PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening

Chris de Graaf,[1] Nicolas Foata,[1] Ola Engkvist,[2] and Didier Rognan[1]*

[1] Bioinformatics of the Drug, CNRS UMR 7175-LC1, Université Louis Pasteur Strasbourg I, Illkirch F-67401, France
[2] Global Compound Sciences, Lead Generation DECS, AstraZeneca R&D Mölndal, Pepparedsleden 1, Mölndal 431 83, Sweden

ABSTRACT

The current study describes the validation of high-throughput modeling procedures for the construction of the second extracellular loop (ecl2) of all nonolfactory human G Protein-coupled receptors. Our modeling flowchart is based on the alignment of essential residues determining the particular ecl2 fold observed in the bovine rhodopsin (bRho) crystal structure. For a set of GPCR targets, the dopamine D2 receptor (DRD2), adenosine A3 receptor (AA3R), and the thromboxane A2 receptor (TA2R), the implications of including ecl2 atomic coordinates is evaluated in terms of structure-based virtual screening accuracy: the suitability of the 3D models to distinguish between known antagonists and randomly chosen decoys using automated docking approaches. The virtual screening results of different models describing increasingly exhaustive receptor representations (seven helices only, seven helices and ecl2 loop, full model) have been compared. Explicit modeling of the ecl2 loop was found to be important in only one of three test cases whereas a loopless model was shown to be accurate enough in the two other receptors. An exhaustive comparison of ecl2 loops of 365 receptors to that of bRho suggests that explicit ecl2 loop modeling should be reserved to receptors where loop building can be guided by experimental restraints.

Proteins 2008; 71:599–620.

Key words: homology modeling; second extracellular loop; automated docking; virtual screening; interaction fingerprint scoring; G-protein coupled receptors.

INTRODUCTION

G protein-coupled receptors (GPCRs) constitute a superfamily of transmembrane proteins of utmost pharmaceutical importance. 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.[1] GPCRs share a common topology, with an extracellular N terminus, a cytoplasmic C terminus, and 7 transmembrane helices (TMs) connected by 3 intracellular (icls) and 3 extracellular loops (ecls).[2,3] The importance of ecls for accommodating high molecular weight GPCR ligands (peptides and proteins) is widely accepted,[4] but recent studies indicate that not only the 7-TM domain[5], but also the ecls (and specifically ecl2) in GPCRs can play an important role in binding lower molecular weight (drug-like) ligands[6,7] (additional references in the Supplementary Table I). Specifically, in the bovine Rhodopsin (bRho) crystal structure,[8,9] 6 of the 24 residues in close contact (<5 Å) to *cis*-retinal are located in ecl2 (Fig. 1). The ecl2 of bRho is structured in an antiparallel β-sheet and deeply folds into the centre of the TM receptor core. This tight fold of ecl2 into the TM cavity is facilitated by a disulfide link between two highly conserved cysteines located in ecl2 (C45.50) and TM3 (C3.25) in nearly all GPCRs[12] (see residue numbering scheme in the "Computational Methods" section). Numerous studies have shown that this disulfide bond is critical for GPCR folding and surface localization and affects ligand binding (Supplementary Table I). Although it has been argued that the presence of ecl2 in the transmembrane domain may be a feature unique to bRho,[13,14] studies based on the substituted-cysteine accessibility method (SCAM) have provided evidence for a comparable ecl2 fold in the dopamine D2 (DRD2) receptor.[6] Furthermore, many site-directed mutagenesis studies have reported the effect of mutations of ecl2 residues other than C45.50 on GPCR

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887–3585/suppmat.
Grant sponsor: AstraZeneca (Mölndal, Sweden).
*Correspondence to: Didier Rognan, Bioinformatics of the Drug, CNRS UMR 7175-LC1, Université Louis Pasteur Strasbourg I, 74 route du Rhin, Illkirch F-67401, France.
E-mail: didier.rognan@pharma.u-strasbg.fr
Received 3 May 2007; Revised 3 July 2007; Accepted 11 July 2007
Published online 30 October 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21724

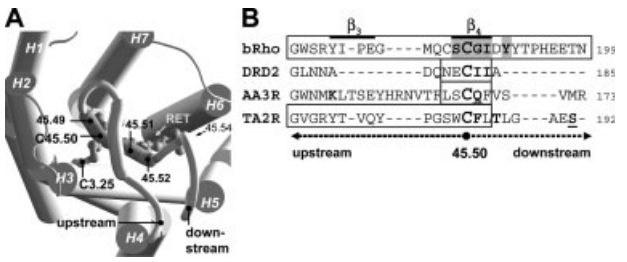


Figure 1

*The second extracellular loop (between transmembrane helices (H) 4 and 5, panel **A**) of bRho is structured in an antiparallel β-sheet (thickened coil) and contains 13 residues upstream and 12 residues downstream of C45.50, the cysteine residue which forms a disulfide bond with C3.25. Five out of the 16 residues in close contact (up to 4 Å) to retinal (depicted in balls- and -sticks) are located in the ecl2 of bRho (at positions 45.49, upstream, and 45.51, 45.52, and 45.54 downstream, and C45.50 itself). Seven TMs are depicted as cylinders. N-terminus (on top of ecl2), helix 8, and all icls are not depicted. The ecl2s of the three GPCR ecl2 modeling test cases described in the current study are aligned with the ecl2 of bRho in panel **B**, and the number of the residue at the downstream end is indicated. Residues reported to be involved in antagonist binding are in bold, and the portions of the ecl2 modelled according to the flowchart depicted in Figure 2 are boxed. The Figure in panel A (as well as Figs. 5–7 and 9) has been prepared with MOLSCRIPT[10] and rendered with Raster3D.[11]*

agonist binding as well as antagonist binding. An overview of references describing these site-directed mutagenesis studies is provided in the Supplementary Table I Interestingly, GPCR specific ecl2 residues that were found to be critical for agonist-induced receptor activation did not necessarily play an important role in agonist binding, and often the residues identified to be involved in agonist binding are different from the residues involved in antagonist binding.[15–20] These specific effects on agonist and antagonist binding suggest that these ligands bind to ecl2 via different binding modes, and possibly also to distinct ecl2 loop conformations. This picture is confirmed by recent studies of Baneres *et al.* which showed that agonist binding to the serotonin 5HT4A receptor is associated with rearrangements in ecl2, while antagonist binding does not induce any structural changes of ecl2.[7] In addition to its importance in structural integrity and ligand binding, a novel role for ecl2 as negative regulator of GPCR activation was recently postulated by Klco *et al.*[21] Site-directed mutagenesis studies in the complement C5a receptor (C5AR) indicated that also in the absence of ligands, ecl2 might be involved in stabilizing the inactive state of the receptor.[21]

Our original in-house database of high-throughput human GPCR models,[22] as well as many other GPCR homology models reported in literature[23,24] only include the 7-TMs. Such loopless GPCR models have already been shown to be suitable for *in silico* inverse screening purposes,[22] and for detecting key residues that drive ligand selectivity.[5] Consideration of the ecls in GPCRs might however affect virtual screening accuracy. Modeling ecls is generally achieved either on a case-by-case basis by restrained knowledge-based procedures and further molecular mechanics/dynamics refinement[25–27] or by more sophisticated but low-throughput *ab initio* simulations[28] unfortunately unsuitable for addressing the whole GPCR human proteome. Automated modeling of GPCRs with explicit loop modeling has only been reported in one recent study using a threading assembly refinement method,[29] but its suitability for structure-based design has not been established so far.

The primary aim of the current study was to evaluate the effect of different ecl2 modeling strategies on structure-based virtual screening for GPCR antagonists. We have set up a high-throughput modeling procedure for the construction of the second extracellular loop of human GPCRs. Our loop modeling flowchart is based on the alignment of essential residues determining the particular ecl2 fold observed in the bRho crystal structure. For a set of unrelated GPCR targets, the dopamine D2 receptor (DRD2), the adenosine A3 receptor (AA3R), and the thromboxane A2 receptor (TA2R), the implications of including ecl2 is evaluated in terms of structure-based virtual screening accuracy: the suitability of the 3D models to distinguish between known antagonists and randomly chosen drug-like compounds using automated docking approaches. These three GPCRs were selected because of the availability of solid experimental evidence for the involvement of ecl2 in antagonist binding[6,17,20,30] (see Supplementary Table I), the fact that they represent different GPCR clusters (biogenic amines for DRD2, adenosines for AA3R, prostanoids for TA2R), their wide range in ecl2 loop upstream and downstream lengths (C45.50 being considered here as the central ecl2 residue), and finally their therapeutical relevance.[31–33] The virtual screening results of four different models have been compared: (i) one model including only the seven transmembrane (7-TM) helices; two models containing 7-TM helices and (parts of) the ecl2 constructed using different high-throughput ecl2 modeling procedures; (ii) loop threading (L1); (iii) comparative loop modeling by satisfaction of spatial restraints (L2); (iv) one full GPCR model containing all ecls constructed using a tailored loop modeling procedure including GPCR-specific SAR/pharmacophore and site-directed mutagenesis data (F).

COMPUTATIONAL METHODS

Residue numbering and nomenclature

The Ballesteros–Weinstein residue numbering scheme[34] was used throughout this manuscript for GPCR TM helices. To apply it to the ecl2 loop, corresponding residues have been labeled 45.*x*, 45 indicating the location between TMs 4 and 5 (therefore ecl2) and *x* a number relative to

the conserved cysteine residue which is assigned number 50. Numbers lower than 50 thus describe upstream residues located between TM 4 and C45.50. Numbers higher than 50 describe downstream residues located between C45.50 and TM5. For explicitly numbering ecl2 residue in specific receptors, the UniProt[35] residue number is given before the ecl2 number in superscript (e.g. $C182^{45.50}$ for DRD2).

Ecl2 amino acid sequence alignment and analysis

The 7-TM amino acid sequence of 365 human GPCRs were aligned in earlier studies using the in-house GPCRmod program.[5,22] On the basis of this TM alignment (in which TM4 was assumed to end at position 4.62 and TM5 was assumed to begin at position 5.35), ecl2 sequences where extracted for each receptor from our GPCR database (http://bioinfo-pharma.u-strasbg/hGPCRlig). A full amino acid sequence alignment of all the ecl2s as well as separate sequence alignments of the ecl2s of 22 GPCR clusters[5] were realized with the T-coffee program[36] showed multiple gaps (data not shown). Therefore, ecl2 sequences were first analyzed in terms of the number of upstream and downstream residues from C45.50. For receptors containing more than one cysteine residue in their ecl2, experimental data alone[37–40] or in combination with alignment to other closely related receptors[9,37–48] was used to determine C45.50. For each cluster, receptors with 12–14 residues upstream from C45.50 were aligned against an ecl2 profile alignment of the opsins cluster (alignment "OPSINS" including 9 human opsin receptors and bRho), resulting in alignment "RECEPTOR." For each GPCR cluster, the receptors showing no insertions/deletions in the β3-β4 sheet region in this RECEPTOR alignment, a separate GPCR cluster profile alignment was made (alignment "CLUSTER"). Subsequently, a profile alignment was performed between this CLUSTER alignment and the OPSINS alignment, yielding a CLUSTER_OPSINS_PROFILE alignment.

Receptors not selected by the above-described procedure were addressed considering the feasibility of modeling ecl2. On the basis of the ecl2 sequence alignment and analysis, the 365 human GPCRs in our database were classified according to their modeling feasibility as depicted in Figure 2. In steps 1 and 2, receptors containing a cysteine residue at position 3.25 and a cysteine resi-

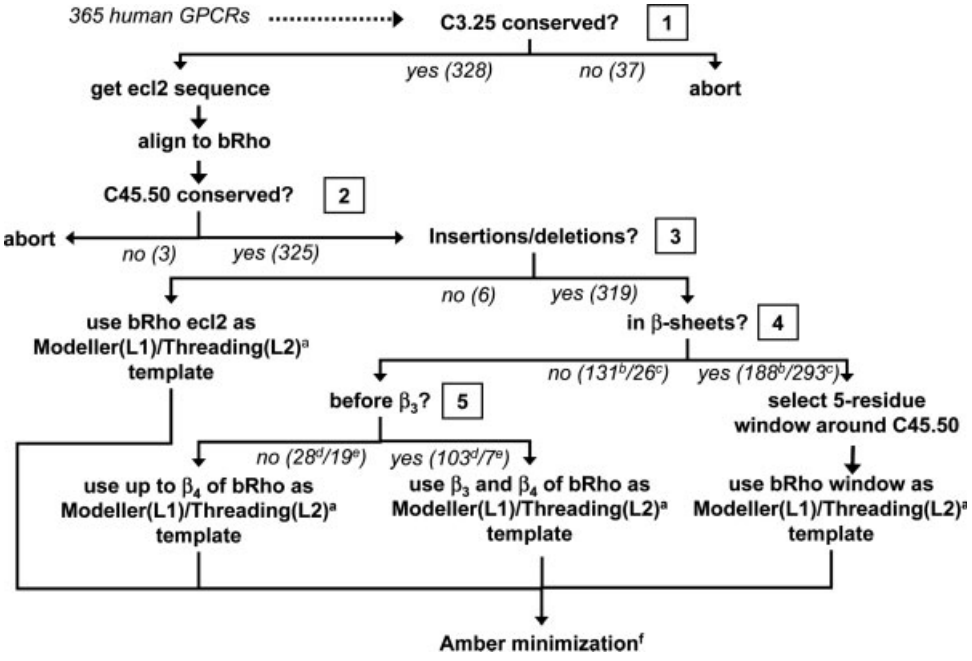


Figure 2

*ecl2 modeling flowchart based on the alignment of essential residues determining the particular ecl2 fold observed in the bRho crystal structure. The number of receptors at each classification step is indicated in parentheses. Legend: (**a**) The "Modeller" and "Threading" loop modeling approaches are described in detail in the "Computational Methods" section; (**b**) strict criterion: GPCR cluster ecl2 profile alignment against opsin ecl2 profile showing no insertions/deletions in β sheets; (**c**) mild criterion: ecl2s containing 12 or 14 residues upstream from C45.50; (**d**) strict criterion: GPCR cluster ecl2 profile alignment against opsin ecl2 profile showing no insertions/deletions before β_3; (**e**) mild criterion: ecl2s containing 13 residues upstream from C45.50; (**f**) The three test cases described in the current study (DRD2, AA3R, TA2R) the final energy minimization is performed with the presence of a known ligand (see Tables III–V) docked under pharmacophore restraints using GOLD, the other 322 GPCRs are minimized without a ligand.*

