective* predictions. In fact, we believe that the strength and added value of GPCR models especially lies not only in mapping ligand-binding sites but also in identifying new ligands from large chemical databases, as supported by the many prospective virtual screening success stories presented in section **5**.

Aligning the amino acid sequences of the extracellular loops of GPCRs is obviously less straightforward (step **5** in Fig. (**1**)) [46], especially given the larger structural diver-gence in these regions when comparing GPCR crystal structures (Fig (**2**)). However, also in this case it might be possible to use small loop segments of GPCR crystal structures for construction of the loop of the target, but only if this is supported by strong additional experi-mental data [46], as discussed in section **4.5**. Finally, even when the amino acid sequence alignment between target and template is established, alternative kinks in the TM helices should sometimes be considered (section **4.3**).

4.2. Customizing Ligand-Biased GPCR Cavities

Figure (**3**) shows how receptor-ligand binding mode hypotheses can be derived from and/or rationalized by ligand-based (SAR) and receptor-based (SDM) experi-mental data. Ligand binding modes in bRho, ADRB2, and AA2AR crystal structures have already been described in the context of SDM data in section **3**. Below will follow a description of the ligand binding orientations in models of S1PR1 (a lipid/class A receptor), GRM5 (a glutamate-like/class C receptor), and CRFR1 (secretin-like/class B receptor) and how experimental data are used to guide their construction and validation.
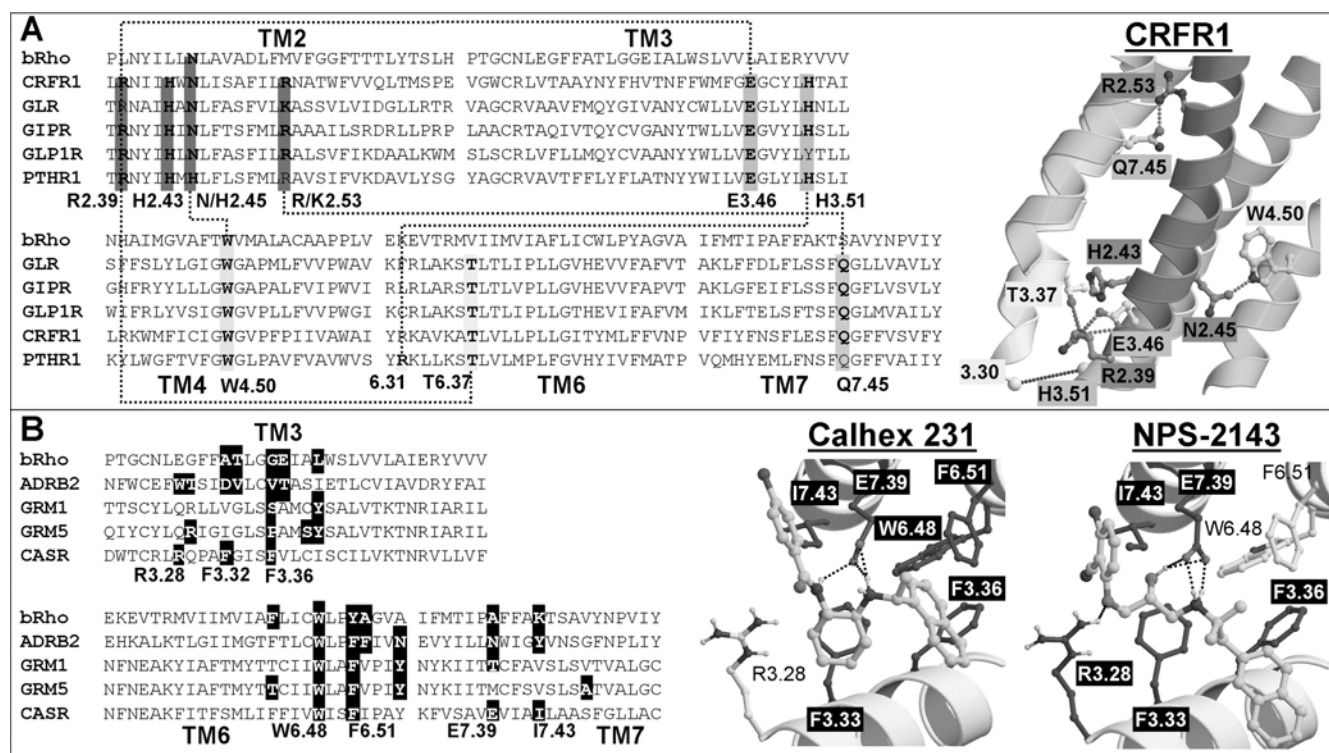
Fig. (4). The use of experimental anchors to guide GPCR TM sequence alignments (step 1 in Fig. (**1**)). A) Amino acid sequence alignment between the TM domains of class B receptors and bRho. The alignment on the left is in line with interhelical interactions displayed in the CRFR model on the right (see text). B) Amino acid sequence alignment between the TM domains of class C receptors, bRho, and ADRB2. Residues in contact with ligands in the bRho and ADRB2 crystal structures are marked and align with residues proposed to be involved in ligand binding based on SDM data (metabotropic glutamate receptors 1 (GRM1) and 5 (GRM5, see also Fig. (**3**)) the calcium-sensing receptor (CASR)). Carbon atoms of residues in the CASR pocket identified by SDM to be involved in binding of the specific ligand are coloured dark grey, residues which are not important for binding of the ligand are coloured light grey.

Essential features of S1PR1 ligands are negatively charged and positively charged groups in the polar "head" as well as a long hydrophobic "tail" [91]. SDM data suggest that the S1PR1 receptor binds its ligands through the conserved residues $R^{3.28}$ and $E^{3.29}$ [92]. The negatively charged phosphate oxygens of *S*-FTY-720-P [93] forms an ionic link to positively charged $R^{3.28}$ residue, while the positively charged protonated amine of the ligand makes a salt bridge to the carboxylate moiety of $E^{3.29}$. The hydroxyl group of *S*-FTY-720-P interacts with one of these carboxylate oxygens *via* another H-bond. The essential aromatic ring of *S*-FTY-720-P [91] stacks between the aromatic rings of $W^{6.48}$ and $F^{3.33}$, while the apolar alkyl chain of the ligand dives down into the relatively deep hydrophobic channel between transmembrane helices (TMs) 3, 4, 5, and 6 (subpocket i [40]). Many residues lining this hydrophobic channel (among which $F^{5.47}$, $F^{6.44}$, and $W^{6.48}$) have recently been identified to be involved in ligand binding and receptor activation in S1PR1 [94]. Residue $L^{6.55}$ at the top of TM6 makes a hydrophobic contact with the aromatic ring of *S*-FTY-720-P as well. Mutation of this residue into a phenyl-alanine only slightly affected the potency of *S*-FTY-720-P in S1PR1 [95].

Negative allosteric modulators of GRM5 are medium-sized molecules with terminal polar and apolar ring systems connected by a rigid straight linker [96]. SAR studies indicated that the internal H-bond in fenobam and its analogues (keeping the linker straight) is essential for their potency [96]. The ligand pharmacophore corresponds to a binding mode in line with SDM data [75], in which fenobam binds deep down in the TM binding pocket forming an essential H-bond with $S^{3.39}$ [75] (an unusual ligand binding residue position in GPCRs [40]) and positioning its hydrophobic aromatic ring in a hydrophobic pocket between $P^{3.36}$, $Y^{3.40}$, $T^{6.44}$, and $W^{6.48}$ [75]. The latter residue forms an aromatic cluster with $F^{6.51}$ and $Y^{6.55}$, shown to be involved in receptor activation [75]. Residue $A^{7.47}$ whose mutation diminishes fenobam binding, is modeled close to the proposed ligand binding pocket as well and might be involved in ligand entrance between TM 1 and 7, one of the putative retinal entrance channels for bRho/Ops* [13], or have an indirect role in stabilization of the pocket. Mutation of residue $R^{3.29}$, in the direct proximity of $Y^{6.55}$, also has a small effect on fenobam potency, while mutation of $L^{5.43}$ only affects binding of another negative allosteric modulator [75].

The CRFR1 antagonist pharmacophore is a hetero-cyclic ring bearing a critical hydrogen-bond acceptor nitrogen and a orthogonal aromatic ring [97]. CRFR1 ligands are highly hydrophibic (polar substitutions of the pendant ring are not tolerated [97]) and the orthogonal orientation of the aromatic ring systems is essential [98]. In the CRFR1 receptor model,

the sp2-hybridized nitrogen in the heterocyclic ring of antalarmin forms a hydrogen bond with the protonated imidazole nitrogen of $H^{3.36}$, while the orthogonally oriented aromatic rings, in line with X-ray and NMR ligand confor-mation studies [98, 99], fit in an "aromatic cage" between $H^{3.36}$, $Y^{3.40}$, $Y^{5.43}$, $M^{5.47}$, $Y^{6.48}$, and $F^{6.52}$. Site-directed muta-genesis studies have indeed identified $H^{3.36}$ [100] and $M^{5.47}$ [88, 100] to be involved in interactions with non-competitive antagonists.

In our GPCR modeling protocols, we have used the above described binding mode hypotheses as experimental anchors to guide the refinement of 3D receptor models, shaping the binding pocket into a proper mould for structure-based virtual screening studies along the different steps in the GPCR modeling process (steps **1**-**5** in Fig. (**1**)). Already the selection of the modeling template (step **1**) can be driven by such experimental anchors, by considering for example TM-TM distances (Fig. (**2A**), section **3**), helical kinks (section **4.3**), and extracellular loop confor-mations (Fig. (**2B**), section **4.5**). For the construction of the GRM5 model, for example, we used ADRB2 as a modeling template because in this structure the distance between TM3 and TM6 at the intracellular side of $W^{6.48}$ is just large enough to accommodate fenobam (Fig. (**3**)), while the bRho structure is not compatible with this constraint [101]. Initial docking studies of known ligands in the receptor model have been guided by pharmacophore constraints [23, 30, 46, 50, 102-104] to satisfy experi-mentally determined (or hypothesized) receptor-ligand interactions and sometimes requires manual changes of rotameric states of residue sidechains. For example, the construction of a S1PR1 model required to rotate $F^{6.44}$ side chain to accommodate the binding of the long aliphatic tail of *S*-FTY-720-P (Fig. (**3**)). Pharmacophore and/or other geometrical constraints linking atoms or groups between receptor and ligand or within receptor or ligand themselves (e.g., distance and angle constraints to satisfy a specific H-bond) have been included in energy minimization and MD/MC refinement procedures (steps **3**-**5**) as well [30, 46] to satisfy experimental data. Such MD/MC simulations are generally performed in a fully hydrated phospholipid bilayer [46, 105, 106]. As a result of such a refinement protocol the shape of the binding pocket as well as the orientation of key residues are adapted to a known ligand in an experimentally supported binding mode. It should be noticed that not only modeling of the TM cavity, but (especially) also loop modeling requires experimental constraints (step **5** and discussed in more detail in section **4.5**). Receptor binding pocket expansion procedures without the presence of a ligand have also been described [107], but only few GPCR models applied for virtual screening (VS) have not been optimized in the presence of a known ligand. However, it should be stated that there is a danger to bias a ligand-binding cavity towards a single chemotype if too many constraints are given as input. Conformational samp-ling of the receptor structure affords the selection of conformers able to accommodate known ligands according to experimental data and discriminate them from non-binders. In several GPCR modeling studies, alternative conformations of the receptor (refined in the presence of different ligand and/or different cluster representatives from MD or MC trajectories) have been i) tested by retrospective

virtual screening to select the "best" model for prospective VS [50, 86], ii) used as an ensemble during the actual VS run [28, 105], or iii) considered as different states of the receptor and used to identify different ligand types [20, 24, 30]. Similar ligand-receptor interaction steered GPCR modeling strategies as described above like Mobile [108] and RED [50] are presented in section **5**.

4.3. Questioning Kinks: Alternative Conformations and Orientations of TM Helices

In principle, the receptor refinement protocols (steps **2**-**5** in Fig (**1**)) described in section **4.2** aims at refining not only the side chains but also the backbone conformation of the receptor to shape the binding pocket to an experi-mentally defined ligand-receptor binding mode. It is however questionable whether large backbone changes, associated to high energy barriers, can be sampled by MD or MC simu-lations. Therefore, alternative TM kinks and orientations should probably be considered at an earlier stage (step **1** in Fig. (**1**)) as well.