due located at least three positions from the N-terminal or C-terminal ends were separated from receptors unable to make the conserved TM3-ecl2 disulfide link. Receptors containing exactly the same number of residues upstream and downstream from C45.50 were separated from receptors showing insertions/deletions in their ecl2 alignment with bRho in step 3. The full ecl2 of bRho can be used as a template for modeling the ecl2s of these receptors. In step 4, receptors showing no insertions/deletions in the β-sheet can be identified according to a strict or milder classification criterion. The strict classification criterion demand the receptors to have no insertions/deletions in the β-sheet region (see Fig. 1) of their CLUSTER_OPSINS_PROFILE alignment (see above). The milder classification criterion states that the receptors should have between 12 and 14 residues upstream from C45.50. In step 5, receptors showing no insertions/deletions up to the end of the β_4 strand can be identified, again according to a strict or milder classification criterion. The strict classification criterion demands the receptors to have no insertions/deletions up to the end of β_4 in their CLUSTER_OPSINS_PROFILE alignment. The milder classification criterion states that the receptors should have 13 residues upstream from C45.50. The bRho ecl2 $\beta_3$and β_4 strands can be used for constructing ecl2s passing the strict or mild criteria of step 4, but not passing the criteria of step 5. The bRho ecl2 segment stretching from the beginning of ecl2 up to the end of β_4 can be used for building the ecl2s passing the criteria of both steps 4 and 5. The 5-residue window around C45.50 of bRho can be used for constructing the ecl2s of receptors possessing the TM3-ecl2 disulfide link, but not passing any of the earlier described criteria.

Modeling antagonist-bound GPCR models

A ground-state homology model of the DRD2 receptor has been earlier reported by our group.[22] To get ground-state models of the AA3R and TA2R receptors, we mainly followed a previously-defined five-step protocol.[25,26]

Step 1: construction of a preliminary TM model

First, a preliminary high-throughput receptor model was generated using the GPCRgen program,[22] including only the seven TM helices. The amino acid sequence alignments used for constructing the receptor models are shown in the Supplementary Figure 1. The preliminary AA3R and TA2R models were derived from the previously validated DRD2[25] and OPRX[22] models, respectively, since the later templates were the closest among our template collection[22] to the target structures. For the TA2R receptor, an alternative conformation of TM5 was constructed as the original conformation was not in satisfactorily agreement with site-directed mutagenesis studies of this receptor and other prostanoid receptors.[20,49] This was not considered to be the result of a wrong alignment of TM5 of this receptor class against bRho, but rather the result of structural differences caused by the absence of the conserved P5.50 residue in prostanoid receptors. Despite the lack of the conserved "FxxPxxxxxY" motif in the TM5 of prostanoid receptors,[22] the alignment with a cysteine residue at position 5.57 (shown to be essential for PI2R stability[38]) and an asparagine residue at position 5.58 (a conserved alternative to a tyrosine residue at this position)[12] indicate the validity of the TM5 alignment. TM5 of the prostanoid receptor cluster does not contain the highly conserved proline residue at position 5.50, inducing an opening in TM5 stabilized by a bulky aliphatic residue at position 3.40 in the bRho crystal structure. Instead, TA2R contains glycine residues at positions 5.47, 5.48, and 5.52, offering space for the L3.40 residue. The part of the TM5 helix of TA2R between G5.48 and F5.43 was unkinked by changing φ and ψ angles using the Biopolymer module implemented in the Sybyl 7.2 package,[50] in such a way that the F5.43 obtained approximately the same position as L5.42 in the original TM5 helix while preserving the α-helical structure intact. This preliminary TA2R model was minimized with AMBER 8[51] using the AMBER03 force field to relax the structure and remove steric bumps. The minimizations were performed by 1000 steps of steepest descent followed by conjugate gradient until the rms gradient of the potential energy was lower than 0.05 kcal/(mol Å). A twin cut-off (12.0, 15.0 Å) was used to calculate nonbonded electrostatic interactions and the nonbonded pair-list was updated every 25 steps.

Step 2: docking of a known antagonist into the preliminary TM model

In a second step, a known antagonist (*N*-methylspiperone **1** for DRD2, antagonist **2** for AA3R, and SQ29,462 **3** for TA2R, Fig. 3) was docked into this preliminary model using the Gold 3.1 program.[52] For each of the 15 independent genetic algorithm (GA) runs, a maximum number of 1000 GA operations were performed on a single population of 50 individuals. Operator weights for crossover, mutation, and migration were set to 100, 100, and 0, respectively. The active site centre determined by the PASS program[53] was taken as the starting position of the GOLD flood fill algorithm. To allow for poor nonbonded contacts at the start of each GA run, the maximum distance between hydrogen donors and fitting points were set to 5 Å, and nonbonded van der Waals energies were cut off at a value equal to k_{ij} (well depth of the van der Waals energy for atom pair i,j). The Goldscore scoring function[54] was used for ligand docking to the DRD2 and TA2R receptor. The Chemscore scoring function as implemented in the Gold program[52] was used for ligand docking into AA3R, as preliminary docking studies showed that Gold docking simulations guided by this

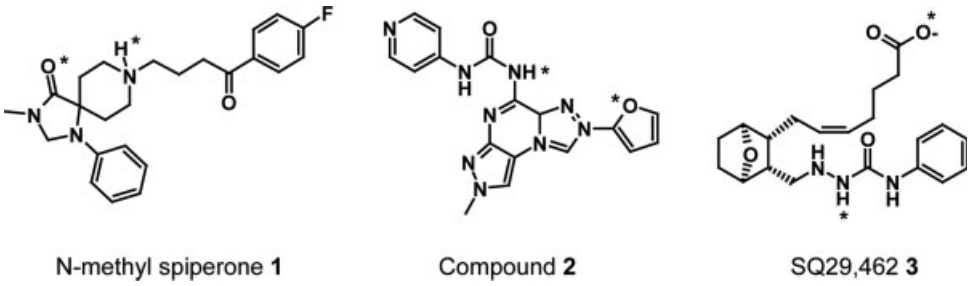


Figure 3

*Structure of reference antagonists (N-methylspiperone **1**, compound **2**, SQ29,452 **3** used to refine current DRD2, AA3R, and TA2R models, respectively. Atoms implicated in H-bond restrains to guide automated docking in preliminary TM receptor models (see "Computational Methods" section) are indicated with an asterisk.*

scoring function resulted in higher docking and virtual screening accuracies than Gold simulations guided by the Goldscore scoring function (results not shown). To further speed up the calculation, the GA docking was stopped when the top three solutions were within 1.5 Å rmsd. Experimentally-driven H-bond constraints[17,30,55,56] were used to preliminary guide the docking process in each receptor (see Fig. 3):

- DRD2: (1) between the protonable tertiary amine of *N*-methylspiperone and both D3.32 carboxylate oxygen atoms; (2) between the carbonyl oxygen of *N*-methylspiperone and the hydroxyl moiety of T7.39.
- AA3R: between the exocyclic amino groups of the [1,2,4]triazolo[1,5-c]pyrimidine ring of antagonist **2** (see Fig. 3), and the N6.55 sidechain carboxamide.
- TA2R: (1) between a carboxylate oxygen atom of SQ29,462 and the NE2 guanidine atom of R7.40 and (2) between an amide nitrogen atom of SQ29,462 (see Fig. 3), and the hydroxyl group of T7.39.

Step 3: energy minimization refinement of the preliminary TM-antagonist complex

In a third step, the antagonist–receptor complex was minimized with AMBER 8[51] using the AMBER03 force field to relax the structure and remove steric bumps as described earlier.

Step 4: construction and energy minimization refinement of the extracellular and intracellular loops

In a fourth step, a 5 residue window around C45.50 was added to the TM model by threading this part of ecl2 onto the bRho crystal structure and changing the residues in the respective residues of the AA3R and TA2R receptors. The Q167$^{45.51}$ residue in AA3R was manually rotated to form an intermolecular H-bond network with N6.55. The W182$^{45.49}$ residue in the ecl2 of TA2R was manually rotated into the large cavity between TM1 and TM2. The rest of ecl2 was constructed using two subsequent Modeller 8v1[57] runs with explicit disulfide bridge constraints. In the first run, the bRho crystal structure (PDB code 1U19.pdb)[8] was used to model the part upstream of ecl2. Of the 30 generated models, the model with highest Modeller and DOPE scores and ecl2 loop conformations properly accommodating the original antagonist binding orientation in the original TM model were selected as input for a second Modeller run. In this second run, the ecl2 segment downstream from the β_4 sheet was constructed. One out of 30 models was again selected based on the criteria described earlier.

After optimization of ecl2 conformation, ecls 1 and 3, icls 1 and 2, as well as helix 8 were modeled based on the bRho crystal structure (PDB code 1U19.pdb)[8] using Modeller 8v1.[57] The N-terminus and C-terminus were not included in any of the three models. The third intracellular was only included in the TA2R model. Extracellular and icls were aligned against bRho with T-coffee using standard settings.[36] Extracellular loop 3 (ecl3) of TA2R is relatively large (22 residues), has low sequence similarity with the ecl3 of bRho (13 residues), and initial simultaneous modeling of all the loops yielded models with ecl3 either flipped into the binding pocket or back flipped onto TM6 and TM7. Therefore, the ecl3 loops of the model (out of 30 models) with the best Modeller score were pasted on the initial 7TM model and this structure was then used as template structure for a next Modeller run to model ecl3. The final receptor model was energy minimized with the initially minimized antagonist docking pose as described earlier.

Step 5: molecular dynamics refinement of the TM-antagonist complex

In the final step, the full antagonist–receptor complex was embedded in a pre-equilibrated lipid bilayer consisting of 77 (TA2R) or 66 (AA3R) molecules of 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) and solvated with 10,811 (TA2R) or 8190 (AA3R) TIP3P water mole-

cules (box dimensions: 88.1 Å × 85.1 Å × 79.7 Å (TA2R) or 87.7 Å × 75.5 Å × 74.2 Å (AA3R)) as described by Urizar *et al.*[58] A short minimization was applied to the complex embedded in the hydrated lipid bilayer using AMBER 8 and applying a positional harmonic constraint of 50 kcal/(mol Å) on Cα carbon atoms. The entire system was then subjected to a 500 ps constant pressure molecular dynamics (MD) simulation. All bonds involving hydrogen atoms were frozen with the SHAKE algorithm. During the first 250 ps, the Cα carbon atoms were constrained as previously described and the temperature was linearly increased from 0 to 300 K. During the last 250 ps, the temperature was kept constant at 300 K and 1 bar, using a coupling constant of 0.2 and the Berendsen approach. Interactions were calculated according to the AMBER 03 force field, using particle-mesh-ewald (MPE) summation to include the long range electrostatic forces. Van der Waals interactions were calculated using a cut-off of 8.0 Å. The receptor-antagonist H-bond constraints earlier used for docking between were transformed into 3.5 Å upper-bound distance restraints during the MD-simulations. Additional distance restraints, supported by site-directed mutagenesis studies,[17,20,30] were defined between:

- AA3R: (1) the oxygen atom of the furan ring of antagonist **2** and the Q167$^{45.2.51}$ amide nitrogen atom, the Q167$^{45.51}$ amide nitrogen atom and the N6.55 amide oxygen atom, (2) the Q167$^{45.51}$ amide oxygen atom and the N6.55 amide nitrogen atom.
- TA2R: (1) the hydroxyl oxygen atom of S191$^{45.58}$ and the OD2 carboxylate oxygen of D5.36, (2) the NE guanidinium nitrogen of R173$^{45.40}$ and the OD2 carboxylate oxygen of D5.36, (3) the NH2 guanidinium nitrogen of R173$^{45.40}$ and the hydroxyl oxygen atom of S191$^{45.58}$.

Antagonist force-field parameters were derived using the Antechamber program[51] and partial charges for the substrates were derived using the AM1-BCC procedure in Antechamber.

Automated high-throughput ecl2 modeling

Preliminary TM receptor models of DRD2, AA3R, and TA2R, lacking icls and ecls, were generated by the earlier described GPCRmod program[22] by including the new optimized ground-state TM receptor models of AA3R and TA2R to the GPCR templates. The binding orientations of the antagonists in the F models were translated to these TM models and a short energy minimization was performed to relax the structure and remove steric bumps. The antagonist-free protein structure was considered as the TM model of the receptor. Two different modeling approaches were applied for automatically constructing partial ecl2 loops for the DRD2, AA3R, and TA2R. The ecl2 loop of model L1 was built using comparative distance restraint modeling with disulfide bridge constraints, using the Modeller8v1 program. The ecl2 loop of model L2 was constructed by threading onto the bRho crystal structure ecl2 backbone and modifying the bRho side chains into the corresponding ecl2 residues of the receptor. In accordance with the ecl2 modeling flow scheme described above and depicted in Figure 2, five-residue windows around C45.50 were automatically modeled for DRD2 and AA3R, while a 15-residue ecl2 segment up to the end of β_4 was modelled for TA2R (Fig. 1).

Ligand database preparation

For the evaluation of the virtual screening performance against the four different models (F, TM, L1, and L2) of the three different GPCR targets (DRD2, AA3R, TA2R), a database was prepared consisting of 990 drug-like compounds randomly selected from our in-house collection of commercially available compounds and 10 known receptor-specific antagonists shown in Tables III–V. The 10 known antagonists of each receptor were manually selected among existing chemotypes for each activity class[6,20,59–65] for their specificity and high affinity, and chosen to span the broadest chemical diversity for the different receptors. To avoid biasing virtual screening results, caution was given to select 990 drug-like decoys covering similar property ranges as the true actives (see property ranges in Supplementary Table II) but structurally different from any known active (the highest similarity coefficient, expressed by the Tanimoto coefficient on SciTegic ECFP_4 circular fingerprint,[66] of any decoy to any true active is 0.42).

DRD2, AA3R, and TA2R antagonists were manually sketched in Isis Draw.[67] Starting from the Isis Draw sketch, 2D sd structures were subsequently protonated using Filter2[68] and converted into 3D mol2 files with Corina 3.1.[69]

Automated docking based virtual screening

The ligand database was automatically docked into each refined receptor models using Gold-Goldscore (in the DRD2 and TA2R receptors) and Gold-Chemscore (AA3R receptor) as described above in step 2 of the "Modeling Antagonist-Bound GPCR Models" section.

Interaction fingerprint scoring

The binding poses of the three receptor-specific reference antagonists (Fig. 3) in each of the four different minimized receptor-complexes (models F, TM, L1, and L2 as defined above) of each of the three receptors (DRD2, AA3R, TA2R), were used to generate reference interaction fingerprints (IFPs) as previously described.[70] Seven different interaction types (negatively charged, positively charged, H-bond acceptor, H-bond donor, aro-

matic face-to-edge, aromatic-face-to-face, and hydrophobic interactions) were used to define the IFP. The cavity used for the IFP analysis consisted of the 30 residues earlier proposed to define a consensus TM binding pocket[5] plus two additional residues at positions 3.37 and 7.40 (which were added as they were accessible in the ligand binding pockets of the DRD2, AA3R, and TA2R receptor models and shown to be involved in antagonist binding in AA3R[17] and TA2R[20]):

DRD2

Y1.35, L1.39, L1.42, I1.46, V2.57, M2.58, V2.61, E2.65, F3.28, V3.29, D3.32, V3.33, C3.36, T3.37, I3.40, I4.56, L4.60, F5.38, V5.39, S5.42, S5.43, S5.46, F6.44, W6.48, F6.51, F6.52, H6.55, Y7.35, T7.39, W7.40, Y7.43, N7.45.

AA3R

Y1.35, E1.39, I1.42, A1.46, V2.57, M2.58, A2.61, S2.65, M3.28, T3.29, L3.32, L3.33, T3.36, H3.37, I3.40, V4.56, P4.60, M5.38, V5.39, S5.42, F5.43, W5.46, F6.44, W6.48, L6.51, S6.52, N6.55, L7.35, I7.39, L7.40, S7.43, N7.45.

TA2R

S1.35, A1.39, F1.42, G1.46, T2.57, G2.58, V2.61, H2.65, M3.28, G3.29, M3.32, I3.33, G3.36, L3.37, L4.56, P4.60, F5.39, G5.40, F5.43, S5.44, G5.47, A6.44, W6.48, L6.51, L6.52, I6.55, L7.35, L7.39, R7.40, T7.43, N7.45

Notice that the binding pocket residues in the TM5 of TA2R have different numbers as the TM5 helix was unkinked in the new homology model of this receptor. Standard IFP scoring parameters, and a Tanimoto coefficient (Tc-IFP) measuring IFP similarity with the reference antagonist pose,[70] were used to rank the docking poses of 10 known antagonists (shown in Tables III–V) and 990 drug-like molecules generated in virtual-screening studies against 12 different receptor models (four models F, TM, L1, and L2 for each receptor).

Virtual screening analysis

Virtual screening accuracies are determined in terms of yield (Y), and the area under the curve (AUC) of receiver-operator characteristic (ROC) plots.[71] The yield is defined as the percentage of true positive hits retrieved by our virtual screening protocol at the top 5% of the hit list (50 highest ranked hits of the 1000 molecules):

$$\text{Yield} = (t/T) \times 100$$

where t the number of true hits found in the hit list, and T the total number of true hits in the full database (10). The yield parameter was used as an indicator of the sensitivity of the screening (how many true positives are recovered), while the AUROC value (area under the ROC curve) was used as an indicator of both the sensitivity and the specificity (how many false positives are recovered) of the VS docking-scoring strategy.

RESULTS

Sequence alignment and analysis of 365 human GPCR ecl2s

The TM sequences of 365 human GPCRs were aligned in earlier studies using the GPCRmod program.[5,22] On the basis of this TM alignment, ecl2 sequences were extracted for each receptor and separate ecl2 alignments were constructed for each GPCR cluster as defined by Surgand *et al.*[5] The results of the ecl2 sequence alignment and analysis per GPCR cluster are summarized in Table I and exhaustively shown in Supplementary Table III.

With the exception of 2 adhesion receptors (GP128, GPR97), and one glutamate receptor (GP158), all GPCRs under investigation contained at least one conserved cysteine residue less than three positions away from the N-terminal or C-terminal ends of ecl2. However, 38 GPCRs (10% of the database) did not possess the conserved cysteine at position 3.25. Among these GPCRs lacking the conserved TM3-ecl2 disulfide link are all receptors belonging to the MAS-related (11) and melanocortins (5) clusters, all receptors belonging to the lipids cluster with the exception of GP119 (13), some receptors belonging to the glutamate cluster (4), and 5 receptors considered as singletons in our previous GPCR classification study.[5] Several GPCRs contain two (36), three (ADRB1-3, AA2AR, P2Y11), or even four (AA2BR) cysteines in ecl2. For some of these receptors an additional disulfide link either with the N-terminal domain[38,39] or within ecl2[40] have been postulated. In 31 of 42 cases, experimental data alone[37–40] or in combination with the GPCR CLUSTER alignment (see "Computational Methods") to other closely related receptors[9,37–48] could be used to assign the disulfide bridge-forming cysteine residue in ecl2 (C45.50). The furthest downstream located ecl2 cysteine residue was indicated to be the one connected to C3.25 in most of these 31 cases. For the 11 remaining receptors, C45.50 was assigned by either the CLUSTER_OPSINS_PROFILE alignment (i.e. for the 3 acid receptors), or CLUSTER alignment alone (see Supplementary Table III).

Table I shows that for some GPCR clusters, the physicochemical properties of at least two of three residues in the close vicinity of C45.50 (one position downstream (45.49), one (45.51) and two positions (45.52) upstream from C45.50 are conserved (among more than 80% of the cluster entries): frizzled, prostanoids, glycoproteins, and adhesion receptors. For other receptors at least one position is conserved for more than 80% of the entries: opsins, melanotonins, and opiates receptors. For many

Table I
Upstream and Downstream ecl2 Loop Lengths for Different Human GPCR Clusters

GPCR cluster (number of receptors)	Residue physicochemical property			# Residues (stdev)	
	Upstream	Downstream		Upstream	Downstream
	45.49	45.51	45.52		
Acids (7)			S	11 (2)	9 (2)
Adenosine (6)	L		F	21 (5)	8 (2)
Adhesion (31)		W	L/I/M/V	14 (4)	4 (4)
Amines (42)			I/L/V	15 (5)	5 (1)
Brain-gut peptides (10)				23 (14)	9 (1)
Chemoattractants (16)				13 (1)	12 (6)
Chemokines (23)				12 (2)	11 (2)
Frizzled (11)	V/I/L	F/Y	V	12 (0)	9 (0)
Glutamate (18)				16 (5)	3 (1)
Glycoproteins (8)	I/L/V		P	12 (3)	9 (2)
Melatonin (7)		V/I		14 (3)	7 (2)
Opiates (13)			L/V/M/I	12 (2)	11 (2)
Opsins (10)	S/T/C			13 (0)	11 (1)
Peptides (26)				16 (4)	10 (1)
Prostanoids (8)	W	F	I/L	13 (1)	10 (4)
Purines (35)			D/E	14 (3)	11 (1)
Secretin (15)		W		9 (0)	4 (1)
SREB (6)				12 (1)	8 (3)
Vasopeptides (7)		W	A	15 (2)	10 (1)
bRho	S	G	I	13	12
all GPCRs (325)[a]				14 (5)	9 (9)

Conserved residues one position upstream (45.49) and one and two positions downstream (45.51, 45.52) from C45.50 are coloured according to the degree of similarity among the receptor cluster entries (white foreground/black background, 100%; white foreground/grey background, >80%; black foreground/grey background, >60%; following the similarity assignment in Surgand et al.[4]). Columns of the amino acid positions for which the physicochemical property is not conserved (<60% of the entries for each cluster) are left blank. The standard deviations of the average upstream and downstream ecl2 loop lengths are depicted in parentheses.
[a]Including the results for 25 singletons (not included in the chemogenomic classification of Surgand et al.)[4] and GP119 (belonging to the lipids cluster).

other GPCR clusters the 45.49, 45.51, and 45.52 residues show some similarity (at least on positions conserved for 60–80% of the cluster entries: vasopeptides, adenosine, acids, biogenic amines, purines) or no significant similarity (SREB, glutamate, peptides, chemoattractants, chemokines, brain-gut peptides). The supplementary Table III provides the sequence analysis results for each receptor individually.

The current GPCR sequence analysis study, in agreement with an earlier study[12] show that the total ecl2 loop length is rather variable. Our current analysis of 325 nonolfactory nonredundant human GPCRs containing both C3.25 and C45.50, however, shows that this variability mainly occurs in the part of ecl2 downstream from C45.50 (Table I and Fig. 4). The average number of upstream residues is 12, and most of the GPCRs possess 11–15 residues upstream from C45.50 (66%), which is close to the number of upstream residues in bRho (13). In comparison, the same percentage of GPCRs possess 6–14 residues downstream (9 residues on average and 12 in bRho) and 20–33 residues in the total ecl2 loop (24 on average and 26 in bRho). From Table I, it can be derived that secretin receptors have a significantly shorter upstream ecl2 loop, while adenosine and brain-gut peptide receptors have significantly longer upstream ecl2 loops compared to bRho. Secretin, adhesion, glutamate, SREB, adenosine, amine, and melatonin receptors have significantly shorter downstream ecl2 loops compared to bRho. Only few GPCRs have a significantly longer downstream ecl2 loop than bRho, among which C3AR with the longest downstream and total ecl2 loop length of 160 and 173 residues, respectively.

The three receptor targets selected as ecl2 modeling and virtual screening test cases in the current study, DRD2, AA3R, and TA2R, span a relatively wide range of upstream and downstream ecl2 loop lengths. Compared to the bRho crystal structure template, DRD2 has a relatively short upstream (7), as well as downstream (3) ecl2 loop. AA3R has a relatively long upstream (18) and medium sized downstream (7) ecl2 loop, while TA2R has the same upstream loop length (13) and a somewhat smaller downstream ecl2 loop (9), compared to bRho (13 residues upstream, 12 residues downstream).

Classification of GPCRs in terms of ecl2 modeling feasibility

On the basis of the ecl2 sequence alignment and analysis, the 365 human GPCRs in our database were classified according to their modeling feasibility as depicted in

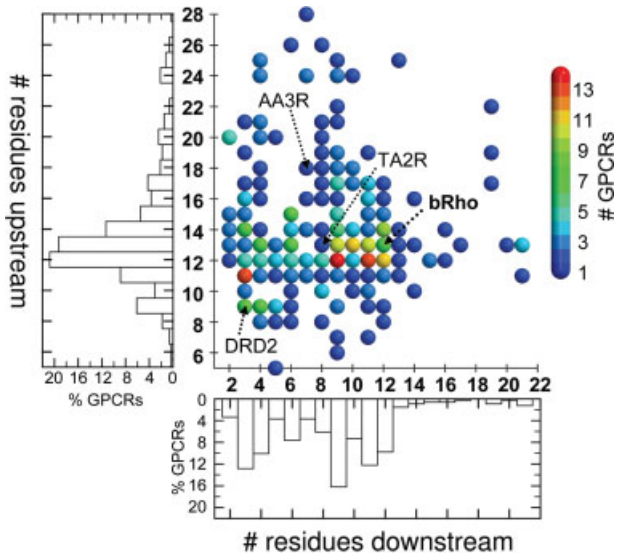


Figure 4

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. The upstream ecl2 loop length (showing an optimum at 12–13 residues) is less variable then the downstream ecl2 loop length (showing optima between 3 and 12 residues). Compared to bRho, many GPCRs have approximately the same number of residues upstream and downstream (e.g. TA2R), but other GPCRs have a somewhat shorter (e.g., DRD2) or longer (e.g. AA3R) upstream ecl2 loop, and/or significantly shorter (DRD2) or longer (TA2R) downstream ecl2 loop (see also Table II). Note that CML1 (27 upstream, 12 downstream), C3AR (160, 12), GP150 (35, 11), DRD5 (39, 4), and MTLR (59, 8) are not depicted in the plot. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Figure 2. The results of this ecl2 modeling classification are summarized in Table II. The supplementary Table III provides the classification results for each receptors individually. Five ecl2 modeling classes are considered:

1. *No C3.25–C45.50 disulfide link*: receptors lacking the TM3-ecl2 disulfide link for which no obvious ecl2 model can be generated;
2. *5-residue window*: receptors with insertions/deletions in the ecl2 β-sheet region, for which a 5-residue window cantered on C45.50 can be generated;
3. *β_3-β_4 aligned*: receptors without any insertion/deletion in the ecl2 β-sheet region, but with insertions/deletions before strand β_3, for which β_3 (4 residues) and β_4 (4 residues) sheets can be generated;
4. *Up to end β_4 aligned:* receptors without any insertion/deletion before or in the ecl2 β-sheet region, but with insertions/deletions after strand β_4, for which an ecl2 segment stretching from the beginning of ecl2 until the end of β_4 (16 residues) can be generated;
5. *Fully aligned*: receptors without any insertion/deletion in the whole ecl2 loop, for which a full ecl2 loop (26 residues) can be generated.