As discussed in section **3**, many of the structural differences between the helical bundles in GPCR crystal structures can be explained by specific sequence motifs (involving Pro/Gly/Thr/Ser residues) and/or TM-TM inter-actions which induce and/or stabilize helical kinks. The prediction of TM kinks based only on sequence motifs [41], however, can be rather difficult and it is very unlikely that all TM conformations are already covered in the set of currently available GPCR crystal structures. Moreover, it has been demonstrated that kinks may be reminiscent of ancestral features which disappeared along evolution [109].

Nevertheless some attempts have been made to model alternative helical kinks, guided by experimental data supporting the involvement of specific residues in ligand binding, but not facing the ligand binding pocket in initial receptor models based on bRho (or any other GPCR crystal structure) [7, 8, 20, 23, 26, 110, 111] (step **1** in Fig. (**1**)). The best way to test alternative TM conformations, however, is to challenge them in retrospective, or even better, prospec-tive structure-based virtual screening experiments as the ones described below.

For the construction of receptor models of the CCR5 receptor (a chemokine receptor, rhodopsin-like GPCR class) and of CRFR1 (secretin-like GPCR) we have consi-dered alternative helical kinks, which improved prospec-tive and retrospective virtual screening results, respec-tively. The $T^{2.56}XP^{2.58}$ motif in CCR5 and other chemokine receptors has been proposed to induce an alternative kink in TM2, as supported by SDM data probing the TM2-TM3 interface [111] and receptor-ligand interactions [89, 112, 113]. This alternative kink directs the residue at position 2.60 (Trp in most chemokine receptors) into the binding pocket, instead of towards the membrane layer as in the bRho, ADRB2, and AA2AR crystal structures. We were able to identify new CCR5 agonists by structure-based virtual screening using a CCR5 model containing this alternative TM2 kink [26], but not with a model based on bRho. Analogously, Secretin-like receptors presents two conserved proline residues ($P^{5.46}$, $P^{6.42}$) on different posi-tions than rhodopsin-like receptors ($P^{5.50}$ and $P^{6.50}$). Site-directed mutagenesis studies in secre-

tin-like receptors showed that $P^{6.42}$ is important for receptor activity and ligand binding [114, 115]. In accordance with modeling studies on CCR5 (TM2 and TM3) [26, 111] and 5HT1A (TM3) [110], we performed MD-simulations with model TM peptide stretches of the secretin-like receptor CRFR1, containing Ser, Thr, Pro, and Gly residues conserved among secretin-like receptors and alanine at the other positions to determine the trends of helix bending for the different TMs. As the relative orientation of most intracellular TM regions were shown to be in line with experimental data (see section **4.1**) these regions were used to fit the MD snapshots (Fig. (**1**)). Especially the extracellular top of TMs 5 and 6, which surround the postulated non-peptide ligand binding pocket in secretin-like receptors, show divergent TM bending compared to bRho and ADRB2 (Fig. (**5**)). Interestingly, a retrospective virtual screening accuracy of the alternative TM-reoriented CRFR1 model was significantly higher than the accuracy of bRho-based and ADRB2-based models. The ligand binding pocket of the TM-reoriented CRFR1 model is discussed in more detail in section **4.2**.

Salo *et al.* [23] modeled an alternative conformation of TM2 of the cannabinoid CB2 receptor (CNR2) to position residue C2.59, which is found to be facing the ligand binding pocket of CNR2 based on experimental data [116] but is directed towards the membrane in bRho and other GPCR crystal structures. The kink in the TM2 of bRho induced by the $G^{2.56}G^{2.57}$ motif was straightened, position 2.57 in CNR2 was aligned to position 2.58 of bRho. TM5 was modeled as a straight helix (like in a previous CNR1 model [117]), because this helix does not contain a proline residue at position 5.50 which induces a kink in bRho and all other GPCR crystal structures and because the location of residue $Y^{5.39}$ in a straight helix more consistent with experimental data [23]. The receptor model was refined by MD and subsequently successfully used to retrieve a new CNR2 agonist, as described in more detail in section **5**.

In one of the earliest reported retrospective VS studies [20], the TM6 helix in bRho-based models of the adrenergic beta 2 (ADRB2), dopamine D2 (DRD2), and delta opioid (OPRD) receptors was rotated counter-clockwise from 30° (when viewed from the extracellular domain), to model the active state of the receptor (based on what was known at that time regarding the activation of bioaminergic receptors [4]). These models were further refined in the presence of sets of known agonists manually docked according to experimental data and successfully used to retrieve known full agonists with hit rates very similar to those previously observed for antagonists in inactive state receptor models [20]. The same approach was used by others to construct an agonist-bound model of the orphan C-X-C chemokine receptor type 7 (CXCR7) and successfully applied this model to find agonists for this receptor [25]. Recently the inactive inverse agonist bound ADRB2 crystal structure [30] was customized to enable *selective* retrieval of full and partial agonist by retrospective VS. Instead of rotating TM helices, only minor conformational changes (rotation of two serine residues known to specifically H-bond to full/partial agonists) were required to model what probably represents an early intermediate state in agonist binding (Fig. (**6**)). Rearrangement of the receptor-agonist binding mode *via* either a rotamer toggle switch [31] or rigid-body shifts of see-saw motions of transmembrane segments [32] probably involves conformational changes like bending [31, 118] and/or rigid body movements [32, 119] of TM helices which are likely to be ligand-dependant and therefore difficult to predict.

As an alternative, fully template-independent *de novo* GPCR modeling approaches such as Predict [8] and Membstruck [7] can be used to model alternative TM orientations and kinks. Predict [8] is a multi-step computational protocol that identifies TM sequences, proposes alternative packing geometries for the seven helices (decoys) into 2D space, optimizes the relative rotational orientation of the helices, converts the most likely decoys into a simplified 3D representation, optimizes and clusters the best solutions,
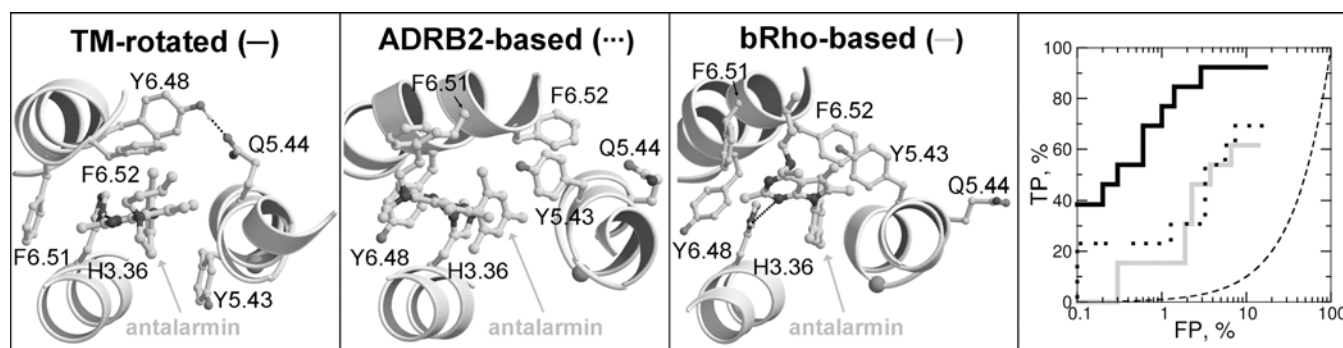

Fig. (5). The binding mode of antalarmin in different models of the CRFR1 receptor. A model in which the individual TM helices are modeled by MD (step 1 in Fig. (**1**), see description in text) and models derived by homology modeling from bRho and ADRB2 crystal structures. Important ligand binding residues in the CRFR1 pocket are depicted as balls and sticks. Hydrogen bonds are indicated by black dotted lines. On the far right panel: Enrichment in virtual screening of a focused database of 987 drug-like compounds (false positives, FP) and 13 known CRFR1 antagonists (true positives, TP) against: TM-reoriented (solid black line), ADRB2-based (dotted black line), and bRho-based (solid grey line) models of CRFR1. Docking simulations are performed with GOLD-Goldscore and docking poses are filtered and ranked according the Interaction fingerprint (IFP) topological scoring function [128]. The thin dashed black line represents the fraction of actives expected by random picking.

minimizes all-atoms models of each cluster repre-sentative and finally refines the most stable model by MD. Predict models have been successfully used to retrieve antagonists as well as agonists both in retrospective [8] and prospective [48, 49] virtual screening studies, as will be discussed in more detail in section **5**. Membstruck [7] is another de novo GPCR modeling method which works according to a similar protocol as Predict. The program was used to construct both antagonist and agonist-based models of ADRB2, and an agonist-based model was able to discriminate known ADRB2 ligands from a database of decoys [47].

4.4. Docking Ligands into GPCR Models

Along the different GPCR modeling steps different alternative receptor conformations should be primarily be selected and challenged by their ability to accommodate known ligands in a binding mode satisfying experimen-tally-determined receptor-ligand interactions. In the final valida-tion step, one can test and apply the receptor not only for its ability to reproduce experimentally plausible ligand binding poses or to use the binding modes to explain and predict SAR and SDM data (structure-based design), but also to retrieve ligands from large chemical databases. The most widely used technique to predict receptor-ligand binding orientations in a rapid manner is molecular docking [120], which combines an algorithm to generate different docking poses with a scoring function to rank them. Selection of the most appropriate docking/ scoring protocols for binding mode prediction and structure-based virtual screening is very much dependent on physicochemical details of target-ligand interactions [121-124] and fine details of the protein structure [125, 126]. It is therefore necessary to evaluate different docking-scoring approaches or even to optimize scoring functions for protein/ligand training sets before applying them to unknown test cases. Post-processing strate-gies for selecting and ranking docking poses have recently received much attention as an alter-native approach to solve the problem of protein-ligand docking and scoring accuracy [127].

In the past few years we have been using a topological scoring function (IFP) based on protein-ligand interaction fingerprints [128] for ranking docking poses in structure-based virtual screening runs [30, 46, 129]. Receptor-ligand binding modes derived from known receptor-ligand X-ray structures [30, 128, 129] and/or supported by SAR/ pharmacophore and site-directed mutagenesis data [30, 46] can be used as IFP references. Docking low-molecular weight compounds, notably in open protein cavities (such as in GPCR receptor models lacking ecl2 on top of the TM binding pocket), can yield multiple different binding modes with comparable binding energies according to classical energy-based scoring functions [128]. Furthermore, omitting ecl2 from GPCR receptor models can provoke docking solutions of large, highly flexible "inactive" compounds that are artificially oriented into the unoccupied region [46, 103]. The IFP scoring protocol however, is able to discriminate between irrelevant ligand docking poses and docking poses comparable to that of known inverse agonists/antagonists [30, 46], and has even been successfully used to discriminate between inverse agonists/full antagonists and full/partial agonists [30], as exemplified in Fig. (**6**).

As previously reported [30] and demonstrated in Fig. (**6**), the topological IFP scoring function outperforms Gold scoring [130] in retrieving ADRB2 inverse agonists/ antago-nists in the original crystal structure (e.g., timololol (**A**)) and is better suited for selectively retrieving agonists in the customized ADRB2 structure than Gold and Surflex scoring [131] (e.g., procaterol (**B**) and carvedilol (**C**)). Interestingly, the top-ranked pose of known ADRB2 ligands according to the native scoring function often has also one of the highest IFP similarity value. This means that the scoring functions perform relatively well in terms of binding mode prediction. The docking score itself of the ligand pose is however relatively poor, resulting in low rankings in the hit lists (Fig. (**6D**)). The poor scores of these poses stem primarily from steric rather than polar terms of the scoring functions. Interestingly, the IFP scoring function can easily overcome scoring problems associated with ligand confor-mational energies [132] or small changes in the protein conformation (ligand-induced fit) [126] even without fine tuning the settings of these docking programs. Post-pro-cessing poses by local/full energy refinement of the corresponding complex will partly solve the problem but not entirely unless ensemble docking is performed on a set of previously-generated protein conformers.