Mild and strict classification criteria were defined for defining "β_3-β_4 aligned" and "up to end β_4 aligned" ecl2 modeling classes, based on upperstream ecl2 loop length, and GPCR cluster profile alignments with the opsins cluster, respectively (see "Computational Methods" section). According to the mild as well as the strict classifi-

Table II

ecl2 Modeling Feasibility for 365 Human GPCRs

Modeling class[a]	%[b]	GPCR clusters[c]	# res upstream[d]
no C3.25-C45.50 link	11	Lipids,[e] melanocortins,[f] MAS-related,[f] glutamate,[e,f] adhesion (GP128, GPR97)[e]	—
5 residue window	80 (*40*)	Brain-gut peptides, acids, purines, chemokines, chemo-attractants, opiates, lipids (GPR19), peptides, melatonin, amines, adenosine, vasopeptides, SRB (GPR27), glutamate, adhesion, secretin, frizzled	3–11, >14
β_3-β_4 aligned	2 (*29*)	acids, vasopeptides (Q6W5P4, OXYR), brain-gut peptides (MCHR1, MCHR2) *(chemokines, opiates,[g] chemoattractants,[g] purines,[g] peptides,[g] amines,[g] SRB,[g] glycoproteins,[g] glutamate,[g] adhesion,[g] frizzled[g])*	12,14
Up to end β_4 aligned	5 (*18*)	Opsins, prostanoids, SRB, singletons (GPR87) *(adhesion,[h] peptides,[h] melatonins (GPR63, GPR45),[h] amines,[h] chemokines,[h] opiates,[h] chemoattractants,[h] purines[h])*	13
Fully aligned	2	Opsins	13

[a]As described in the Computational Methods section and Figure 2.
[b]Percentage in parentheses in italics is based on upstream ecl2 length only, percentage in parentheses is after application of T-coffee profile alignment (as described in the Computational Methods section).
[c]In cases were clusters are represented by less than 3 entries, the GPCRs are mentioned explicitly in parentheses.
[d]Number of residues upstream from C3.25.
[e]Lacking cysteine residue at least two residues upstream from the N-terminal end or two residues downstream from the C-terminal end of ecl2.
[f]C3.25 lacking.
[g]Clusters in parentheses in italics classified into the "β_3-β_4 aligned" group based on upstream ecl2 length compatibility with bRho, but are classified into the "5 residue segment" group when applying T-coffee profile alignment against the opsin cluster (as described in the Computational Methods section).
[h]Clusters in parentheses in italics classified into the "up to end β_4 aligned" aligned group based on upstream ecl2 length only, but are classified into the "5 residue segment" group when applying T-coffee profile alignment against the opsin cluster (as described in the Computational Methods section).

cation criteria, most of the receptors were classified in the "5-residue window" group (40 and 80%, respectively). According to the mild classification criteria, many receptors however could be classified in the "β_3-β_4 aligned" and "up to β_4 aligned" groups (in total 29% and 18% of the receptors, respectively), including receptors belonging to the frizzled (11 out 11), prostanoids (8/8), SRB (5/6), chemoattractants (16/18), purines (24/35), chemokines (15/23), melatonins (5/8), opiates (8/13), and opsins (4/10) GPCR clusters, some of the peptides (8/26), biogenic amines (14/42), and vaso-peptides (2/7) clusters, together with some singletons (11/30). According to the strict classification criteria, however, only 2% of the receptors were classified in the "β_3-β_4 aligned" group (some acids (3), some vasopeptides (2), and some brain-gut peptides (2)), and only 5% of the receptors (prostanoids (11), some opsin receptors (4), and most SRB receptors (4)) were classified in the "up to β_4 aligned" group. The ecl2s of most of the opsins receptors could be aligned to the ecl2 of bRho without any insertions, nor deletions, classifying them in the "fully aligned" group (accounting for 2% of GPCRs).

Construction and validation of antagonist-bound models of DRD2, AA3R, and TA2R receptors

To get 3D coordinates of the antagonist-bound state of the DRD2, AA3R, and TA2R receptors, we chose *N*-methylspiperone (**1**, DRD2 receptor), antagonist **2** (AA3R receptor), and SQ29,452 (**23**, TA2R receptor), docked them according to known experimental data into their respective receptors, energy minimized the complexes, and used the receptor–ligand interactions observed in the minimized complexes as interaction fingerprint (IFP) reference for scoring ligands in our virtual screening exercise. These three reference ligands (Fig. 3) were selected in the light of their conformational freedom (as rigid as possible), high binding affinity[6,20,72], and other known experimental (site-directed mutagenesis) data to guide us in the initial docking.[6,17,20,73–75] Moreover, mutating residues one to two positions downstream from C45.50 in DRD2 (I183$^{45.51}$C, I184$^{45.52}$C)[6] and TA2R (F184$^{45.52}$A/Y)[20], was shown to affect *N*-methylspiperone and SQ29,452 binding, respectively. MRS1220 (Compound **14**, Table IV), a selective AA3R antagonist structurally related to the AA3R reference antagonist **2**, was shown to be affected by a point mutation 14 residues upstream from C45.50 in AA3R (K152$^{45.36}$A)[17], while CGS15943, a nonselective adenosine receptor antagonist, was shown to be affected by point mutations one residue downstream from C45.50 (Q167$^{45.51}$A/E/R).[30] Later follows a detailed description and validation of our DRD2, AA3R, and TA2R receptor homology models and receptor–antagonist binding modes.

DRD2

The proposed binding mode of *N*-methylspiperone to the full DRD2 model (Fig. 5, panel F) is consistent with known experimental data. Like many aminergic GPCRs, the dopamine D2 receptor (DRD2), binds its ligands through the conserved aspartic acid residue D3.32. Site directed mutagenesis studies have further identified V2.61, F3.28, V3.33, C3.36, F5.47, W6.48, F6.51, F6.52, Y7.35, Y7.43, I183$^{45.51}$, and I184$^{45.52}$ as antagonist binding partners.[6,73,76] Three serine residues in TM5 (S5.42, S5.43, and S5.46) are involved in binding of agonist and some antagonists,[59] and mutation of H6.55 only affects binding of specific antagonists.[77] Furthermore, mutation of T7.39 was shown to significantly affect agonist and antagonist binding to the closely related dopamine D3 receptor.[55] Figure 5 shows the typical binding mode of dopamine D2 receptor antagonists observed in the current study, in line with the reported site-directed mutagenesis studies, and earlier reported homology modeling studies.[22,78,79] The protonatable amine of the antagonist forms a salt bridge with the negatively charged car-

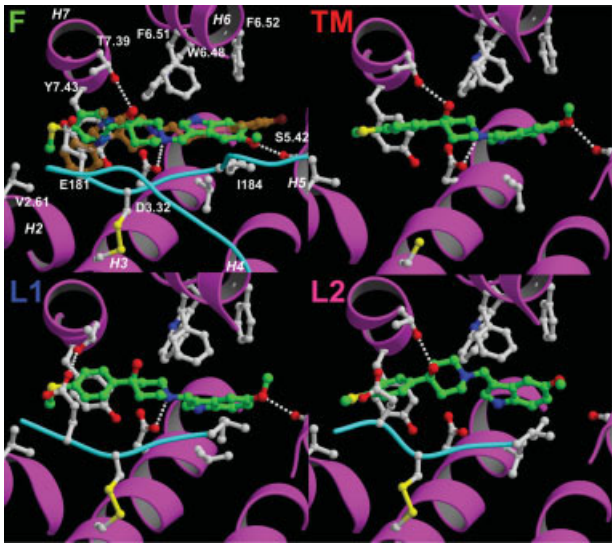


Figure 5

The binding mode of antagonist ***6***^{61} *(green carbon atoms, see Table III for 2D-representation) in different models of the dopamine D2 receptor (DRD2): the F model with all icls and ecls (except the N-terminal and C-terminal loops), the TM model without any loops, the L1 model with only a 5 residue segment of ecl2 modeled with Modeller, the L2 model with only a five residue segment of ecl2 modeled by direct threading onto the bRho crystal structure backbone (see the "Computation Methods" section). The binding mode of the template antagonist N-methylspiperone (**1**, see Fig. 3 for 2D-representation) is depicted in model F (transparent orange carbon atoms), and is similar to the binding modes in the other three models. Important ligand binding residues in the DRD2 pocket (magenta backbone ribbon) are depicted as balls and sticks and the ecl2 backbone is depicted in cyan. Oxygen, nitrogen, sulphur, and fluor atoms are coloured red, blue, yellow, and brown, respectively. The figures [as well as Figs. 6,7, and 10(A–C)] were made using Molscript*[10] *and Raster3D.*[11] *[Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

boxylate group of D3.32, and additional H-bonds are formed between antagonist hydroxyl and/or amide groups and the T7.39 (and Y7.35) hydroxyl group(s). Dopamine D2 antagonists occupy both subpockets i (TMs 1, 2, 3, and 7) and ii (TMs 3, 4, 5, 6),[5] making aromatic interactions (F3.28, F6.51, F6.52, and Y7.43) and hydrophobic contacts (V2.61, V3.33, C3.36, I4.56, W6.48, I183$^{45.51}$, and I184$^{45.52}$) with the receptor. Especially, interactions in subsite i have been shown to be responsible for D2/D4 selectivity.[80] The above described binding mode of the template antagonist *N*-methylspiperone was used for refining our DRD2 receptor homology model and rank database ligands with our IFP-scoring function as described in the "Computational Methods" section.

AA3R

Like many adenosine receptors, the adenosine A3 receptor (AA3R), accommodate its ligands through a conserved asparagine residue N6.55.[17] Mutational studies have further identified H3.37 and Q167$^{45.51}$ as binding partners for both agonists and antagonists.[17,30] Mutation of K152$^{45.36}$ and W6.48 was only shown to significantly affect antagonist binding, while having only slightly effects on agonist binding.[17] E1.39 and H7.43 were found to be important for constraining the inactive ground structure of AA3R by forming an intramolecular H-bond network between TM1 and TM7.[81] Furthermore, Y7.53 was shown to be involved in agonist binding,[82] while mutation of S6.52 showed only very minor effects on agonist and antagonist binding.[17] These site-directed mutagenesis studies and recent molecular modeling studies[75] suggest that distinct binding domains exist in AA3R for agonists and (non-nucleoside) antagonists. Both agonists and antagonists are proposed to form H-bond interactions with N6.55 and Q167$^{45.51}$ and bind in pocket ii between TM3 (L3.32, L3.33, T3.36, H3.37, I3.40), TM5 (F5.43, I5.47), TM6 (F6.44, W6.48, L6.51), as is the case for the non-nucleoside antagonists investigated in the current study (Fig. 6). Agonists form additional H-bond interactions with TM7 (S7.42 and H7.43). Additional interactions with non-nucleoside antagonist occur at the upper regions of TM5 (V5.39, F5.43), TM6 (Y6.59), and ecl2 (F168$^{45.52}$), as is shown in Figure 6 for MRE 3008-F20 (compound **15**, Table IV) and template antagonist **12** (Fig. 3). The above described binding mode of compound **12**, in agreement with the reported site-directed mutagenesis studies and earlier proposed binding modes presented in molecular modeling studies,[75] was used for refining our AA3R receptor homology model and rank database ligands with our IFP-scoring function as described in the "Computational Methods" section.

TA2R

Like other prostanoid receptors, the thromboxane A2 receptor (TA2R), recognized its ligands via an ionic bond between a conserved arginine residue R7.40 and the ligand carboxylate group.[56] The involvement of other residues in ligand binding was found to be rather ligand-specific. S5.43 and S6.35 seem to be involved in agonist binding, but not in antagonist binding. Mutation of F184$^{45.51}$, T186 $^{45.53}$, and S191$^{45.58}$ in ecl2 did not affect agonist binding, but did affect antagonist binding.[20] Mutation of C3.22 and D5.36 affected both agonist and antagonist binding. The binding mode of the SQ29,452 antagonist (compound **3**, Fig. 3) we used for refining our TA2R receptor homology model and IFP-scoring was in line with the ligand binding mode proposed for PI2R, based on extensive site-directed mutagenesis studies for this closely related prostanoid receptor.[74] The ligand binds in pocket i, between TM2 (G2.57), TM3 (M3.28, M3.32, I3.33), TM6 (W6.48, L6.51), and TM7 (L7.39, R7.40, T7.43, N7.46), and ecl2 (W182$^{45.49}$, F184$^{45.51}$), as illustrated in Figure 7 for antagonist S-145 (compound **28**, Table V) and template antagonist SQ29,452. Subsite i of prostanoid receptors is delimited by conserved small or medium-sized residues at positions 1.46 (G), 2.58 (S/T/G), 2.61 (V/T/A), and 7.43 (T/S/A), and thereby extends relatively deep between TMs 1 and 2.[5] This leaves room for the bulky W$^{45.49}$ residue, conserved throughout all prostanoid receptors (Fig. 8). The

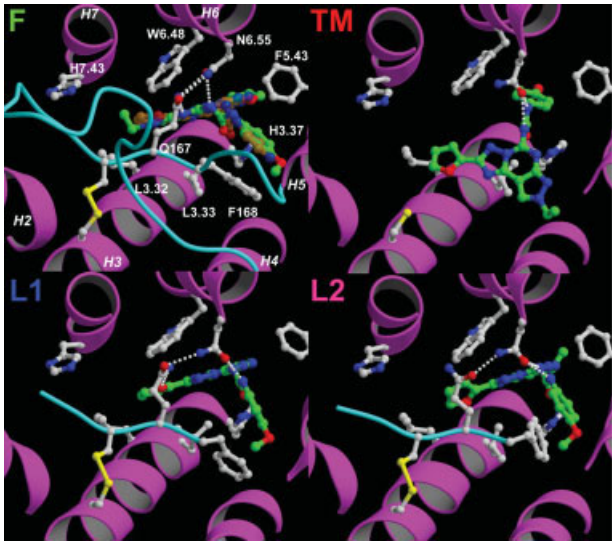


Figure 6

*The binding mode of MRE 3008-F20 (antagonist **15**, green carbon atoms, see Table IV for 2D-representation) in different models of the adenosine A3 receptor (AA3R): the F model with all icls and ecls (except the N-terminal and C-terminal loops), the TM model without any loops, the L1 model with only a 5 residue segment of ecl2 modelled with Modeller, the L2 model with only a 5 residue segment of ecl2 modeled by direct threading onto the bRho crystal structure backbone (see the "Computational Methods" section). The binding mode of the template antagonist **12** (see Fig. 3 for 2D-representation) is depicted in model F (transparent orange carbon atoms), and is similar to the binding modes in the other three models. Rendering and colour coding is the same as in Figure 5. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

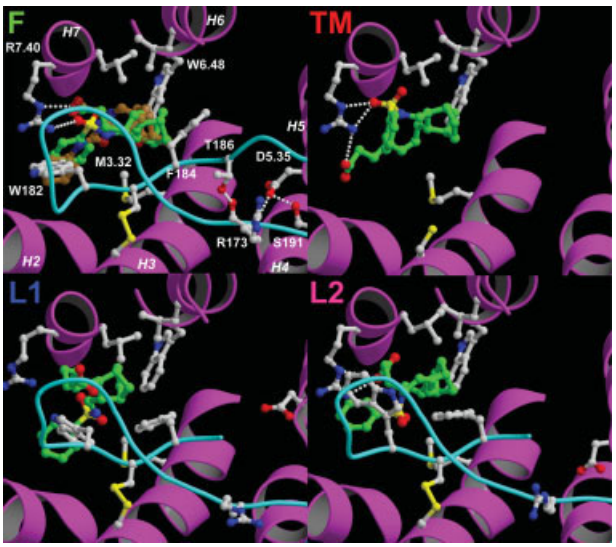


Figure 7

*The binding mode of S-145 (antagonist **28**, green carbon atoms, see Table V for 2D-representation) in different models of the thromboxane A2 receptor (TA2R): the F model with all icls and ecls (except the N-terminal and C-terminal loops), the TM model without any loops, the L1 model with ecl2 up to two residues downstream of C45.50 modeled with Modeller, the L2 model with ecl2 up to two residues downstream of C45.50 modeled by direct threading onto the bRho crystal structure backbone (see the "Computational Methods" section). The binding mode of the template antagonist SQ29,452 (**23**, see Fig. 3 for 2D-representation) is depicted in model F (transparent orange carbon atoms), and is similar to the binding modes in the other three models. Rendering and colour coding is the same as in Figure 5. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

observed hairpin conformation of the receptor-bound antagonists is furthermore supported by earlier modeling studies, indicating this conformation to be very close in energy to more extended conformations.[90] The residues C3.22, D5.36, T186$^{45.53}$, and S191$^{45.58}$ are located far away from this postulated ligand binding pocket, but were nevertheless found to be involved in antagonist binding. C3.22 might be involved in disulfide bonding with C11 in the N-terminal loop, in the same way as a second nonconserved disulfide link (in addition to the conserved C3.25–C45.50 link) was proposed for PI2R, which is essential for the structural integrity of this receptor.[38] Our homology model suggests a similar structural role for D5.36, T186$^{45.53}$, and S191$^{45.58}$, which form an intramolecular H-bond network between ecl2 and TM5 (also involving R173$^{45.40}$), and, in this way, controls the conformation of ecl2 and its interactions with the ligand. In the same way, an intramolecular H-bond network can be observed in the bRho crystal structure between N5.36, R177$^{45.44}$, and D190$^{45.57}$.[9] TM5 of GPCRs in the prostanoid receptor cluster does not contain highly the conserved proline residue at position 5.50, which usually induces an opening in TM5 stabilized by a bulky aliphatic residue at position 3.40 in the bRho crystal structure. Instead, TA2R contains glycine residues at positions 5.47 and 5.52, offering space for the L3.40 residue. The TM5 of TA2R was unkinked, as described in the "Computational Methods" section, accommodating D3.36 to participate in the above described H-bond network. PI2R mutation studies suggesting F150 and Y188 (located at positions 4.56 and 5.43 in our homology models, respectively) to form an aromatic cluster essential for receptor activation,[49] offered further circumstantial evidence for an alternative prostanoid receptor-specific conformation of TM5.

Virtual screening of receptor antagonists

The virtual screening accuracies of the four different models (F, TM, L1, L2) of the three receptors (DRD2, AA3R, and TA2R) were subsequently evaluated in terms of yield and AUROC values (see "Methods") for a virtual screening experiment involving 10 known receptor-specific antagonists and 990 drug-like decoys. Gold-Goldscore (DRD2 and TA2R) and Gold-Chemscore (AA3R) generated docking poses were further ranked using our IFP-scoring program[70] and taking reference IFPs from *N*-methylspiperone (D2DR), compound **2** (AA3R), and SQ29,452 (TA2R). For all receptor models, IFP-scoring outperformed the native Goldscore and Chemscore scoring in terms of VS accuracy (results not shown). The results of the VS studies are shown in Tables III–V and Figure 9 and described in detail below for each receptor separately.

DRD2

The VS accuracy obtained against the full DRD2 receptor (F) was much higher than the VS accuracies of the two models containing an automatically modelled 5-residue window around C45.50 (L1 and L2), and also significantly better than the VS accuracy of the model containing only the seven TMs. Table III and Figure 9 show that in the full DRD2 model hit list, 60% of the known antagonists (6 of 10 molecules) are ranked in the top 5% (50 molecules) of the database, while only 50, 30, and 30% of the true hits are ranked in the top 5% of the hit lists models TM, L1, and L2, respectively. AUROC values also indicate that the F model of DRD2 (AUROC of 0.92) is a significantly more sensitive as well as selective virtual screening target structure than the TM (0.79), L1 (0.81), and L2 (0.79) models.

Figure 5 illustrates for compound **6** (Table V), the effect of DRD2 ecl2 modeling on antagonist binding mode prediction and VS accuracy. All models, excepted for model L1, show an ionic interaction between the positively-charged protonated nitrogen of antagonist **6** and the negatively charged carboxylate group of D3.32, and hydrophobic and aromatic ligand-receptor interactions in binding pocket i (delimited by TMs 1, 2, 3, and 7) and binding pocket ii (delimited by TMs 3, 4, 5, and 6). Additional hydrophobic interactions between antagonist

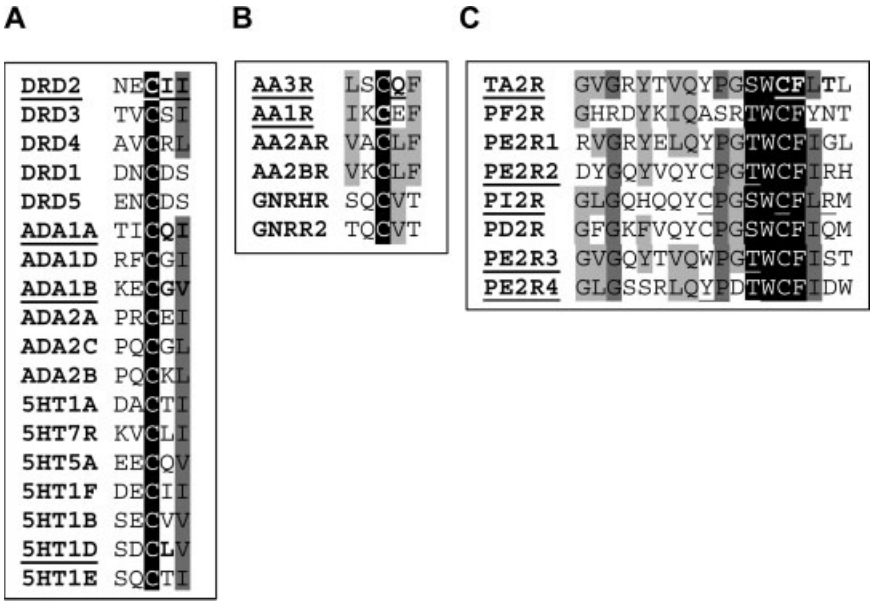


Figure 8

*Amino acid sequence alignments of parts of the second extracellular loop automatically modelled in the current study of a sub class of the biogenic amine receptor cluster (**A**), the adenosine receptor cluster (**B**), and the prostanoids receptor cluster (**C**), coloured according to the degree of similarity (white foreground/black background, 100%; white foreground/grey background, >80%; black foreground/grey background, >60%; following a previous similarity assignment in reference.[5] Residues shown to be involved in antagonist binding (bold)[20,83,84], agonist binding (underlined)[38,85–87], and antagonist binding as well as agonist binding (bold underlined)[6,20,30,88,89] are indicated in the sequences of underlined receptors. Note that the ecl2 residues indicated for ADA1A, ADA1B, PE2R3, and the R45.53 residue indicated for PI2R are shown to be involved in the receptor of nonhuman species (while the other indicated residues are shown to be involved in ligand binding in human GPCRs).*

6 and I184$^{45.52}$ (located in ecl2) are observed in all three loop models. Models F and TM show additional H-bond interactions between the ligand and T7.39 and S5.42 that are not observed in models L1 and L2, respectively. In the L1 model, T7.39 forms an intramolecular H-bond with ecl2 residue E181$^{45.49}$ instead. This loss of ligand–receptor H-bond interactions compared to the binding mode of the *N*-methylspiperone-DRD2 reference complex results in lower IFP ranking of the antagonist **6** in models L1 and L2 (Table III).

AA3R

The VS accuracy of the full AA3R receptor (F) was significantly better than the VS accuracies of the two models containing an automatically modeled 5-residue window around C45.50 (L1 and L2), and much better than the VS accuracy of the TM model. Table IV and Figure 9 show that in the full AA3R model hit list, 40% of the known antagonists are ranked in the top 5% (50 molecules) of the database, while only 30, 30, and 10% of the true hits are ranked in the top 5% of the hit lists models L1, L2, and TM, respectively. AUROC values also indicate that the F model of AA3R (AUROC of 0.81) is a significantly more sensitive as well as selective virtual screening target structure than the L2 (0.72), L1 (0.59), and TM (0.57) models.

Figure 6 illustrates for antagonist MRE 3008-F20 (compound **15**, Table IV) the effect of AA3R ecl2 modeling on antagonist binding mode prediction and VS accuracy. In all models H-bond interactions are observed between the urea moiety of compound **15** and the carbonyl oxygen atom of N6.55, and hydrophobic and aromatic ligand-receptor interactions in binding pocket ii (delimited by TMs 3, 4, 5, and 6). The F and L1 models show additional ligand–receptor H-bond interactions with Q167$^{45.51}$ located in ecl2. Models F, L1, and L2 show intramolecular H-bonds between N6.55 and Q167$^{45.51}$ and π-π stacking of the *O*-methyl substituted aromatic ring of MRE 3008-F20 between the phenyl rings of F5.43 and F168$^{45.52}$. The later ligand–receptor interaction is lacking in the loopless TM model, causing MRE 3008-F20 to bind close to transmembrane helix 4 (in an area occupied by ecl2 in the other models). This results in the loss of hydrophobic and aromatic ligand–receptor interactions observed for the AA3R- antagonist **2** reference complex and a lower IFP ranking of compound **15** (Table IV).

TA2R

The VS accuracy of the full TA2R receptor (F) was higher than the VS accuracies of the two models contain-

Table III
Virtual Screening of a Database of 990 Decoys and 10 Known DRD2 Antagonists

Ligand[a]	Nr	Model F[b]	TM[c]	L1[d]	L2[e]
haloperidol	4				
nemanopride	5				
	6				
mazapertine	7				
pimozide	8				
DM1451	9				
S-butaclamol	10				
domperidone	11				

Table III
(Continued)

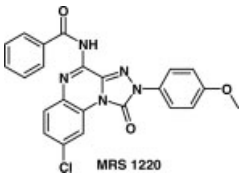
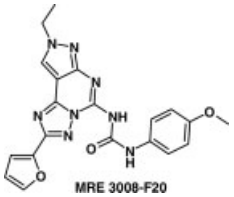
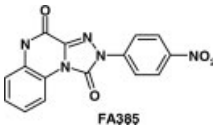
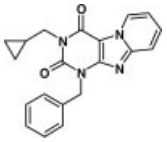
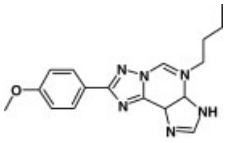
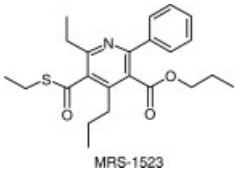
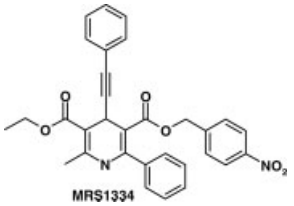
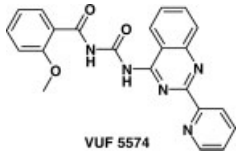
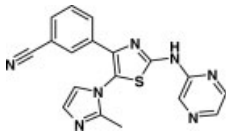
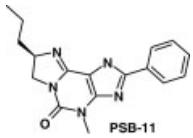
Ligand[a]	Nr	Model F[b]	TM[c]	L1[d]	L2[e]
5-OMe-BPAT	12				
S-sulpiride	13				
Yield[f]		60	50	30	30
AUROC value[g]		0.92	0.79	0.81	0.79

[a]Ligands have been selected from the literature.[6,52–55] Black, dark grey, and light grey shades indicate that the current antagonist has been ranked in top 5, 10, and 20% of the database, respectively.
[b]Full model with all intracellular and extracellular loops (except the N-terminal and C-terminal loops).
[c]Loop less TM.
[d]TM Model with only a 5 residue segment of ecl2 modeled with Modeller.
[e]TM model with only a 5 residue segment of ecl2 modeled by direct threading onto the bRho crystal structure backbone.
[f]Percentage of known D2DR antagonists ranked among the top 5% scorers (IFP scoring[63] using N-methylspiperone **1** (Fig. 3) as reference.
[g]Area under the curve of the ROC plot.

ing an automatically modelled ecl2 loop segment until the end of the β_4 sheet [L1 and L2, see Fig. 1(B)], but slightly lower than the VS accuracy of TM model. Table V and Figure 9 show that in the full AA3R model hit list, 50% of the known antagonists are ranked in the top 5% of the database, while 60, 20, and 20% of the true hits are ranked in the top 5% of the hit lists of models TM, L1, and L2, respectively. AUROC values also indicate that the F and TM models of TA2R (AUCROC of 0.87 and 0.88, respectively) are significantly more sensitive as well as selective virtual screening target structures than the L1 (0.74), L2 (0.71) models.