Fig. (**6**) illustrates how IFP manages to distinguish partial/full agonists from inverse agonists/antagonists. *R,S*-procaterol (panel **B**) and IFP reference *R*-isoproterenol (**B,C**) share most of their interactions with the customized ADRB2 structure thus yielding a high IFP score despite a weak Surflex score and rank (**D**). The antagonist *S*-carvedilol (**C**), on the other hand, has a high docking score as it makes a high number of contacts with the receptor binding pocket. Like the *R*-isoproterenol reference, carve-dilol interacts with the three important serine residues on TM5 (S5.42, S5.43, and S5.46) but through a set of different interactions (hydrophobic contacts instead of H-bonds; panel **D**). While not affecting the Goldscore (carvedilol is ranked first), changing the type of molecular interactions to the receptor yields in a relatively low IFP score and ranking (**D**), as was also determined for almost all other inverse agonists/ antagonists docked in the customized ADRB2 structure [30].

The example of timolol also demonstrates, in agreement with a recent comparative evaluation of different virtual screening methods [133], that IFP scoring can be used for various scaffold hopping scenarios. Timolol and carazolol share the same 1-(isopropylamino)-3-aryloxy-propan-2-ol scaffold, but marked differences in the aryl moiety (Fig. (**6A**)) are disfavourable to a good 2D-similarity rank, while IFP scoring (as well as carazolol-based 3D-similarity methods) still rank timolol among the top scorers (Fig. (**6D**)). Alternatively, it was observed in the same study that when only disconnected fragments (maximum common edge subgraph MCE) are common to two ligands the 3D-similarity ranking is logically affected, while IFP scores (and 2D-similarity scores) are not [30].

4.5. Laying the Loops

As discussed earlier in section **3**, bRho seems to be a relatively suitable modelling template for modelling the upstream ecl2 segment, while the long upstream in
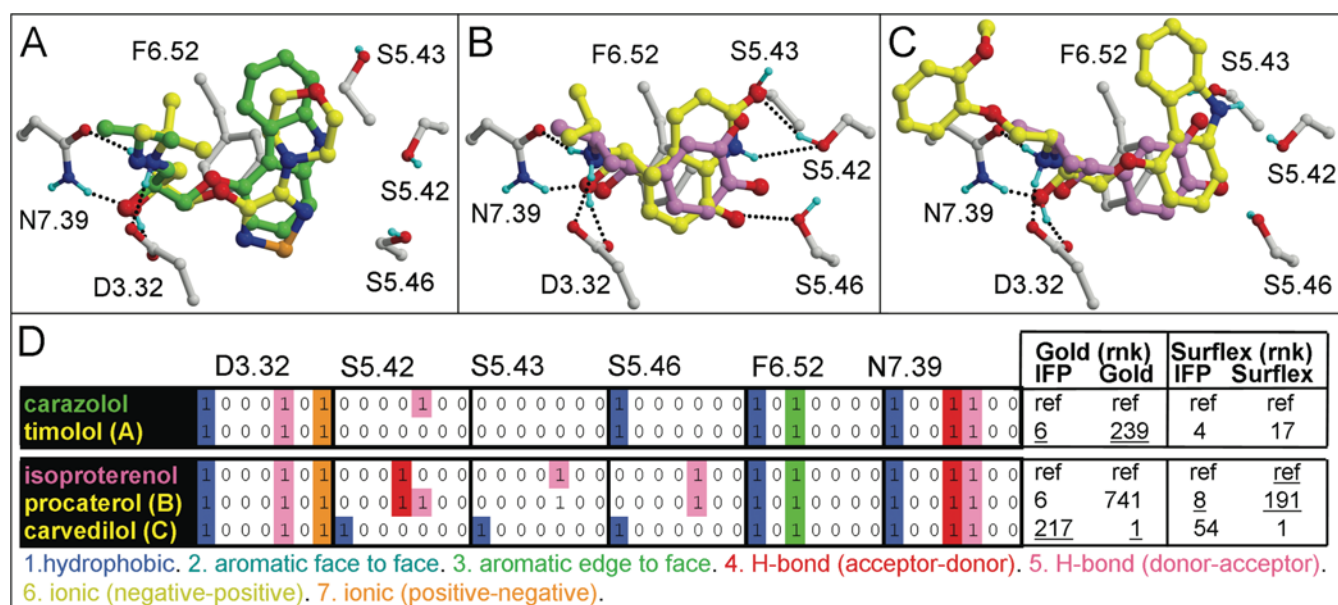
Fig. (6). Docking poses selected by IFP of: (A) S-timolol (yellow carbon atoms, docked with Gold) in the ADRB2 crystal structure compared to the reference binding mode of S-carazolol (green carbon atoms), (B) R,S-procaterol (yellow carbon atoms, docked with Surflex) and (C) S-carvedilol in the customized ADRB2 structure compared to the reference binding mode of R-isoproterenol (green carbon atoms). H-bonds between the docking pose and the receptor are depicted by black dots. The IFP bitstrings of receptor-bound timolol (A), R,S-procaterol (B), and carvedilol (C) are compared to the reference IFPs of carazolol and isoproterenol in panel D. For reasons of clarity, the bit strings of only six residues (out of 33) are shown as an example. For each pose and corresponding IFP bit-string, ranks in a retrospective virtual screening excercise [30] are indicated for Gold/Surflex docking ranked by IFP, Goldscore and Surflex score.

combination with short downstream ecl2 loops of ADRB2 and AA2R are relatively rare among GPCRs (Fig. (**2C**)). Nevertheless, the different extracellular loop structures displayed in the currently available GPCR crystal structures tell us to approach GPCR loop modeling (step **5** in Fig. (**1**)) with much caution and that it should be reserved for cases where loop building can be guided by experimental restraints, rather than carried out in a high-throughput fashion and derived directly from bRho (or any other GPCR crystal structure) [46]. The minimal experimental constraint imposed on modeling of the second extracellular loop should at least be the conserved disulfide bridge between TM3.25 and C45.50, as derived from amino acid sequence alignments (section **4.1**). For receptors lacking any of these cysteine residues, its is better to omit ecl2 from the model. Other experimental constraints, like H-bonds and contacts between the extracellular loop and the ligand (or other parts of the receptor) can be used at the loop modeling step itself [134] (step **5** in Fig (**1**)), and during the refinement of the loop conformation as described in section **4.2** (step **4**).

For a set of GPCR targets, the dopamine D2 receptor (DRD2), adenosine A3 receptor (AA3R), and the thromboxane A2 receptor (TA2R), we have evaluated the implication of including ecl2 atomic coordinates in terms of structure-based virtual screening accuracy: the suitability of the 3-D models to distinguish between known antagonists and randomly-chosen chemically-similar decoys using automated docking approaches [46]. Explicit modelling of the ecl2 loop was found to be important in only one (AA3R) out of three test cases whereas a loopless model was shown to be accurate enough in the two other receptors.

Interestingly, the antagonist binding mode we proposed for AA3AR [46] is not in line with the binding orientation of ZM241385 in the recently reported crystal structure homologous AA2AR receptor [12], but could nevertheless be used to achieve significant enrichments over random picking (albeit somewhat lower than in DRD2 and TA2R models) in the retrospective VS study.

Also several other structure-based virtual screening studies have shown that loopless TM models of GPCR receptors can be suitable targets for virtual screening (see Table **1**) [8, 47-50]. This further supports our believe that loop (notably ecl2) modeling should be reserved for cases where loop building can be guided by experimental restraints and the effect of incorporation of extracellular loop(s) can be tested by retrospective VS.

4.6. Model Validation by Prospective Virtual Screening of Compound Libraries

The final refined receptor model can be evaluated and validated in various ways. Many models have been used to rationalize site-directed mutagenesis data and/or structure-activity relationships, mostly in a retrospective manner. We believe that the strength and added value of GPCR models especially lies in its application to identify new ligands from large chemical databases by prospective virtual screening studies (step **7**, Fig. (**1**)). Fig. (**7**) gives an overview of a general structure-based virtual screening flow chart.

The initial compound library should be filtered to remove undesirable compounds exhibiting chemically reactive moieties [135], scaffold-inherited toxicity [136], or poor oral
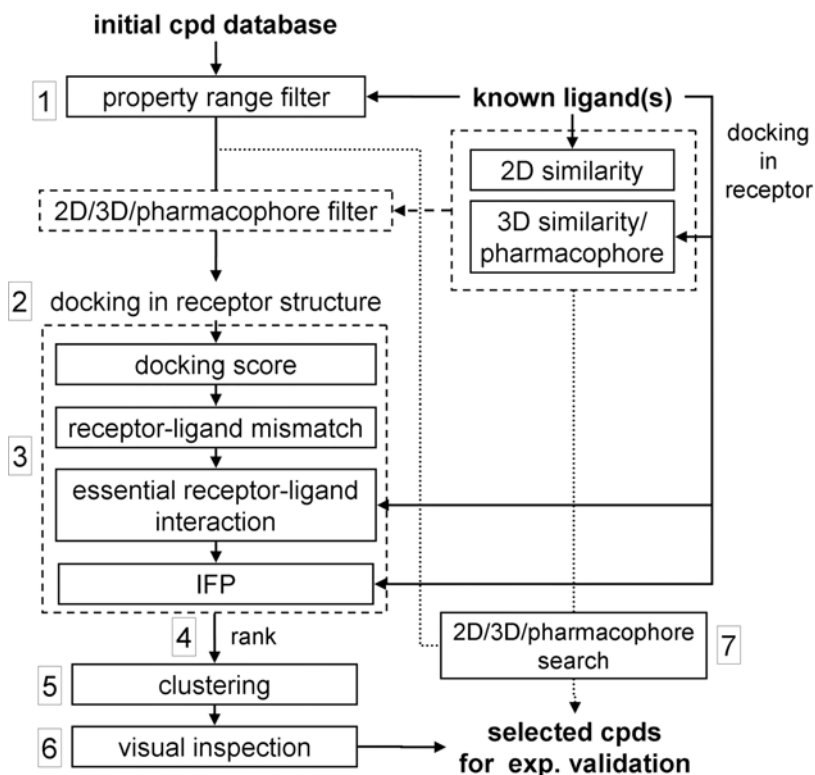
Fig. (7). Structure-based virtual screening flow chart.

bioavailability [137]. An overview of chemical libraries suitable for virtual screening is presented in [138]. In most reported GPCR-based virtual screening studies, such pre-filters are applied to construct the chemical library to be screened (Table **1**). Additional filters, derived from property ranges (e.g., molecular weight, number of rotatable bonds, number of rings, hydrogen bond donor/acceptor counts, number of positively/negatively charged atoms, etc.) determined from a set of known actives (step **1** of Fig. (**7**)), are applied in most GPCR VS studies as well (Table **1**) to obtain a more "lead-like" database. Chemical similarity descriptors and metrics [139] and 3D-shape similarity or pharmacophore models [140] derived from known ligands can be used to narrow down the number of compounds to be handled in automated docking simulations (step **2**) even further.

In step **2** the chemical database is automatically docked in the receptor model. Numerous automated docking programs [138, 141] and scoring functions [138, 141] based on different physicochemical approximations are available. In several GPCR VS studies [86, 102, 104, 142], docking simulations have been guided by pharmacophore constraints [143, 144].

In steps **3** and **4**, the docking poses are post-processed and scored, respectively. Many structure-based virtual screening investigations have only employed docking scoring functions to rank docking poses, but more and more structure-based virtual screenings apply additional filters to post-process docking results. The reason for this is that, as already mentioned in section **4.3**, the scoring accuracy of docking-scoring combinations is very much dependent on physicochemical details of target-ligand interactions [121-

124] and fine details of the protein structure [125, 126]. One way to overcome these problems is to use a consensus scoring strategy. One or several scoring functions [121] can be used in combination with one or several docking algorithms [145], applied to one or a set of receptor structures [146], and using different consensus scoring scenarios [20]. Especially the latter consensus scoring approach has been applied in several of the VS studies [20, 24, 29, 86, 102, 103], but also docking into an ensemble of receptor coordinates at the same time [28], or into alternative receptor structures to retrieve only specific ligand types [20, 30] has been applied. Topological filters can be used to filter out poses exhibiting steric or electrostatic mismatches between the ligand and its target [147]. In the case of CRFR1 (Fig. (**5**)), we have discarded docking poses not satisfying important ligand-receptor H-bonds. Such essential receptor-ligand interactions can be derived from experimental data featuring a reference binding mode of a true active, and can be used as *post*-processing filter [29, 30, 50, 86]. In some VS studies, a receptor-ligand interaction finger-print (IFP) scoring metric [128, 148] has even been used to *rank* docking poses [30, 46, 101]. When too many ligands are retrieved along the VS funnel, it is generally wise to cluster virtual hits by chemical diversity (step **5**). The most intuitive way, at least for medicinal chemists, to achieve this kind of classification is to group compounds by chemical scaffolds [24, 26, 149] and prioritize scaffolds rather than individual compounds. Sampling a few representative analogs for each scaffold usually enables a selection of chemically-dissimilar compounds for biological evaluation [26]. Finally, the selected docking poses and compounds should be visually inspected in step **6** for the ultimate selection: no algorithm

yet outperforms the brain of an experienced modeller for such a task.