Figure 7 illustrates, for antagonist S-145 (compound **28**, Table V), the effect of TA2R ecl2 modeling on antagonist binding mode prediction and VS accuracy. In all models S-145 (**28**) binds in pocket i (delimited TMs 1, 2, 3, and 7). Models F, TM, and L2 show an ionic interaction between the negatively charged carboxylate group of S-145 and the positively charged guanidinium group of R7.40. Additional H-bond interactions between R7.40 and the sulfonamide oxygens of S-145 are observed in models F and TM. In model F an intramolecular H-bond network is formed between R173, T186, S191 (all located in ecl2), and D5.35. Aromatic and hydrophobic ligand–receptor interactions are observed with ecl2 residues W182$^{45.49}$, and F184$^{45.51}$, respectively (in F, L1, and L2).

Table IV
Virtual Screening of a Database of 990 Decoys and 10 Known AA3R Antagonists

Ligand[a]	Nr	Model F[b]	TM[c]	L1[d]	L2[e]
 MRS 1220	14				
 MRE 3008-F20	15				
 FA385	16				
	17				
	18				
 MRS-1523	19				
 MRS1334	20				
 VUF 5574	21				
	22				
 PSB-11	23				
Yield[f]		40	10	30	30
AUROC value[g]		0.81	0.57	0.59	0.72

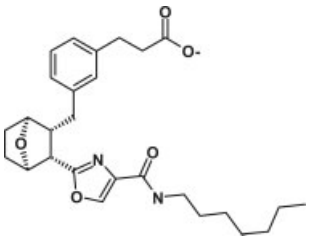
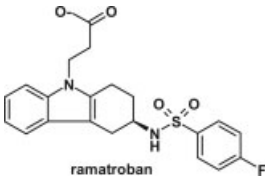
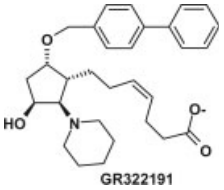
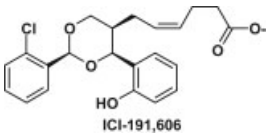
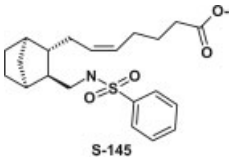
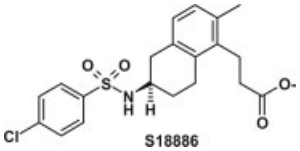
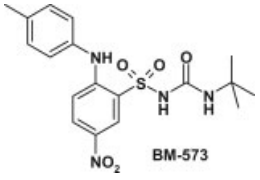
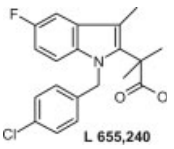
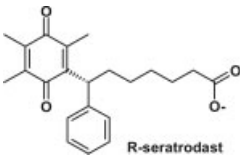
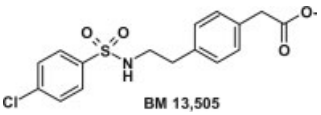
[a]Ligands have been selected from the literature.[32,58,65,83–85] Black, dark grey, and light grey shades indicate that the current antagonist has been ranked in top 5, 10, and 20% of the database, respectively.
[b]Full model with all intracellular and extracellular loops (except the N-terminal and C-terminal loops).
[c]Loop less TM.
[d]TM Model with only a 5 residue segment of ecl2 modeled with Modeller.
[e]TM model with only a 5 residue segment of ecl2 modeled by direct threading onto the bRho crystal structure backbone.
[f]Percentage of known AA3R antagonists ranked among the top 5% scorers (IFP scoring[63] using compound **2** (Fig. 3) as reference.
[g]Area under the curve of the ROC plot.

R7.40 is less accessible in model L1, and the orientations of the two bulky W182$^{45.49}$ and F184$^{45.51}$ residues decrease the volume of pocket i, as compared to models F, TM, and L2. This results in the loss of ionic/H-bond, hydrophobic, and aromatic ligand–receptor interactions observed for the SQ29542-TA2R reference complex and a lower IFP ranking of compound **28** (Table V).

DISCUSSION

The primary aim of the current study was to evaluate the effect of different ecl2 modeling strategies on structure-based virtual screening for GPCR antagonists. For a set of 3 unrelated GPCR targets (DRD2, AA3R, and TA2R) for which the second extracellular loop is experimentally known to affect antagonist binding, the implications of including ecl2 was evaluated in terms of structure-based virtual screening accuracy: the suitability of

Table V
Virtual Screening of a Database of 990 Decoys and 10 Known TA2R Antagonists

Ligand[a]	Nr	Model F[b]	Model TM[c]	Model L1[d]	Model L2[e]
	24				
 ramatroban	25				
 GR322191	26				
 ICI-191,606	27				
 S-145	28				
 S18886	29				
 BM-573	30				
 L 655,240	31				
 R-seratrodast	32				
 BM 13,505	33				
Yield[f]		50	60	20	20
AUROC value[g]		0.87	0.88	0.74	0.71

[a]Ligands have been selected from the literature.[20,56,57] Black, dark grey, and light grey shades indicate that the current antagonist has been ranked in top 5, 10, and 20% of the database, respectively.
[b]Full model with all intracellular and extracellular loops (except the N-terminal and C-terminal loops).
[c]Loop less TM.
[d]TM Model with only a 5 residue segment of ecl2 modeled with Modeller.
[e]TM model with only a 5 residue segment of ecl2 modeled by direct threading onto the bRho crystal structure backbone.
[f]Percentage of known TA2R antagonists ranked among the top 5% scorers (IFP scoring[63] using compound **4** (Fig. 3) as reference.
[g]Area under the curve of the ROC plot.

the 3D models to distinguish between known antagonists and randomly chosen drug-like compounds using automated docking approaches.

Bovine rhodopsin is an "all-round" template for modeling the ecl2 of GPCRs

Our amino acid sequence analysis of the ecl2s of 325 non-olfactory nonredundant human GPCRs containing the conserved C3.25-C45.50 disulfide link shows that the ecl2 loop length downstream from C45.50 is less variable than the total ecl2 loop length (Fig. 4, Table II). Moreover, most of the receptors have an ecl2 loop length similar (11–15) to that of bRho (13). Although this does not necessarily mean that the structure of the upstream ecl2 loop is conserved among GPCR receptors, this suggests that bRho is a relatively suitable structural template for constructing ecl2s for other GPCRs. Nevertheless, automated high-throughput modeling of ecl2s should be performed with great caution. Despite the conservation of upstream ecl2 loop length, automatic alignment of GPCR ecl2s is not straightforward since ecl2 sequen-

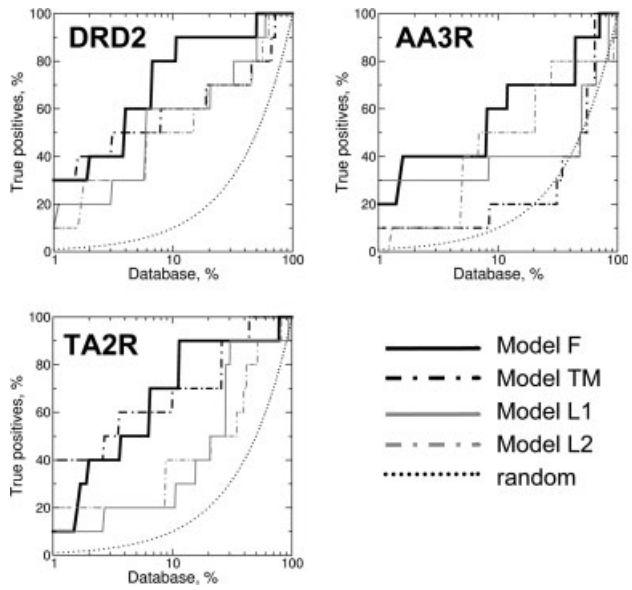


Figure 9

Enrichment in virtual screening of a database of 990 drug-like compounds[91] and 10 known antagonists (true positives, see Tables III–V) against: F (solid black line, model with all full icls and ecls), TM (dash-dot black line, model without any loops), L1 (solid grey line, model with (partial) ecl2 automatically modeled with Modeller), and L2 (dash-dot grey line, model with (partial) ecl2 automatically modeled by direct threading onto bRho crystal structure backbone) models of DRD2, AA3R, and TA2R. Docking simulations are performed with GOLD-Goldscore (DRD2 and TA2R) or GOLD-Chemscore (AA3R) in combination with IFP scoring. The dotted black line represents the fraction of actives expected by random picking.

ces show homology for only some GPCR receptor clusters (opsins, prostanoids, glycoproteins, secretin, adhesion) separately, and primarily in the region close to C45.50. Secondly, bRho seems to be a less suitable template for modeling the downstream part of the ecl2 loop, as it is relatively long (12 residues) compared to most of the other GPCRs (Table I). Furthermore, the ecl2 conformation can be greatly influenced by interactions with other parts of the receptor. In bRho, the N-terminal part is in direct contact with ecl2, while in some GPCRs, possibly additional disulfide links exist between ecl2 and other cysteine residues than C3.25.[38–40]

High-throughput modeling of the TM helices has shown to yield structures suitable for *in silico* inverse screening purposes,[22] and for detecting key residues that drive ligand selectivity.[5] A reasonable way to model GPCR ecl2s in a high-throughput fashion would be to only model the biggest possible loop segment one has the most confident in. The herein proposed ecl2 modeling flowchart utilizes several strategies according to local homology of the ecl2 target sequence to that of bRho (Fig. 2). Decreasing homology scenarios ranging from full sequence homology (with absolute length and C3.25–C45.50 conservation) to local homology on a C45.50-centered five residue window (see classification in Table II) enable the selection of the most appropriate modeling strategy for the longest possible loop segment.

Picking the pose: interaction fingerprint scoring

Selection of the most appropriate docking/scoring protocol in structure-based virtual screening is very much dependent on physicochemical details of target-ligand interactions[91–94] and fine details of the protein structure.[95,96] This makes it 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 strategies for selecting and ranking docking poses have recently received much attention as an alternative approach to solve the problem of protein–ligand docking and scoring accuracy.[97] In the current study, we have used a recently developed Tanimoto metric measuring the similarity between protein–ligand IFP for ranking docking poses.[70] For each of the three receptors (DRD2, AA3R, TA2R), binding modes of known high-affinity antagonists in line with available SAR/pharmacophore and site-directed mutagenesis data were used as IFP references. Docking simulations of low-molecular weight compounds, but also docking simulations 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 fast energy-based scoring functions.[70] 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.[98] Our IFP scoring protocol, however, could discriminate between irrelevant ligand docking poses and docking poses comparable to that of high affinity reference antagonists, as exemplified for the TA2R receptor in Figure 10. In the current study, IFP-scoring performed significantly better than the native Goldscore/Chemscore scoring functions in distinguishing between known antagonists and randomly chosen drug-like molecules. Of course, IFP ranking assumes the choice of a reference compound and corresponding binding mode. In the present report, a single reference antagonist was manually selected for each target receptor, based on its high affinity and selectivity, and the availability of site-directed mutagenesis data (in ecl2) to guide the definition of the reference binding mode. Its is however advised, in a prospective VS experiment, to define as many different IFP references as possible and later use machine learning algorithms (e.g. Bayesian modeling) to discriminate relevant from irrelevant poses in the hit triage step.[70]

Loopless transmembrane models of GPCR receptors can be suitable targets for virtual screening

As the total number of compounds in the database (1000 molecules), the number of true positives (10

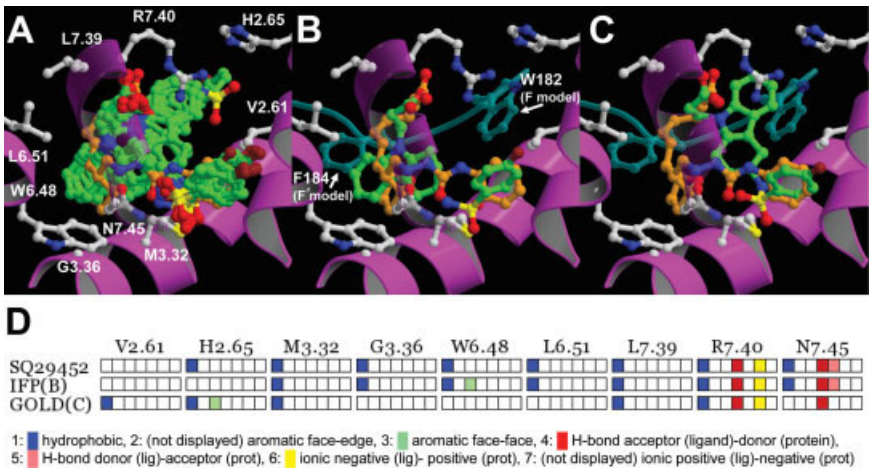


Figure 10

*Interaction Fingerprint (IFP) scoring of docking poses of ramatroban (compound **25**, Table V) in the TM model of TA2R (panel **A**), using SQ29,452 (orange carbon atoms, panel **B** and **C**) as a reference. Rendering and colour coding is the same as in Figure 5. Only parts of TMs 2, 3, 6, and 7 are shown, of which the last one is shown semitransparent. Seven different interaction types (hydrophobic interactions, aromatic face-to-edge, aromatic face-to-face, negatively charged, positively charged, H-bond acceptor, H-bond donor) were used to define the IFP. The cavity used for the IFP analysis consist of the 30 residues used earlier to define the TM binding pockets[5] and two extra residues at positions 3.37 and 7.40 (see "Computational Methods"). The IFP bit-strings (panel **D**) of the ramatroban docking poses with the highest Tc-IFP Tanimoto score (panel B) and highest Goldscore (panel C) are compared to the IFP of the SQ29452 reference pose. For reasons of clarity, the bit strings of only seven residues are shown as an example. The ecl2 loop and ecl2 residues W182$^{45.49}$ and F183$^{45.51}$ as located in the F model of TA2R (Fig. 7) are depicted semitransparent in cyan (panels B and C). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

known antagonists), and the number of compounds in the hit list considered for analyzing our virtual screening studies (top 50-ranked molecules), is the same for all models, the enrichment factor over random picking can be directly derived from the reported yields at the top 5% (Enrichment Factor = Yield/5). The enrichment factors (8–12) and yields (40–60%) obtained in our virtual screening studies using full (F) receptor models are comparable to the enrichment factors and yields obtained with optimal single and double consensus scoring protocols in previous antagonist screening studies against DRD3, ACM1, and V1AR homology models[25], as well as the enrichment factors and yields obtained with other retrospective structure-based screening studies in GPCRs.[23,79] In fact, for two of the three receptor test cases, DRD2 and TA2R, receptors for which residues in ecl2 have experimentally been shown to be involved in antagonist binding,[6,20] the virtual screening results obtained with the loopless TM models were as good as that obtained with the full receptor model containing all tailor-made ecls (F). For the AA3R receptor, however, explicit modeling of the ecl2 loop was shown to be essential for successful virtual screening.

The effects of ecl2 on structure-based virtual screening are GPCR specific

To avoid biasing virtual screening results, caution was given to select 990 drug-like decoys covering similar property ranges as the true actives but structurally different from any known active. Therefore, it can be assumed that the ecl2 loop affected virtual screening results by introducing specific additional receptor–ligand interaction points, rather than merely acting as a "steric" filter discriminating between low and high molecular weight compounds. In DRD2 the interactions between antagonists and ecl2 are of hydrophobic nature (I183$^{45.51}$ and I184$^{45.52}$). The residues in the ecl2 of TA2R (W182$^{45.49}$ and F184$^{45.51}$) interact with antagonists via hydrophobic/aromatic interactions. In the AA3R receptor, ecl2 residues serve as aromatic (F168$^{45.52}$) as well as H-bond donor/acceptor (Q167$^{45.51}$) interaction points. The involvement of I183$^{45.51}$ and I184$^{45.52}$ in DRD2-antagonist binding, F184$^{45.51}$ in TA2R-antagonist binding, and Q167$^{45.51}$ in AA3R-antagonist binding are supported by site-directed mutagenesis studies.[6,20,30] The roles of W182$^{45.49}$ and F184$^{45.51}$ in TA2R- and AA3R-ligand binding, however, remain to be investigated experimentally. Although another residue involved in AA3R-ligand H-bond interactions is located in TM5 (N6.55), it seems that Q167$^{45.51}$ is very important for successful virtual screening studies, driving ligands containing multiple H-bond donors and acceptors such as adenosine receptor antagonists into the proper binding mode. Residue F168$^{45.52}$, however, seems to be equally important, serving as an aromatic pi-stacking binding site for the aromatic moieties of antagonists together with F5.43. Another reason for the erroneous binding modes and low virtual screening accuracies of antagonists in the loopless AA3R TM model is that by omitting ecl2, an area normally occupied by this loop

might serve as a biased interaction surface for docking poses. This suggests that the possible effects of ecl2 residues on antagonist binding might not only depend on the physicochemical properties of the residues in direct proximity of C45.50, but also on the binding mode of the ligand with the ecl2 loop.

The three GPCR receptors investigated in the current study all show different antagonist binding orientations in the TM binding pocket (Figs. 5–7). Most DRD2 antagonists bind in both subpocket i (TMs 1, 2, 3, 7) and ii (TMs 3, 4, 5, 6), while most antagonists in AA3R and TA2R bind in subpockets ii and i, respectively. Omitting ecl2 from the GPCR receptor model generally does not only make the residues facing the binding pocket at the top of TMs 1, 2, 3, 5, 6, and 7 more accessible, it also makes TM4 (located in subpocket ii) much more accessible compared to a receptor model containing ecl2. This means that for ligands binding both in subpockets i and ii (DRD2, Fig. 5), or only in subpocket i (TA2R, Fig. 7), excluding ecl2 from the receptor model will probably have less drastic effects on the potential ligand–receptor interaction surface, than for ligands binding only in pocket ii (AA3R, Fig. 6).

Residues in the second extracellular loop of GPCRs as potential receptor-specific interaction points

Aminergic, prostanoid, and adenosine receptors show high similarity in the residues lining the TM binding pocket.[5] In Figure 8, segments of the ecl2s of the prostanoid cluster, the adenosine cluster, and a subgroup of the amine cluster are aligned. The ecl2 residues of prostanoid receptors are generally well conserved, showing a conserved "(S/T)WCF" motif around C45.50. In the amine receptor subgroup depicted in Figure 8, most receptors have an aliphatic residue at position 45.52, with the exception of DRD1 and DRD5 which contain a serine residue at this position. Other residues in the close proximity of C45.50 (e.g., positions 45.49 and 45.51), however, show large variability for the different aminergic receptors. The four adenosine receptors (AA1R, AA2AR, AA2BR, and AA3R) contain a conserved phenylalanine residue at position 45.52, while the two gonadotropin-releasing hormones receptors of the adenosine cluster (GNRHR, GNRR2) possess a serine residue at this position. Two adenosine receptors (AA3R and AA1R) have a polar residue (glutamine and glutamate, respectively) at position 45.51, while the other members of the adenosine cluster possess an aliphatic residue at this position. The residue at position 45.49 varies across all members of the adenosine receptor cluster.

The above exemplified similarity and dissimilarity of ecl2 regions involved in ligand binding offer possibilities to identify ecl2 residues which can potentially drive subtype-specific ligand selectivity. Our ecl2 sequence analysis of most human GPCRs further indicated that only few GPCRs show high similarity in the direct proximity of C45.50 (frizzled, prostanoids, glycoproteins, and adhesion receptors), some receptors show some conserved similarity (opsins, melanotonins, and opiates receptors), while most other GPCR clusters show little (vasopeptides, adenosine, acids, biogenic amines, purines) or no significant similarity (SREB, glutamate, peptides, chemoattractants, chemokines, brain-gut peptides). Careful and detailed consideration of possible receptor-specific ecl2-ligand interactions could therefore facilitate the design of new subtype-specific drugs.

It should be noted that the effect of ecl2 on ligand binding also depends on the binding mode of the antagonist in the TM binding pocket towards ecl2. Ligands binding in subpocket i for example will probably primarily interact with residues at positions 45.49 and 45.51 (like in TA2R), and ligands binding in subpocket ii will possibly only be able to interact with residues at positions 45.51 and 45.52 (like AA3R). Ligands binding in pocket i and ii (like DRD2) might even interact with all three residues (45.49, 45.50, and 45.51). To complicate things even further, it should be stated that also residues not in the direct vicinity of C45.50 might affect antagonist binding, directly (interacting with the ligand) or indirectly (e.g., stabilizing the ecl2 conformation via intramolecular H-bonding with TM helices like in TA2R). Modeling these residues in a high-throughput fashion is highly speculative as they are located in portions of the ecl2 loop which cannot directly be derived from the bRho structural template. A tailored loop modeling procedure, guided by experimental (SAR, site-directed mutagenesis) data, however, can yield GPCR models with increased value for virtual screening studies and interpretation of experimental site-directed mutagenesis and ligand binding studies.

CONCLUSIONS

The current study shows that consideration of the second extracellular loop in GPCR homology models can lead to an increase in structure-based virtual screening accuracy, but that this effect is rather receptor-specific. Sequence analysis of the ecl2 of most human GPCRs indicated that bRho is a relatively suitable modeling template for modeling the upstream ecl2 segment up to the end of β_4. 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. Moreover, loopless TM models of GPCR receptors can be suitable targets for virtual screening, using proper post-processing strategies, such as interaction fingerprint scoring, to select automated docking poses in line with experimentally known ligand–receptor interactions. The

ecl2s of the three receptor test cases covered a wide range in loop lengths upstream and downstream from C45.50, were different with respect to the physicochemical properties of the amino acid residues in the ecl2 region close to C45.50, and showed very different ecl2-antagonist binding modes. This might explain why for two of the three receptor test cases, DRD2 and TA2R, virtual screening accuracies in the loopless TM models were comparable to the virtual accuracies obtained with full receptor models, while for the AA3R receptor, the full model outperformed the TM model in terms of virtual screening accuracy. Automated docking studies with ligands binding primarily in sub pocket ii (between TM3, TM4, TM5, TM6, and TM7) and interacting with ecl2 residues downstream of C45.50 via H-bond interactions might have a higher chance of yielding biased docking poses when ecl2 is omitted from the GPCR receptor model. Nevertheless our ecl2 sequence analysis indicated that many GPCR receptors sharing high similarity in the residues lining the TM cavity, show low ecl2 sequence homology. As a conclusion, explicit modeling of ecl2 for structure-based *in silico* screening is only justified either in the presence of strong sequence homology to bRho or by use of appropriate ligand-derived topological restraints. In the absence of such conditions, we strongly advise to use loopless TM models.

ACKNOWLEDGMENTS

The authors wish to thank AstraZeneca (Mölndal, Sweden) for numerous discussions with many scientists worldwide. The Centre Informatique National de l'Enseignement Supérieur (CINES, Montpellier, France) and the Institut de Recherches Informatique Scientifique (IDRIS, Orsay, France) are acknowledged for allocation of computing time. Nathanael Weill is acknowledged for writing the IFP image creator tool used for making panel D of Figure 10.

REFERENCES

1. Klabunde T, Hessler G. Drug design strategies for targeting G-protein-coupled receptors. Chembiochem 2002;3:928–944.
2. Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G. GPCRDB information system for G protein-coupled receptors. Nucleic Acids Res 2003;31:294–297.
3. 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.
4. Ji TH, Grossmann M, Ji I. G protein-coupled receptors. I. Diversity of receptor-ligand interactions. J Biol Chem 1998;273:17299–17302.
5. Surgand JS, Rodrigo J, Kellenberger E, Rognan D. A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. Proteins 2006;62:509–538.
6. Shi L, Javitch JA. The second extracellular loop of the dopamine D2 receptor lines the binding-site crevice. Proc Natl Acad Sci USA 2004;101:440–445.
7. Baneres JL, Mesnier D, Martin A, Joubert L, Dumuis A, Bockaert J. Molecular characterization of a purified 5-HT4 receptor: a structural basis for drug efficacy. J Biol Chem 2005;280:20253–20260.
8. Li J, Edwards PC, Burghammer M, Villa C, Schertler GF. Structure of bovine rhodopsin in a trigonal crystal form. J Mol Biol 2004;343:1409–1438.
9. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: a G protein-coupled receptor. Science 2000;289:739–745.
10. Kraulis PJ. Molscript–a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:946–950.
11. Merritt EA, Murphy MEP. Raster3d Version-2.0–a program for photorealistic molecular graphics. Acta Crystallogr D Biol Crystallogr 1994;50:869–873.
12. Mirzadegan T, Benko G, Filipek S, Palczewski K. Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. Biochemistry 2003;42:2759–2767.
13. Bourne HR, Meng EC. Structure. Rhodopsin sees the light. Science 2000;289:733–734.
14. Menon ST, Han M, Sakmar TP. Rhodopsin: structural basis of molecular physiology. Physiol Rev 2001;81:1659–1688.
15. 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;49:683–691.
16. Wurch T, Pauwels PJ. Coupling of canine serotonin 5-HT(1B) and 5-HT(1D) receptor subtypes to the formation of inositol phosphates by dual interactions with endogenous G(i/o) and recombinant G(alpha15) proteins. J Neurochem 2000;75:1180–1189.
17. Gao ZG, Chen A, Barak D, Kim SK, Müller CE, Jacobson KA. Identification by site-directed mutagenesis of residues involved in ligand recognition and activation of the human A3 adenosine receptor. J Biol Chem 2002;277:19056–19063.
18. Archer-Lahlou E, Tikhonova I, Escrieut C, Dufresne M, Seva C, Pradayrol L, Moroder L, Maigret B, Fourmy D. Modeled structure of a G-protein-coupled receptor: the cholecystokinin-1 receptor. J Med Chem 2005;48:180–191.
19. Lei B, Morris DP, Smith MP, Svetkey LP, Newman MF, Rotter JI, Buchanan TA, Beckstrom-Sternberg SM, Green ED, Schwinn DA. Novel human α1a-adrenoceptor single nucleotide polymorphisms alter receptor pharmacology and biological function. Naunyn Schmiedebergs Arch Pharmacol 2005;371:229–239.
20. Khasawneh FT, Huang JS, Turek JW, Le Breton GC. Differential mapping of the amino acids mediating agonist and antagonist coordination with the human thromboxane A2 receptor protein. J Biol Chem 2006;281:26951–26965.
21. Klco JM, Wiegand CB, Narzinski K, Baranski TJ. Essential role for the second extracellular loop in C5a receptor activation. Nat Struct Mol Biol 2005;12:320–326.
22. 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;44:1162–1176.
23. Becker OM, Marantz Y, Shacham S, Inbal B, Heifetz A, Kalid O, Bar-Haim S, Warshaviak D, Fichman M, Noiman S. G protein-coupled receptors: in silico drug discovery in 3D. Proc Natl Acad Sci USA 2004;101:11304–11309.
24. Jacobsen RB, Sale KL, Ayson MJ, Novak P, Hong J, Lane P, Wood NL, Kruppa GH, Young MM, Schoeniger JS. Structure and dynamics of dark-state bovine rhodopsin revealed by chemical cross-linking and high-resolution mass spectrometry. Protein Sci 2006;15:1303–1317.
25. 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.