Parallel to the docking-based virtual screening run, 2D-similarity, 3D-similarity, or pharmacophore searches can be performed to complement the docking-based hit list [23, 27, 50, 122, 142, 150]. The receptor bound conformation of reference compounds can be derived from docking simulations in the receptor model. In this way, the receptor model and information derived from the receptor-ligand complex are used to set up 3D-similarity searches or pharmacophore models. Alternatively, pure receptor-based pharmacophore models, derived from ligand-receptor interaction hot spots in the binding pocket, can be even used in the absence of a ligand [151].

5. GPCR VIRTUAL SCREENING SUCCESS STORIES

As described in the previous section, many possible problems and pitfalls have to be overcome in the GPCR modeling process. However, even despite the possible structural inaccuracies of the final refined GPCR models, these models have been efficiently used to find new ligands, as will be described in this final section.

Table **1** gives an overview of recent prospective structure-based VS studies on GPCR models. Figure (**8**) presents agonists as well as antagonists identified by prospective structure-based virtual screening studies in GPCR models ([8, 23, 24, 26-28, 49, 50, 103, 104, 152, 153], see Table (**1**)) and the previously known ligands used to refine the model used for the structure-based VS run. Notice that structure-based virtual screening often yields new chemical scaffolds, but still contain essential functional groups like positively or negatively charged atoms, often because these chemical features/properties were used as filters/constraints to set up the initial ligand database or score/rank docking poses. Interestingly, novel receptor agonists have been found using antagonist-bound receptor models and vice versa (see also Table (**1**)). Fig. (**8**) shows that most of the studies have focused on bioaminergic receptors, but that more and more successful prospective structure-based virtual screenings are reported also for other rhodopsin-like GPCRs (brain-gut peptide, chemokine, lipid, peptide, purine), while the first models of glutamate-like and secretin-like receptors have been successfully challenged in retrospective virtual screening exercises, opening up new challenges in the field of structure-based VS for GPCR ligands. Only prospective (and not retrospective) VS studies on GPCRs will be discussed in more detail in the following section.

The very first prospective virtual screening study on a GPCR model was conducted by Varady *et al.* on the identification of dopamine D3 (DRD3) receptor ligands [105]. A homology model was derived from the bRho crystal structure and refined by MD in a fully hydrated phospholipid layer. Snapshots from the MD-trajectory were clustered into four families out of which one representative was selected for docking studies. First a pharmacophore was derived from a set of known DRD3 ligands (antagonists and partial agonists (Fig. (**8**)) and used as a query for retrieving a first hit list of 6 727 compounds from an initial database of 250 000 compounds. This hit list was then docked to each of the four receptor cluster representatives. A total of 2 478

compounds, ranked within the top 30% of the library for at least two receptor conformations, were finally selected. From this second hit list, 1 314 molecules showing a Tanimoto similarity index lower than 80% to any of the 10 known DRD3 ligands were selected. Twenty molecules were finally selected for biological evaluation, eight com-pounds (Fig. (**8**)) exhibited submicromolar binding affi-nities.

A similar strategy was used by Evers *et al.* to identify a new neurokinin-1 receptor antagonist [104]. Ligand-biased modeling of the receptor was first realized with the Mobile program. A set of crude homology models was first obtained from the bRho crystal structure and a known NK1R antagonist (Fig. (**9**)) was docked into the receptor model. Four poses satisfying key interactions were selected for generating a new set of receptor models scored by a knowledge-based scoring function to satisfy the receptor-ligand constraints. The selected model was refined by energy minimization. A 1D-filter (removing large and flexible structures) was used to discard 50% of the initial database of 800.000 compounds. From this reduced database, 131 967 compounds were selected satisfying a 2D-pharmacophore query, from which subsequently 36 704 compounds were selected satisfying a 3D pharmacophore hypothesis. Using an implicit definition of forbidden volumes, the size of the hit list was further reduced to 11 109 compounds. Only these molecules were docked to the refined receptor model (imposing an H-bond to an important residue in NK1R) and scored by the same knowledge-based scoring function. Energy minimization refinement and visual inspection of the 1000 best-ranked ligands resulted in the final selection of seven candidates. One of these molecules (Fig. (**8**)) proved to bind to NK1R with a submicromolar inhibition constant. The same approach was used to identify alpha 1A adrenergic receptor (ADA1A) antagonists [103]. Homology models of ADA1A were constructed in a similar fashion as for NK1R and the one that best suited the known binding mode of a reference antagonist was selected for structure-based virtual screening. A series of filters of increasing complexity (satisfying topological and pharmacophore restraints) were then applied to select 22 950 compounds (from the initial 728 000 compound stock collection) for docking simulations in the receptor model. An optimal docking-scoring strategy, derived from a retrospective virtual screening evaluation, was used to select 300 compounds which were then clustered and a diverse subset of 80 molecules was finally evaluated for receptor binding. Thirty seven compounds exhibited Ki values lower than 10 μM, of which 24 molecules bound in the submicromolar range and three compounds (Fig (**8**)) below 10 nM.

An alternative ligand-steered homology method was recently reported by Cavassoto *et al.* to find new antagonists of the melanin-concentrating hormone receptor 1 (MCHR1) [50]. First, a homology model was built based on the bRho crystal structure. In a second step, 20 receptor-ligand models were generated by docking four known ligands (Fig. (**8**)) into the receptor model using an experimentally supported receptor-ligand distance constraint, ranked by a scoring function and clustered. These 20 structures were then subjected to MC simulations (without receptor-ligand distance constraint) and the top 20 best-energy complexes were selected and merged with the original 20 structures.

Table 1.   **Overview Prospective Receptor-Based Virtual Screening (VS) on GPCR Models**

Receptor[a]	Template[b]	Construction Method[c]	ecls[d]	Refin. Protoc.[e]	Database Pre-Filter[f]	Conf. Search[g]	Post-Processing + Scoring[h]	Prediction Retrospec. Ligand	Prediction Prospective Initial db[i]	Prediction Prospective Hits[j] (tested)	Ref[k]
amines											
ADA1A	bRho	Mobile	2	em	dl+2D+3D	ad	clust+score	ant	827.000	37 ant (80)	[103]
ADRB2	ADRB2	X-ray	1,2,3	-	?	ad	score	ant	4.000.000	19 ant (56)	[17,156]
DRD1	DRD3	GPCRgen	1,2,3	em	dl+2D	ad	clust+c-score	-	310.000	3 ant (17)	[24]
DRD2	DRD3	GPCRgen	1,2,3	em	dl+2D	ad	clust+c-score	-	310.000	3 ago+2 ant (17)	[24]
DRD2	de novo	Predict	no	em+md	dl+ll	ad	score	ant/ago	1.600.0000	7 ant (42)	[8,49]
DRD3	bRho	thread	1,2,3	em+md	dl+3D	ad	score+2D	-	250.000	8 lig/ago (20)	[105]
HRH4	bRho	Modeller	1,2,3	em	dl	(cnstr) ad	c-score	lig	5.066.235	16 (255)	29,86]
5HT1A	de novo	Predict	no	em+md	dl+ll	ad	clust+score+2D	ago	1.600.0000	16 lig/ago (87)	[48,49]
5HT4R	de novo	Predict	no	em+md	dl+ll	ad	clust+score+2D	ago	1.600.0000	19 lig/ago (93)	[49]
Brain gut peptides											
MCHR1	bRho	thread	no	em	dl+2D+3D	3D	clust	-	615.000	19 ant (795)	[161]
TRFR1/2	bRho	3D-map	1,2,3	em+md	3D	ad	score+clust.	-	1.000.000	8 ant (~100)	[28]
Chemokines											
CCR3	de novo	Predict	no	em+md	dl+ll	ad	clust+score+2D	-	1.600.0000	5 lig/ant (43)	[8,49]
CCR4	bRho	thread	1,2,3	em+md	dl+ll	ad	score	-	450.000	16 ant (116)	[154]
CCR5	bRho	thread+kink (2)	1	em	dl+ll+2D	ad	clust+score+2D	ant	1.600.0000	6 ago + 1 ant (59)	[26]
Lipids											
CNR2	bRho	thread+kink (2+5)	1,2,3	em+md	none	ad+3D	3D+score	-	55.600	1 ago (68)	[23]
peptides											
FPR1R	bRho	thread	no	em	none	ad+3D	3D	-	480.000	52 ant (4342)	[150]
NK1R	de novo	Predict	no	em+md	dl+ll	ad	score	lig	1.600.000	53 ant (8)	[8,49]
NK1R	bRho	Mobile	2	em	dl+2D+3D	cnstr ad	norm score	-	827.000	1 ant (7)	[104]
purines											
FFAR1	bRho	thread	1,2,3	em+mc	dl+2D+clus	ad+3D	score+3D+clust.	-	2.600.000	13 ago+2 ant (52)	[27,152]

[a]Receptors clustered according to Surgand *et al*. [40]; [b]bRho (bovine rhodopsin) and ADRB2 are crystal structures, other receptors are models, *de novo*: no template used; [c]Construction methods of the TM helical backbone (GPCRgen [34], Mobile [108], Predict , Membstruck) are described in more detail in the text; [d]Construction of extracellular loops (ecls) explicitely guided by experimental data are underlined; [e]Energy minimization (em) and molecular dynamics (md) refinement protocols are generally performed in the presence of a ligand; [f]Consecutive filters (dl (drug-like physicochemical properties), ll (leadlike physicochemical properties), 2D (two-dimensional topological/chemical similarity/pharmaphoric features/sub-groups), 3D (three-dimensional pharmacophore)) used to compile database for docking/3D conformer search; [g]Conformer search method: ((H-bond) constr(ained)) automated docking (ad), protein-based or docked ligand-based 3D pharmacophore search (3D); [h]Method to score, rank and/or filter conformers: clust. (scaffold clustering), (c-)score ((consensus) docking scoring function), 2D (two-dimensional topological/chemical similarity/pharmaphoric features/sub-groups), 3D (three-dimensional pharmacophore); [i]Prospective validation: initial databatabase (db) and [j]number of experimentally confirmed hits with detectable affinity/activity (of the total number of tested compounds); [k]References describing model construction and refinement as well as virtual screening are reported; [k]not reported.
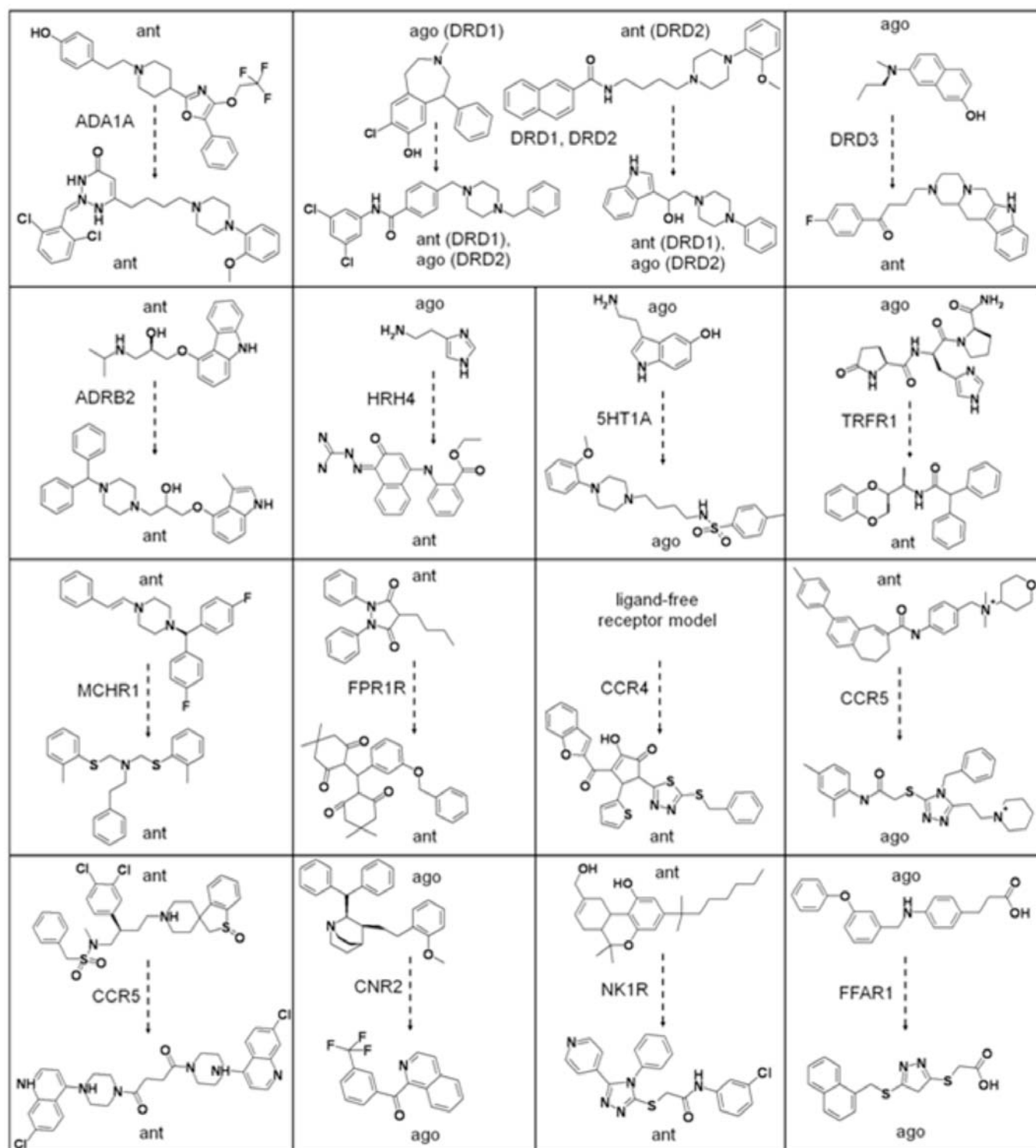
Fig. (8). Examples of agonists (ago) and antagonists (ant) identified by prospective virtual screening studies in GPCR models ([8, 23, 24, 26-28, 49, 50, 103, 104, 152, 153], see Table (**1**)) refined with known ligands (upper structure). It should be noticed that only one representative hit (and one reference compound) is presented for each VS study and that compounds of some prospective VS studies reported in literature (Table **1**) have not been disclosed (DRD3, 5HT4, NK1R, and CCR3 [49], and CXCR7 [25]).

These structures were clustered and visually inspected for the presence of experimentally supported receptor-ligand interactions, yielding 8 models which were challenged by retrospective virtual screening. The model retrieving the

highest diversity of known ligand chemotypes in the top-ranked list was selected and used for a prospective docking-based virtual screen of a filtered database of 187 084 compounds. A set of filters was imposed to post-process the

docking poses requiring the absence of ligand-receptor clashes and presence of a receptor-ligand H-bond. The remaining ca. 7 000 compounds were clustered by chemical similarity and the highest scoring compounds per cluster were chosen. Molecules with a total charge of +1 were given priority for biological evaluation, resulting in a set of 281 compounds of which 129 were commercially available. Six of these exhibited Ki values in the micromolar range (a hit enrichment of 12 fold compared to random screening of a corporate collection).

In a relatively straightforward structure-based VS studies of Bayry *et al*. [154] and Liu *et al*. [153], antagonists were found by docking studies on bRho-based homology models of chemokine receptors 4 (CCR4) and 5 (CCR5), respectively (Fig. (**8**)). About 13 000 compounds selected from an initial database of 450 00 compounds by a pharmacophore filter were docked in CCR4 and scored by a single scoring function, yielding a hit list of 116 top-ranked compounds; 16 of these inhibited CCR4-mediated cell migration with IC50 values lower than 10 nM. A database of 80 000 compounds was docked in CCR5 and ranked with a single scoring function, yielding a hit list of 150 compounds from which 95 were purchased and tested; one compound exhibited an IC50 value of 2 µM which was further optimized by structure-based design (see later).

Edwards *et al*. [150] identified novel formylpeptide receptor (FPR1R) antagonists by virtual screening using a pharmacophore model derived from antagonist docking poses in a FPR1R receptor model (Fig. (**8**)). Exclusion spheres derived from the receptor binding pocket were added to the pharmacophore model and the top-ranked 4 234 compounds were selected from a ~480.000 compound library; 52 of these compounds were confirmed hits, of which 30 had Ki values in the micromolar range.

While in the above mentioned VS studies new antagonists were found by docking into receptor models based on the inactive bRho structure and refined using known antagonists (Fig (**8**)), more and more reports show that also agonists can be found by structure-based VS in such inactive receptor models [23-27]. In some of these studies [23, 27], the initial bRho-based ground state model was refined by true agonists, but in other cases [24, 26] the inactive model was even refined by antagonists. The other way around, *agonist*-biased models have also been successfully applied to find antagonists [24, 29]. Examples of these studies will be discussed in more detail in the following section.

Kellenberger *et al*. identified new CCR5 agonists by virtual screening against an antagonist-customized receptor model [26] (Fig. (**8**)). A homology model was derived from the bRho crystal structure, but an alternative bend in TM2 induced by the $T^{2.56}XP^{2.58}$ motif (conserved among chemokine receptors) was modeled separately (see section **4.3**). This receptor model was validated by retrospective virtual screening using two different docking programs. A library of 1.6 million compounds was first filtered by 1D (drug likeness) and 2-D pharmacophore filters, decreasing the numbers of hits to 431 029 and 44 524 compounds, respectively. This database was docked into the receptor model by two different docking programs in parallel. The top 5% of the two hit lists were then independently classi-fied based on

common scaffolds. Scaffold classes enriched enough in well-scored molecules were retained and a final list of 77 molecules was selected by visual inspection of the binding mode. Of these, 59 could be purchased and were biologically tested; 10 compounds exhibited a detectable binding affinity for CCR5, of which four molecules had an IC50 in the high micromolar range, and of which three were surprisingly characterized as agonists.

Salo *et al*. found a novel agonist of the cannabinoid CB2 receptor (CNR2) by structure-based VS in receptor model derived from the ground state bRho crystal structure and refined in the presence of a known CNR2 agonists (Fig. (**8**)). The conformations of TM2 and TM5 were customized to position certain residues shown to be involved in ligand binding by experimental data into the binding pocket, as described in section **4.3**. The model was then subjected to a constant temperature MD run with positional constraints on the helix backbone atoms and a subsequent simulated annealing protocol applied to the side chains of the ligand binding pocket. Five structures were extracted from these simulations and the best snapshot was selected based on the quality of crude CoMFA models derived from docking poses of known agonists. This final model was used to derive three different pharmacophore models: one based on the pharmacophoric points of known agonists in the receptor model, a second model based on the binding cavity of the docked ligands, and a third model based on the combination of features taken from both the receptor and bound ligands.

Tikhonova *et al*. identified new agonists for the free fatty acid receptor 1 (FFAR1) [27]. The putative binding cavity [40] of an initial FFAR1 model, derived from the bRho crystal structure, was subjected to MC simulation [152]. Representatives of 12 structural clusters from the MC-trajectory were selected according to the orientation of charged residues in the binding pocket and used for auto-mated docking studies of a known agonist. Docking results in combination with solvent accessible surface area analysis and molecular interaction field analysis of the binding pocket and amino acid sequence analysis, were used to generate binding mode hypotheses which were then experimentally corroborated by SDM studies [152]. The model best in line with the SDM data was further refined by MC and QM energy minimization to optimize a part of the receptor-ligand complex. An initial database of ca. 2.6 million compounds was reduced to 70 477 molecules by a topological filter [27]. This set of compounds was docked into the refined receptor model and also subjected to a 3D pharmacophore search based on the refined receptor-ligand complex. From the docking simulations 3 131 compounds were selected ranked by a docking scoring function, while 1 581 compounds passed the pharmacophore filter. From these hit lists compounds were clustered by chemical diversity and visually inspected. From the subset of 183 compounds obtained both by docking and the pharmacophore search, 32 compounds were selected, appended by 10 unique compounds from each of the different hit lists. Experimental testing identified six active compounds, of which five displayed agonistic activity with EC$_{50}$ values around 10 µM. One of the confirmed hits was only identified by docking and another one only selected by the pharmacophore search, while four hits were found by both methods.

Kiss *et al*. [29] found new ligands for the histamine H4 receptor (HRH4). Receptor models based on the bRho crystal structure were separately refined in the presence of known agonists and antagonists [86]. These optimized models, as well as a set of ligand-free receptor models were challenged by a docking-based virtual screening study. The highest enrichment of known antagonists and agonists in the top-ranked scorings lists was achieved by the agonist-bound model (Fig. (**8**)) and these enrichments increased further when the docking simulation was guided by a receptor-ligand interaction pharmacophore constraint. This model was later used in a prospective VS study [29]. A large database of 5 066 235 compounds were all docked into the validated receptor model and ranked using an optimized scoring functions combination [86]. The top 2 000 hits were visually inspected, discarding poses having a wrong tautomeric state and not positioned in the binding cavity and preferring compounds showing experimentally supported receptor-ligand interactions, yielding a final list of 128 compounds of which 66 were purchased. The top 45 000 ranked docking poses were automatically filtered consi-dering a specific receptor-ligand interaction, clustered by chemical diversity, yielding a selection of 229 compounds of which 189 (including 23 analogues) were purchased. Out of the total of 250 biologically tested compounds 16 com-pounds showed a significant (> 20%) displacement of 10n M [^{3}H]--histamine at a 5 μM concentration (Fig. (**8**)).

Jones *et al*. [25] identified C-X-C chemokine receptor type 7 (CXCR7) agonists. A homology model was derived from bRho and customized by twisting TM6, like described in section **4.3**. A database containing an unknown number of compounds was docked into this model and 1000 com-pounds were selected based on docking score and chemical diversity criteria. Of these, 392 compounds were available for screening and two of these were experimentally confir-med to have agonist activity at high concentrations (ca. 100 μM).

Engel *et al*. found new antagonists of thyrotropin-releasing hormone receptors (TRFR1/2) [28] using an agonist-biased receptor model (Fig. (**8**)). A homology model of TRFR1 based upon a 3D projection map of bRho [155] was refined and validated by numerous prospective experi-mental studies. An initial database of ca. 1 million com-pounds was first screened using a receptor-based pharma-cophore model derived from the refined and validated receptor-agonist complex, yielding a hit list of ca. 100 000 compounds from which 10% was selected based on chemical diversity. These ca. 10 000 molecules were then docked to an ensemble of five different receptor structures, considering flexibility of the binding pocket. The top 10% ranked compounds were again clustered by chemical diversity and a final selection of 100 compounds (appended by several related compounds identified by a nearest neighbour search) were tested experimentally. Eight com-pounds were experi-mentally confirmed to be TRFR1 antago-nists (no agonists were found), including five structurally diverse chemical classes.

The recently published carazolol-bound ADRB2 crystal structure has recently been used for prospective virtual screening by Sabio *et al*. [156]. In-house proprietary and commercial databases of ca. 400 000 and 4 million com-pounds were docked in the ADRB2 crystal structure and scored with a single docking scoring function. Out of 150 tested compounds, 8 exhibited micromolar, and 23 sub-micromolar Ki values (of which 4 in the subnanomolar range). Most of the reported virtual hits were however chemically similar to carazolol (Fig. (**8**)).

Becker *et al*. [49] described the use of (crystal structure) template-independent models generated by the *de novo* Predict method [8]. The top 10% of an initial database of 1.6 million compounds passing a receptor-specific topological filter was docked into 3D models of three aminergic recep-tors (5-hydroxytryptamine 1A (5HT1A), 5-hydroxytryp-tamine 4 (5HT4R), and dopamine D2 (DRD2)), the peptide receptor NK1R, and the chemokine receptor type 3 (CCR3) and rescored using several scoring functions. A series of cut-off values for each score was used to reduce the size of the hit list. The remaining compounds were then filtered by using 3D principle component analysis based on the 3D properties of docked solutions. Molecules describing the same 3D space as known ligands were finally retrieved and clustered by diversity, yielding a list of ca. 100 representa-tive virtual hits for each receptor. Hit rates between 12 and 21% were reported at a 5 micromolar cut-off for four (5HT1A, 5HT4, DRD2, NK1R) out of five targets (see Table **1**).

In most of the studies described above, the final hit selection is performed by scoring and post-processing recep-tor-ligand docking poses (steps 2-6 in Fig. (**8**)). In some cases [23, 27, 50, 150], models of receptor-ligand complexes have been used to construct pharmacophore models and/or derive exclusion spheres for pharmacophore searches (step **7** in Fig. (**8**)). Whereas ligand-based models can only reveal binding features which are already present in the reference ligands, the inclusion of complementary information from the receptor cavity allows for a n understanding of the molecular recognition process [142]. In fact, some of the hits found in the structure-based VS runs have been successfully optimized by structure-based design, similarity searches, and analysis of the receptor-ligand docking pose. Comparative docking studies of a 5HT1A VS hit (Fig. (**8**)) in models of the 5HT1A and ADA1A receptors, as well as the hERG ion channel were used to design analogues with higher selectivity for 5HT1A over ADA1A and reduced affinity for hERG [48]. Fragments of an initial VS hit found for CCR5 (Fig. (**8**)) and a known CCR5 antagonist were assembled to design an antagonist with higher affinity [153]. Similarity searches were performed to find close neighbors of experi-mentally confirmed VS hits for FFAR1 (Fig. (**8**)), and these analogues were used to derive structure-activity relation-ships and further validate the receptor model [27]. Nearest neighbour and pharmacophore model searches were also performed based on initial VS hits for the MCHR1 receptor (Fig. (**8**)), resulting in many additional hits for this receptor [50]. Similarity analysis was also used to identify more agonists based on an initial hit found by structure-based VS in CXCR7 [25]. Predicted ligand stereoselec-tivity (Fig. (**8**)) and roles of ligand binding residues in TRFR1 were experimentally corroborated to further validate the receptor model [28].

6. CONCLUSIONS

The last two years have brought the scientific community high resolution structural details on how a diffusible ligand is recognized by various class A GPCRs. In some cases (ADRB2, ADRB1), the experimentally-dertermined binding mode of the small molecular-weight ligand has not been a real surprise to experienced GPCR modelers because of the high directionality of protein-ligand interactions and the massive amount of previously-known experimental data on analogous ligands. In one case (AA2AR), the binding mode was almost impredictable from known data and might suggest inexperienced modelers that predicting fine details of GPCR-ligand interactions is out of reach. The peculiar nature of the Adenosine A2A receptor, notably the exceptional features of its second extracellular loop makes this example more an exception than the rule. The recently-described structure of the ligand-free opsin also provides interesting hints for ligand entry and exit from the TM cavity that may be common to many liphophilic GPCR ligands. The availability of multiple structures of the 7-TM helical bundle will undoubtedly facilitate the modeling of sequence-specific structural features of future GPCRs (bents, kinks). Unfortunately, it has not been possible yet to achieve this level of fine molecular details with receptors in an activated state. Therefore, the structure-based rational design of partial/full GPCR agonists still remains a difficult task in which experimental data are of utmost importance to restrict the number of possible models. We also anticipate that other ligand-binding sites (e.g. allosteric sites [157], receptor dimers [158] , G-protein interface [159, 160]) will be investigated in a near future to provide alternative solutions in the design of GPCR-regulating bioactive com-pounds.

ABBREVIATIONS

GPCR	=	G protein-coupled receptor
TM	=	Transmembrane helix
bRho	=	Bovine rhodopsin
ADRB1	=	Beta 1 adrenergic receptor
ADRB2	=	Beta 2 adrenergic receptor
AA2AR	=	A2A adenosine receptor
Ops*	=	Ligand-free opsin
EM	=	Energy minimization
MD	=	Molecular dynamics
CRFR1	=	Corticotropin-releasing factor receptor 1
ecl	=	Extracellular loop
SDM	=	Site-directed mutagenesis
SAR	=	Structure-activity relationships
S1PR1	=	Sphingosine 1-phosphate receptor 1
VS	=	Virtual screening
IFP	=	Interaction fingerprint

REFERENCES

[1]   Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G. GPCRDB information system for G protein-coupled receptors. Nucleic Acids Res 2003; 31: 294-7.

[2]   Kristiansen K. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. Pharmacol Ther 2004; 103: 21-80.

[3]   Klabunde T, Hessler G. Drug design strategies for targeting G-protein-coupled receptors. Chembiochem 2002; 3: 928-44.

[4]   Gether U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. Endocr Rev 2000; 21: 90-113.

[5]   Bjarnadottir TK, Gloriam DE, Hellstrand SH, Kristiansson H, Fredriksson R, Schioth HB. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. Genomics 2006; 88: 263-73.

[6]   Chung S, Funakoshi T, Civelli O. Orphan GPCR research. Br J Pharmacol 2008; 153 Suppl 1: S339-46.

[7]   Freddolino PL, Kalani MY, Vaidehi N, Floriano WB, Hall SE, Trabanino RJ, *et al*. Predicted 3D structure for the human beta 2 adrenergic receptor and its binding site for agonists and antagonists. Proc Natl Acad Sci U.S.A. 2004; 101: 2736-41.

[8]   Shacham S, Marantz Y, Bar-Haim S, Kalid O, Warshaviak D, Avisar N, *et al*. PREDICT modeling and in-silico screening for G-protein coupled receptors. Proteins 2004; 57: 51-86.

[9]   Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, *et al*. Crystal structure of rhodopsin: A G protein-coupled receptor. Science 2000; 289: 739-45.

[10]   Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, *et al*. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science 2007; 318: 1258-65.

[11]   Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola VP, Chien EY, *et al*. A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor. Structure 2008; 16: 897-905.

[12]   Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, *et al*. The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. Science 2008; 322: 1211-7.

[13]   Park JH, Scheerer P, Hofmann KP, Choe HW, Ernst OP. Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature 2008; 454: 183-7.

[14]   Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, *et al*. Crystal structure of opsin in its G-protein-interacting conformation. Nature 2008; 455: 497-502.

[15]   Murakami M, Kouyama T. Crystal structure of squid rhodopsin. Nature 2008; 453: 363-7.

[16]   Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, *et al*. Structure of a beta1-adrenergic G-protein-coupled receptor. Nature 2008; 454: 486-91.

[17]   Topiol S, Sabio M. Use of the X-ray structure of the Beta2-adrenergic receptor for drug discovery. Bioorg Med Chem Lett 2008; 18: 1598-602.

[18]   Martinelli A, Tuccinardi T. Molecular modeling of adenosine receptors: new results and trends. Med Res Rev 2008; 28: 247-77.

[19]   Yuzlenko O, Kiec-Kononowicz K. Molecular modeling of A(1) and A(2A) adenosine receptors: Comparison of rhodopsin- and beta(2)-adrenergic-based homology models through the docking studies. J Comput Chem 2008.

[20]   Bissantz C, Bernard P, Hibert M, Rognan D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? Proteins 2003; 50: 5-25.

[21]   Gouldson PR, Kidley NJ, Bywater RP, Psaroudakis G, Brooks HD, Diaz C, *et al*. Toward the active conformations of rhodopsin and the beta2-adrenergic receptor. Proteins 2004; 56: 67-84.

[22]   Kobilka B, Schertler GF. New G-protein-coupled receptor crystal structures: insights and limitations. Trends Pharmacol Sci 2008; 29: 79-83.

[23]   Salo OM, Raitio KH, Savinainen JR, Nevalainen T, Lahtela-Kakkonen M, Laitinen JT, *et al*. Virtual screening of novel CB2 ligands using a comparative model of the human cannabinoid CB2 receptor. J Med Chem 2005; 48: 7166-71.

[24]   Rognan D. In *Ligand Design for G Protein-coupled Receptors*, D. Rognan, ed.; Wiley-VCH: Weinheim, **2006**, pp. 183-203.

[25]   Jones SW, Brockbank SM, Mobbs ML, Le Good NJ, Soma-Haddrick S, Heuze AJ, *et al*. The orphan G-protein coupled receptor RDC1: evidence for a role in chondrocyte hypertrophy and

articular cartilage matrix turnover. Osteoarthritis Cartilage 2006; 14: 597-608.

[26]   Kellenberger E, Springael JY, Parmentier M, Hachet-Haas M, Galzi JL, Rognan D. Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening. J Med Chem 2007; 50: 1294-303.

[27]   Tikhonova IG, Sum CS, Neumann S, Engel S, Raaka BM, Costanzi S, *et al*. Discovery of novel agonists and antagonists of the free fatty acid receptor 1 (FFAR1) using virtual screening. J Med Chem 2008; 51: 625-33.

[28]   Engel S, Skoumbourdis AP, Childress J, Neumann S, Deschamps JR, Thomas CJ, *et al*. A Virtual Screen for Diverse Ligands: Discovery of Selective G Protein-Coupled Receptor Antagonists. J Am Chem Soc 2008; 130: 5115-23.

[29]   Kiss R, Kiss B, Konczol A, Szalai F, Jelinek I, Laszlo V, *et al*. Discovery of novel human histamine H4 receptor ligands by large-scale structure-based virtual screening. J Med Chem 2008; 51: 3145-53.

[30]   de Graaf C, Rognan D. Selective structure-based virtual screening for full and partial agonists of the beta2 adrenergic receptor. J Med Chem 2008; 51: 4978-85.

[31]   Shi L, Liapakis G, Xu R, Guarnieri F, Ballesteros JA, Javitch JA. Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. J Biol Chem 2002; 277: 40989-96.

[32]   Schwartz TW, Frimurer TM, Holst B, Rosenkilde MM, Elling CE. Molecular mechanism of 7TM receptor activation--a global toggle switch model. Annu Rev Pharmacol Toxicol 2006; 46: 481-519.

[33]   Jongejan A, Bruysters M, Ballesteros JA, Haaksma E, Bakker RA, Pardo L, *et al*. Linking agonist binding to histamine H1 receptor activation. Nat Chem Biol 2005; 1: 98-103.

[34]   Bissantz C, Logean A, Rognan D. High-throughput modeling of human G-protein coupled receptors: amino acid sequence alignment, three-dimensional model building, and receptor library screening. J Chem Inf Comput Sci 2004; 44: 1162-76.

[35]   Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. J Mol Biol 2004; 342: 571-83.

[36]   Li J, Edwards PC, Burghammer M, Villa C, Schertler GF. Structure of bovine rhodopsin in a trigonal crystal form. J Mol Biol 2004; 343: 1409-38.

[37]   Nakamichi H, Okada T. Local peptide movement in the photoreaction intermediate of rhodopsin. Proc Natl Acad Sci U.S.A. 2006; 103: 12729-34.

[38]   Salom D, Lodowski DT, Stenkamp RE, Le Trong I, Golczak M, Jastrzebska B, *et al*. Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. Proc Natl Acad Sci U.S.A. 2006; 103: 16123-8.

[39]   Ballesteros J, Weinstein H. Integrated methods for the construction of three-dimensional models and computational probing pf structure-function relations of G protein-coupled receptors. Methods Neurosci 1995; 25: 366-428.

[40]   Surgand JS, Rodrigo J, Kellenberger E, Rognan D. A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. Proteins 2006; 62: 509-38.

[41]   Deupi X, Dolker N, Lopez-Rodriguez ML, Campillo M, Ballesteros JA, Pardo L. Structural models of class a G protein-coupled receptors as a tool for drug design: insights on transmembrane bundle plasticity. Curr Top Med Chem 2007; 7: 991-8.

[42]   Pardo L, Deupi X, Govaerts C, Campillo M. In: Rognan D Ed, Ligand Design for G Protein-coupled Receptors D. Weinheim, Wiley-VCH. 2006, 183-203.

[43]   Ballesteros JA, Shi L, Javitch JA. Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors. Mol Pharmacol 2001; 60: 1-19.

[44]   Altenbach C, Kusnetzow AK, Ernst OP, Hofmann KP, Hubbell WL. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. Proc Natl Acad Sci U.S.A. 2008; 105: 7439-44.

[45]   Kobilka B. Agonist binding: a multistep process. Mol Pharmacol 2004; 65: 1060-2.

[46]   de Graaf C, Foata N, Engkvist O, Rognan D. Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening. Proteins 2008; 71: 599-620.

[47]   Bhattacharya S, Hall SE, Li H, Vaidehi N. Ligand-stabilized conformational states of human beta(2) adrenergic receptor: insight into G-protein-coupled receptor activation. Biophys J 2008; 94: 2027-42.

[48]   Becker OM, Dhanoa DS, Marantz Y, Chen D, Shacham S, Cheruku S, *et al*. An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression. J Med Chem 2006; 49: 3116-35.

[49]   Becker OM, Marantz Y, Shacham S, Inbal B, Heifetz A, Kalid O, *et al*. G protein-coupled receptors: in silico drug discovery in 3D. Proc Natl Acad Sci U.S.A. 2004; 101: 11304-9.

[50]   Cavasotto CN, Orry AJ, Murgolo NJ, Czarniecki MF, Kocsi SA, Hawes BE, *et al*. Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. J Med Chem 2008; 51: 581-8.

[51]   Costanzi S. On the applicability of GPCR homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the beta2-adrenergic receptor. J Med Chem 2008; 51: 2907-14.

[52]   Strader CD, Sigal IS, Candelore MR, Rands E, Hill WS, Dixon RA. Conserved aspartic acid residues 79 and 113 of the beta-adrenergic receptor have different roles in receptor function. J Biol Chem 1988; 263: 10267-71.

[53]   Strader CD, Sigal IS, Register RB, Candelore MR, Rands E, Dixon RA. Identification of residues required for ligand binding to the beta-adrenergic receptor. Proc Natl Acad Sci U.S.A. 1987; 84: 4384-8.

[54]   Liapakis G, Ballesteros JA, Papachristou S, Chan WC, Chen X, Javitch JA. The forgotten serine. A critical role for Ser-2035.42 in ligand binding to and activation of the beta 2-adrenergic receptor. J Biol Chem 2000; 275: 37779-88.

[55]   Suryanarayana S, Kobilka BK. Amino acid substitutions at position 312 in the seventh hydrophobic segment of the beta 2-adrenergic receptor modify ligand-binding specificity. Mol Pharmacol 1993; 44: 111-4.

[56]   Strader CD, Candelore MR, Hill WS, Sigal IS, Dixon RA. Identification of two serine residues involved in agonist activation of the beta-adrenergic receptor. J Biol Chem 1989; 264: 13572-8.

[57]   Wieland K, Zuurmond HM, Krasel C, Ijzerman AP, Lohse MJ. Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta 2-adrenergic receptor. Proc Natl Acad Sci U.S.A. 1996; 93: 9276-81.

[58]   Kikkawa H, Isogaya M, Nagao T, Kurose H. The role of the seventh transmembrane region in high affinity binding of a beta 2-selective agonist TA-2005. Mol Pharmacol 1998; 53: 128-34.

[59]   Jiang Q, Van Rhee AM, Kim J, Yehle S, Wess J, Jacobson KA. Hydrophilic side chains in the third and seventh transmembrane helical domains of human A2A adenosine receptors are required for ligand recognition. Mol Pharmacol 1996; 50: 512-21.

[60]   Kim J, Wess J, van Rhee AM, Schoneberg T, Jacobson KA. Site-directed mutagenesis identifies residues involved in ligand recognition in the human A2a adenosine receptor. J Biol Chem 1995; 270: 13987-97.

[61]   Kim J, Jiang Q, Glashofer M, Yehle S, Wess J, Jacobson KA. Glutamate residues in the second extracellular loop of the human A2a adenosine receptor are required for ligand recognition. Mol Pharmacol 1996; 49: 683-91.

[62]   Moro S, Gao ZG, Jacobson KA, Spalluto G. Progress in the pursuit of therapeutic adenosine receptor antagonists. Med Res Rev 2006; 26: 131-59.

[63]   Pardo L, Deupi X, Dolker N, Lopez-Rodriguez ML, Campillo M. The role of internal water molecules in the structure and function of the rhodopsin family of G protein-coupled receptors. Chembiochem 2007; 8: 19-24.

[64]   Jiang Q, Lee BX, Glashofer M, van Rhee AM, Jacobson KA. Mutagenesis reveals structure-activity parallels between human A2A adenosine receptors and biogenic amine G protein-coupled receptors. J Med Chem 1997; 40: 2588-95.

[65]   Horn F, Lau AL, Cohen FE. Automated extraction of mutation data from the literature: application of MuteXt to G protein-coupled receptors and nuclear hormone receptors. Bioinformatics 2004; 20: 557-68.

[66]   Moller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 2001; 17: 646-53.

[67]   Mirzadegan T, Benko G, Filipek S, Palczewski K. Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. Biochemistry 2003; 42: 2759-67.

[68]   Attwood TK. A compendium of specific motifs for diagnosing GPCR subtypes. Trends Pharmacol Sci 2001; 22: 162-5.

[69]   Bondensgaard K, Ankersen M, Thogersen H, Hansen BS, Wulff BS, Bywater RP. Recognition of privileged structures by G-protein coupled receptors. J Med Chem 2004; 47: 888-99.

[70]   Frimurer TM, Bywater RP. Structure of the integral membrane domain of the GLP1 receptor. Proteins 1999; 35: 375-86.

[71]   Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 2003; 63: 1256-72.

[72]   Petrel C, Kessler A, Dauban P, Dodd RH, Rognan D, Ruat M. Positive and negative allosteric modulators of the Ca2+-sensing receptor interact within overlapping but not identical binding sites in the transmembrane domain. J Biol Chem 2004; 279: 18990-7.

[73]   Bu L, Michino M, Wolf RM, Brooks CL, 3rd. Improved model building and assessment of the Calcium-sensing receptor transmembrane domain. Proteins 2008; 71: 215-26.

[74]   Malherbe P, Kratochwil N, Knoflach F, Zenner MT, Kew JN, Kratzeisen C, et al. Mutational analysis and molecular modeling of the allosteric binding site of a novel, selective, noncompetitive antagonist of the metabotropic glutamate 1 receptor. J Biol Chem 2003; 278: 8340-7.

[75]   Malherbe P, Kratochwil N, Muhlemann A, Zenner MT, Fischer C, Stahl M, et al. Comparison of the binding pockets of two chemically unrelated allosteric antagonists of the mGlu5 receptor and identification of crucial residues involved in the inverse agonism of MPEP. J Neurochem 2006; 98: 601-15.

[76]   Malherbe P, Kratochwil N, Zenner MT, Piussi J, Diener C, Kratzeisen C, et al. Mutational analysis and molecular modeling of the binding pocket of the metabotropic glutamate 5 receptor negative modulator 2-methyl-6-(phenylethynyl)-pyridine. Mol Pharmacol 2003; 64: 823-32.

[77]   Muhlemann A, Ward NA, Kratochwil N, Diener C, Fischer C, Stucki A, et al. Determination of key amino acids implicated in the actions of allosteric modulation by 3,3'-difluorobenzaldazine on rat mGlu5 receptors. Eur J Pharmacol 2006; 529: 95-104.

[78]   Pagano A, Ruegg D, Litschig S, Stoehr N, Stierlin C, Heinrich M, et al. The non-competitive antagonists 2-methyl-6-(phenyl-ethynyl)pyridine and 7-hydroxyiminocyclopropan[b]chromen-1a-carboxylic acid ethyl ester interact with overlapping binding pockets in the transmembrane region of group I metabotropic glutamate receptors. J Biol Chem 2000; 275: 33750-8.

[79]   Gardella TJ, Luck MD, Fan MH, Lee C. Transmembrane residues of the parathyroid hormone (PTH)/PTH-related peptide receptor that specifically affect binding and signaling by agonist ligands. J Biol Chem 1996; 271: 12820-5.

[80]   Di Paolo E, Vilardaga JP, Petry H, Moguilevsky N, Bollen A, Robberecht P, et al. Role of charged amino acids conserved in the vasoactive intestinal polypeptide/secretin family of receptors on the secretin receptor functionality. Peptides 1999; 20: 1187-93.

[81]   Mathi SK, Chan Y, Li X, Wheeler MB. Scanning of the glucagon-like peptide-1 receptor localizes G protein-activating determinants primarily to the N terminus of the third intracellular loop. Mol Endocrinol 1997; 11: 424-32.

[82]   Binet V, Duthey B, Lecaillon J, Vol C, Quoyer J, Labesse G, et al. Common structural requirements for heptahelical domain function in class A and class C G protein-coupled receptors. J Biol Chem 2007; 282: 12154-63.

[83]   Lee EJ, Kotlar TJ, Ciric I, Lee MK, Lim SK, Lee HC, et al. Absence of constitutively activating mutations in the GHRH receptor in GH-producing pituitary tumors. J Clin Endocrinol Metab 2001; 86: 3989-95.

[84]   Petrel C, Kessler A, Maslah F, Dauban P, Dodd RH, Rognan D, et al. Modeling and mutagenesis of the binding site of Calhex 231, a novel negative allosteric modulator of the extracellular Ca(2+)-sensing receptor. J Biol Chem 2003; 278: 49487-94.

[85]   Sheikh SP, Vilardarga JP, Baranski TJ, Lichtarge O, Iiri T, Meng EC, et al. Similar structures and shared switch mechanisms of the beta2-adrenoceptor and the parathyroid hormone receptor. Zn(II)

[86]   bridges between helices III and VI block activation. J Biol Chem 1999; 274: 17033-41.

[86]   Kiss R, Noszal B, Racz A, Falus A, Eros D, Keseru GM. Binding mode analysis and enrichment studies on homology models of the human histamine H4 receptor. Eur J Med Chem 2008; 43: 1059-70.

[87]   Jongejan A, Lim HD, Smits RA, de Esch IJ, Haaksma E, Leurs R. Delineation of agonist binding to the human histamine H4 receptor using mutational analysis, homology modeling, and ab initio calculations. J Chem Inf Model 2008; 48: 1455-63.

[88]   Hoare SR, Brown BT, Santos MA, Malany S, Betz SF, Grigoriadis DE. Single amino acid residue determinants of non-peptide antagonist binding to the corticotropin-releasing factor1 (CRF1) receptor. Biochem Pharmacol 2006; 72: 244-55.

[89]   Kondru R, Zhang J, Ji C, Mirzadegan T, Rotstein D, Sankuratri S, et al. Molecular interactions of CCR5 with major classes of small-molecule anti-HIV CCR5 antagonists. Mol Pharmacol 2008; 73: 789-800.

[90]   Paiva AC, Oliveira L, Horn F, Bywater RP, Vriend G. Modeling GPCRs. Ernst Schering Found Symp Proc 2006: 23-47.

[91]   Parrill AL. Lysophospholipid interactions with protein targets. Biochim Biophys Acta 2008; 1781: 540-6.

[92]   Wang DA, Lorincz Z, Bautista DL, Liliom K, Tigyi G, Parrill AL. A single amino acid determines lysophospholipid specificity of the S1P1 (EDG1) and LPA1 (EDG2) phospholipid growth factor receptors. J Biol Chem 2001; 276: 49213-20.

[93]   Kiuchi M, Adachi K, Tomatsu A, Chino M, Takeda S, Tanaka Y, et al. Asymmetric synthesis and biological evaluation of the enantiomeric isomers of the immunosuppressive FTY720-phosphate. Bioorg Med Chem 2005; 13: 425-32.

[94]   Fujiwara Y, Osborne DA, Walker MD, Wang DA, Bautista DA, Liliom K, et al. Identification of the hydrophobic ligand binding pocket of the S1P1 receptor. J Biol Chem 2007; 282: 2374-85.

[95]   Deng Q, Clemas JA, Chrebet G, Fischer P, Hale JJ, Li Z, et al. Identification of Leu276 of the S1P1 receptor and Phe263 of the S1P3 receptor in interaction with receptor specific agonists by molecular modeling, site-directed mutagenesis, and affinity studies. Mol Pharmacol 2007; 71: 724-35.

[96]   Ceccarelli SM, Jaeschke G, Buettelmann B, Huwyler J, Kolczewski S, Peters JU, et al. Rational design, synthesis, and structure-activity relationship of benzoxazolones: new potent mglu5 receptor antagonists based on the fenobam structure. Bioorg Med Chem Lett 2007; 17: 1302-6.

[97]   Gilligan PJ, Robertson DW, Zaczek R. Corticotropin releasing factor (CRF) receptor modulators: progress and opportunities for new therapeutic agents. J Med Chem 2000; 43: 1641-60.

[98]   Hodge CN, Aldrich PE, Wasserman ZR, Fernandez CH, Nemeth GA, Arvanitis A, et al. Corticotropin-releasing hormone receptor antagonists: framework design and synthesis guided by ligand conformational studies. J Med Chem 1999; 42: 819-32.

[99]   Chen C, Dagnino R, Jr., Huang CQ, McCarthy JR, Grigoriadis DE. 1-Alkyl-3-amino-5-aryl-1H-[1,2,4]triazoles: novel synthesis via cyclization of N-acyl-S-methylisothioureas with alkylhydrazines and their potent corticotropin-releasing factor-1 (CRF(1)) receptor antagonist activities. Bioorg Med Chem Lett 2001; 11: 3165-8.

[100]   Liaw CW, Grigoriadis DE, Lorang MT, De Souza EB, Maki RA. Localization of agonist- and antagonist-binding domains of human corticotropin-releasing factor receptors. Mol Endocrinol 1997; 11: 2048-53.

[101]   Radestock S, Weil T, Renner S. Homology Model-Based Virtual Screening for GPCR Ligands Using Docking and Target-Biased Scoring. J Chem Inf Model 2008; 48: 1104-17.

[102]   Chen JZ, Wang J, Xie XQ. GPCR structure-based virtual screening approach for CB2 antagonist search. J Chem Inf Model 2007; 47: 1626-37.

[103]   Evers A, Klabunde T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. J Med Chem 2005; 48: 1088-97.

[104]   Evers A, Klebe G. Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. J Med Chem 2004; 47: 5381-92.

[105]   Varady J, Wu X, Fang X, Min J, Hu Z, Levant B, et al. Molecular modeling of the three-dimensional structure of dopamine 3 (D3) subtype receptor: discovery of novel and potent D3 ligands through a hybrid pharmacophore- and structure-based database searching approach. J Med Chem 2003; 46: 4377-92.

[106] Urizar E, Claeysen S, Deupi X, Govaerts C, Costagliola S, Vassart G, *et al*. An activation switch in the rhodopsin family of G protein-coupled receptors: the thyrotropin receptor. J Biol Chem 2005; 280: 17135-41.

[107] Kimura SR, Tebben AJ, Langley DR. Expanding GPCR homology model binding sites *via* a balloon potential: A molecular dynamics refinement approach. Proteins 2008; 71: 1919-29.

[108] Evers A, Gohlke H, Klebe G. Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. J Mol Biol 2003; 334: 327-45.

[109] Yohannan S, Faham S, Yang D, Whitelegge JP, Bowie JU. The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. Proc Natl Acad Sci U.S.A. 2004; 101: 959-63.

[110] Lopez-Rodriguez ML, Vicente B, Deupi X, Barrondo S, Olivella M, Morcillo MJ, *et al*. Design, synthesis and pharmacological evaluation of 5-hydroxytryptamine(1a) receptor ligands to explore the three-dimensional structure of the receptor. Mol Pharmacol 2002; 62: 15-21.

[111] Govaerts C, Bondue A, Springael JY, Olivella M, Deupi X, Le Poul E, *et al*. Activation of CCR5 by chemokines involves an aromatic cluster between transmembrane helices 2 and 3. J Biol Chem 2003; 278: 1892-903.

[112] Jensen PC, Nygaard R, Thiele S, Elder A, Zhu G, Kolbeck R, *et al*. Molecular interaction of a potent nonpeptide agonist with the chemokine receptor CCR8. Mol Pharmacol 2007; 72: 327-40.

[113] Wong RS, Bodart V, Metz M, Labrecque J, Bridger G, Fricker SP. Comparison of the potential multiple binding modes of bicyclam, monocylam and non-cyclam small-molecule CXCR4 inhibitors. Mol Pharmacol 2008; 74: 1485-95.

[114] Bailey RJ, Hay DL. Agonist-dependent consequences of proline to alanine substitution in the transmembrane helices of the calcitonin receptor. Br J Pharmacol 2007; 151: 678-87.

[115] Conner AC, Hay DL, Simms J, Howitt SG, Schindler M, Smith DM, *et al*. A key role for transmembrane prolines in calcitonin receptor-like receptor agonist binding and signalling: implications for family B G-protein-coupled receptors. Mol Pharmacol 2005; 67: 20-31.

[116] Zhang R, Hurst DP, Barnett-Norris J, Reggio PH, Song ZH. Cysteine 2.59(89) in the second transmembrane domain of human CB2 receptor is accessible within the ligand binding crevice: evidence for possible CB2 de*via*tion from a rhodopsin template. Mol Pharmacol 2005; 68: 69-83.

[117] Salo OM, Lahtela-Kakkonen M, Gynther J, Jarvinen T, Poso A. Development of a 3D model for the human cannabinoid CB1 receptor. J Med Chem 2004; 47: 3048-57.

[118] Ballesteros JA, Jensen AD, Liapakis G, Rasmussen SG, Shi L, Gether U, *et al*. Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. J Biol Chem 2001; 276: 29171-7.

[119] Farrens DL, Altenbach C, Yang K, Hubbell WL, Khorana HG. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. Science 1996; 274: 768-70.

[120] Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br J Pharmacol 2008; 153 Suppl 1: S7-26.

[121] Bissantz C, Folkers G, Rognan D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem 2000; 43: 4759-67.

[122] Kellenberger E, Rodrigo J, Müller P, Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins 2004; 57: 225-42.

[123] Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP. Comparison of automated docking programs as virtual screening tools. J Med Chem 2005; 48: 962-76.

[124] Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, *et al*. A critical assessment of docking programs and scoring functions. J Med Chem 2006; 49: 5912-31.

[125] Kramer B, Rarey M, Lengauer T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins 1999; 37: 228-41.

[126] Fischer B, Merlitz H, Wenzel W. Receptor flexibility for large-scale in silico ligand screens: chances and challenges. Methods Mol Biol 2008; 443: 353-64.

[127] Kämper A, Rognan D, Lengauer T. In: Lengauer T Ed, Bioinformatics- From Genomes to Therapies. Weinheim, Wiley-VCH. 2007; 651-704.

[128] Marcou G, Rognan D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J Chem Inf Model 2007; 47: 195-207.

[129] Kellenberger E, Foata N, Rognan D. Ranking targets in structure-based virtual screening of three-dimensional protein libraries: methods and problems. J Chem Inf Model 2008; 48: 1014-25.

[130] Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997; 267: 727-48.

[131] Jain AN. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. J Comput Aided Mol Des 2007; 21: 281-306.

[132] Tirado-Rives J, Jorgensen WL. Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding. J Med Chem 2006; 49: 5880-4.

[133] Venhorst J, Nunez S, Terpstra JW, Kruse CG. Assessment of Scaffold Hopping Efficiency by Use of Molecular Interaction Fingerprints. J Med Chem 2008; 51: 3222-9.

[134] Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993; 234: 779-815.

[135] Hann M, Hudson B, Lewell X, Lifely R, Miller L, Ramsden N. Strategic pooling of compounds for high-throughput screening. J Chem Inf Comput Sci 1999; 39: 897-902.

[136] Muegge I. Selection criteria for drug-like compounds. Med Res Rev 2003; 23: 302-21.

[137] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 2001; 46: 3-26.

[138] Höltje H-D, Sippl W, Rognan D, Folkers G. Molecular Modeling: Basics, Principles and Applications. Weinheim, Wiley-VCH. 2008.

[139] Stahura FL, Bajorath J. Virtual screening methods that complement HTS. Comb Chem High Throughput Screen 2004; 7: 259-69.

[140] Langer T, Hoffmann RD. Pharmacophores and Pharmacophore Searches, Weinheim. Wiley-VCH, 2008.

[141] Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. Proteins 2006; 65: 15-26.

[142] Evers A, Hessler G, Matter H, Klabunde T. Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. J Med Chem 2005; 48: 5448-65.

[143] Hindle SA, Rarey M, Buning C, Lengaue T. Flexible docking under pharmacophore type constraints. J Comput Aided Mol Des 2002; 16: 129-49.

[144] Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD, *et al*. Virtual screening using protein-ligand docking: avoiding artificial enrichment. J Chem Inf Comput Sci 2004; 44: 793-806.

[145] Paul N, Rognan D. ConsDock: A new program for the consensus analysis of protein-ligand interactions. Proteins 2002; 47: 521-33.

[146] Vigers GP, Rizzi JP. Multiple active site corrections for docking and virtual screening. J Med Chem 2004; 47: 80-9.

[147] Stahl M, Bohm HJ. Development of filter functions for protein-ligand docking. J Mol Graph Model 1998; 16: 121-32.

[148] Deng Z, Chuaqui C, Singh J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. J Med Chem 2004; 47: 337-44.

[149] Krier M, Bret G, Rognan D. Assessing the scaffold diversity of screening libraries. J Chem Inf Model 2006; 46: 512-24.

[150] Edwards BS, Bologa C, Young SM, Balakin KV, Prossnitz ER, Savchuck NP, *et al*. Integration of virtual screening with high-throughput flow cytometry to identify novel small molecule formylpeptide receptor antagonists. Mol Pharmacol 2005; 68: 1301-10.

[151] Barillari C, Marcou G, Rognan D. Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. J Chem Inf Model 2008; 48: 1396-410.

[152] Tikhonova IG, Sum CS, Neumann S, Thomas CJ, Raaka BM, Costanzi S, *et al*. Bidirectional, Iterative Approach to the Structural Delineation of the Functional "Chemoprint" in GPR40 for Agonist Recognition. J Med Chem 2007; 50: 2981-9.

[153] Liu Y, Zhou E, Yu K, Zhu J, Zhang Y, Xie X, *et al.* Discovery of a novel CCR5 antagonist lead compound through fragment assembly. Molecules 2008; 13: 2426-41.

[154] Bayry J, Tchilian EZ, Davies MN, Forbes EK, Draper SJ, Kaveri SV, *et al.* In silico identified CCR4 antagonists target regulatory T cells and exert adjuvant activity in vaccination. Proc Natl Acad Sci U.S.A. 2008; 105: 10221-6.

[155] Colson AO, Perlman JH, Smolyar A, Gershengorn MC, Osman R. Static and dynamic roles of extracellular loops in G-protein-coupled receptors: a mechanism for sequential binding of thyrotropin-releasing hormone to its receptor. Biophys J 1998; 74: 1087-100.

[156] Sabio M, Jones K, Topiol S. Use of the X-ray structure of the beta(2)-adrenergic receptor for drug discovery. Part 2: Identification of active compounds. Bioorg Med Chem Lett 2008; 18: 5391-5.

[157] Valant C, Gregory KJ, Hall NE, Scammells PJ, Lew MJ, Sexton PM, *et al.* A novel mechanism of G protein-coupled receptor functional selectivity. Muscarinic partial agonist McN-A-343 as a bitopic orthosteric/allosteric ligand. J Biol Chem 2008; 283: 29312-21.

[158] Panetta R, Greenwood MT. Physiological relevance of GPCR oligomerization and its impact on drug discovery. Drug Discov Today 2008; 13: 1059-66.

[159] Taylor CM, Barda Y, Kisselev OG, Marshall GR. Modulating G-protein coupled receptor/G-protein signal transduction by small molecules suggested by virtual screening. J Med Chem 2008; 51: 5297-303.

[160] Smrcka AV, Lehmann DM, Dessal AL. G Protein betagamma Subunits as Targets for Small Molecule Therapeutic Develop-ment. Comb Chem High Throughput Screen 2008; 11: 382-95.

[161] Clark DE, Higgs C, Wren SP, Dyke HJ, Wong M, Norman D *et al.,* A virtual screening approach to finding novel and potent antagonists at the melanin-concentrating hormone 1 receptor. J Med Chem 2004; 47, 3962-71.