26. Rodrigo J, Pena A, Murat B, Trueba M, Durroux T, Guillon G, Rognan D. Mapping the binding site of arginine vasopressin to v1a and v1b vasopressin receptors. Mol Endocrinol 2007;21:512–523.
27. Freddolino PL, Kalani MY, Vaidehi N, Floriano WB, Hall SE, Trabanino RJ, Kam VW, Goddard WA III. Predicted 3D structure for the human β 2 adrenergic receptor and its binding site for agonists and antagonists. Proc Natl Acad Sci USA 2004;101:2736–2741.
28. Mehler EL, Hassan SA, Kortagere S, Weinstein H. Ab initio computational modeling of loops in G-protein-coupled receptors: lessons from the crystal structure of rhodopsin. Proteins 2006;64:673–690.
29. Zhang Y, Devries ME, Skolnick J. Structure modeling of all identified G protein-coupled receptors in the human genome. PLoS Comput Biol 2006;2:e13.
30. Duong HT, Gao ZG, Jacobson KA. Nucleoside modification and concerted mutagenesis of the human A3 adenosine receptor to probe interactions between the 2-position of adenosine analogs and Gln167 in the second extracellular loop. Nucleos Nucleot Nucleic Acids 2005;24:1507–1517.
31. Seeman P. Targeting the dopamine D2 receptor in schizophrenia. Expert Opin Ther Targets 2006;10:515–531.
32. Jacobson KA, Gao ZG. Adenosine receptors as therapeutic targets. Nat Rev Drug Discov 2006;5:247–264.
33. Dogne JM, Hanson J, de Leval X, Pratico D, Pace-Asciak CR, Drion P, Pirotte B, Ruan KH. From the design to the clinical application of thromboxane modulators. Curr Pharm Des 2006;12:903–923.
34. 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.
35. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. The universal protein resource (UniProt): an expanding universe of protein information. Nucleic Acids Res 2006;34:D187–D191.
36. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000;302: 205–217.
37. Schulein R, Zuhlke K, Oksche A, Hermosilla R, Furkert J, Rosenthal W. The role of conserved extracellular cysteine residues in vasopressin V2 receptor function and properties of two naturally occurring mutant receptors with additional extracellular cysteine residues. FEBS Lett 2000;466:101–106.
38. Stitham J, Arehart EJ, Gleim SR, Douville KL, Hwa J. Human prostacyclin receptor structure and function from naturally-occurring and synthetic mutations. Prostaglandins Other Lipid Mediat 2007;82:95–108.
39. Davidson JS, Assefa D, Pawson A, Davies P, Hapgood J, Becker I, Flanagan C, Roeske R, Millar R. Irreversible activation of the gonadotropin-releasing hormone receptor by photoaffinity cross-linking: localization of attachment site to Cys residue in N-terminal segment. Biochemistry 1997;36:12881–12889.
40. Noda K, Saad Y, Graham RM, Karnik SS. The high affinity state of the β 2-adrenergic receptor requires unique interaction between conserved and non-conserved extracellular loop cysteines. J Biol Chem 1994;269:6743–6752.
41. Olah ME, Jacobson KA, Stiles GL. Role of the second extracellular loop of adenosine receptors in agonist and antagonist binding. Analysis of chimeric A1/A3 adenosine receptors. J Biol Chem 1994; 269:24692–24698.
42. Hoffmann C, Moro S, Nicholas RA, Harden TK, Jacobson KA. The role of amino acids in extracellular loops of the human P2Y1 receptor in surface expression and activation processes. J Biol Chem 1999;274:14639–14647.
43. Blanpain C, Doranz BJ, Bondue A, Govaerts C, De Leener A, Vassart G, Doms RW, Proudfoot A, Parmentier M. The core domain of chemokines binds CCR5 extracellular domains while their amino terminus interacts with the transmembrane helix bundle. J Biol Chem 2003;278:5179–5187.
44. Qi LJ, Leung AT, Xiong Y, Marx KA, Abou-Samra AB. Extracellular cysteines of the corticotropin-releasing factor receptor are critical for ligand interaction. Biochemistry 1997;36:12442–12448.
45. Tokita K, Hocart SJ, Coy DH, Jensen RT. Molecular basis of the selectivity of gastrin-releasing peptide receptor for gastrin-releasing peptide. Mol Pharmacol 2002;61:1435–1443.
46. Prossnitz ER, Schreiber RE, Bokoch GM, Ye RD. Binding of low affinity *N*-formyl peptide receptors to G protein. Characterization of a novel inactive receptor intermediate. J Biol Chem 1995;270: 10686–10694.
47. Perez HD, Vilander L, Andrews WH, Holmes R. Human formyl peptide receptor ligand binding domain(s). Studies using an improved mutagenesis/expression vector reveal a novel mechanism for the regulation of receptor occupancy. J Biol Chem 1994;269: 22485–22487.
48. Fowler CB, Pogozheva ID, Lomize AL, LeVine H III, Mosberg HI. Complex of an active μ-opioid receptor with a cyclic peptide agonist modeled from experimental constraints. Biochemistry 2004;43: 15796–15810.
49. Stitham J, Stojanovic A, Ross LA, Blount AC Jr, Hwa J. Clusters of transmembrane residues are critical for human prostacyclin receptor activation. Biochemistry 2004;43:8974–8986.
50. TRIPOS Associates Inc. St-Louis, MO.
51. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem 2004;25:1157–1174.
52. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein–ligand docking using GOLD. Proteins 2003;52: 609–623.
53. Brady GP, Pieter FWS. Fast prediction and visualization of protein binding pockets with PASS. J Comput Aided Mol Des 2000;V14: 383–401.
54. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
55. Lundstrom K, Turpin MP, Large C, Robertson G, Thomas P, Lewell XQ. Mapping of dopamine D3 receptor binding site by pharmacological characterization of mutants expressed in CHO cells with the Semliki Forest virus system. J Recept Signal Transduct Res 1998;18: 133–150.
56. Funk CD, Furci L, Moran N, Fitzgerald GA. Point mutation in the seventh hydrophobic domain of the human thromboxane A2 receptor allows discrimination between agonist and antagonist binding sites. Mol Pharmacol 1993;44:934–939.
57. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
58. Urizar E, Claeysen S, Deupi X, Govaerts C, Costagliola S, Vassart G, Pardo L. An activation switch in the rhodopsin family of G protein-coupled receptors: the thyrotropin receptor. J Biol Chem 2005;280:17135–17141.
59. Coley C, Woodward R, Johansson AM, Strange PG, Naylor LH. Effect of multiple serine/alanine mutations in the transmembrane spanning region V of the D2 dopamine receptor on ligand binding. J Neurochem 2000;74:358–366.
60. Homan EJ, Copinga S, Unelius L, Jackson DM, Wikstrom HV, Grol CJ. Synthesis and pharmacology of the enantiomers of the potential atypical antipsychotic agents 5-OMe-BPAT and 5-OMe-(2,6-di-OMe)-BPAT. Bioorg Med Chem 1999;7:1263–1271.
61. Vangveravong S, McElveen E, Taylor M, Xu J, Tu Z, Luedtke RR, Mach RH. Synthesis and characterization of selective dopamine D2 receptor antagonists. Bioorg Med Chem 2006;14:815–825.
62. Wood MD, Scott C, Clarke K, Westaway J, Davies CH, Reavill C, Hill M, Rourke C, Newson M, Jones DN, Forbes IT, Gribble A. Aripiprazole and its human metabolite are partial agonists at the human dopamine D2 receptor, but the rodent metabolite displays antagonist properties. Eur J Pharmacol 2006;546:88–94.
63. Misra RN, Brown BR, Sher PM, Patel MM, Hall SE, Han WC, Barrish JC, Kocy O, Harris DN, Goldenberg HJ, Michel IM,

Schumacher WA, Webb ML, Monshizadegan H, Ogletree ML. Interphenylene 7-oxabicyclo[2.2.1]heptane oxazoles. Highly potent, selective, and long-acting thromboxane A2 receptor antagonists. J Med Chem 1993; 36:1401–1417.
64. Nicolai E, Goyard J, Benchetrit T, Teulon JM, Caussade F, Virone A, Delchambre C, Cloarec A. Synthesis and structure-activity relationships of novel benzimidazole and imidazo[4,5-b]pyridine acid derivatives as thromboxane A2 receptor antagonists. J Med Chem 1993; 36:1175–1187.
65. Press NJ, Taylor RJ, Fullerton JD, Tranter P, McCarthy C, Keller TH, Brown L, Cheung R, Christie J, Haberthuer S, Hatto JD, Keenan M, Mercer MK, Press NE, Sahri H, Tuffnell AR, Tweed M, Fozard JR. A new orally bioavailable dual adenosine A2B/A3 receptor antagonist with therapeutic potential. Bioorg Med Chem Lett 2005;15:3081–3085.
66. SciTegic Inc., San Diego, CA 92123–1365, USA. 4.2.
67. Molecular Networks GmbH, D-91052 Erlangen, Germany.
68. OpenEye Scientific software, Santa Fe, NM 87507.
69. Bywater RP. Location and nature of the residues important for ligand recognition in G-protein coupled receptors. J Mol Recognit 2005;18:60–72.
70. Marcou G, Rognan D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J Chem Inf Model 2007;47:195–207.
71. Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. J Med Chem 2005;48:2534–2547.
72. Moro S, Gao ZG, Jacobson KA, Spalluto G. Progress in the pursuit of therapeutic adenosine receptor antagonists. Med Res Rev 2006;26: 131–159.
73. Shi L, Javitch JA. The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop. Annu Rev Pharmacol Toxicol 2002;42:437–467.
74. Stitham J, Stojanovic A, Merenick BL, O'Hara KA, Hwa J. The unique ligand-binding pocket for the human prostacyclin receptor. Site-directed mutagenesis and molecular modeling. J Biol Chem 2003;278:4250–4257.
75. Kim SK, Gao ZG, Jeong LS, Jacobson KA. Docking studies of agonists and antagonists suggest an activation pathway of the A(3) adenosine receptor. J Mol Graph Model 2006;25:562–577.
76. Javitch JA, Ballesteros JA, Weinstein H, Chen J. A cluster of aromatic residues in the sixth membrane-spanning segment of the dopamine D2 receptor is accessible in the binding-site crevice. Biochemistry 1998;37:998–1006.
77. Woodward R, Daniell SJ, Strange PG, Naylor LH. Structural studies on D2 dopamine receptors: mutation of a histidine residue specifically affects the binding of a subgroup of substituted benzamide drugs. J Neurochem 1994;62:1664–1669.
78. Shacham S, Marantz Y, Bar-Haim S, Kalid O, Warshaviak D, Avisar N, Inbal B, Heifetz A, Fichman M, Topf M, Naor Z, Noiman S, Becker OM. PREDICT modeling and in-silico screening for G-protein coupled receptors. Proteins 2004;57:51–86.
79. 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;48:5448–5465.
80. Simpson MM, Ballesteros JA, Chiappa V, Chen J, Suehiro M, Hartman DS, Godel T, Snyder LA, Sakmar TP, Javitch JA. Dopamine D4/D2 receptor selectivity is determined by A divergent aromatic microdomain contained within the second, third, and seventh membrane-spanning segments. Mol Pharmacol 1999;56:1116–1126.
81. Jacobson KA, Gao ZG, Chen A, Barak D, Kim SA, Lee K, Link A, Rompaey PV, van Calenbergh S, Liang BT. Neoceptor concept based on molecular complementarity in GPCRs: a mutant adenosine A(3) receptor with selectively enhanced affinity for amine-modified nucleosides. J Med Chem 2001;44:4125–4136.
82. Chen A, Gao ZG, Barak D, Liang BT, Jacobson KA. Constitutive activation of A(3) adenosine receptors by site-directed mutagenesis. Biochem Biophys Res Commun 2001;284:596–601.
83. Zhao MM, Hwa J, Perez DM. Identification of critical extracellular loop residues involved in α1-adrenergic receptor subtype-selective antagonist binding. Mol Pharmacol 1996;50:1118–1126.
84. Wurch T, Colpaert FC, Pauwels PJ. Chimeric receptor analysis of the ketanserin binding site in the human 5-hydroxytryptamine1D receptor: importance of the second extracellular loop and fifth transmembrane domain in antagonist binding. Mol Pharmacol 1998;54:1088–1096.
85. Ruan CH, Wu J, Ruan KH. A strategy using NMR peptide structures of thromboxane A2 receptor as templates to construct ligand-recognition pocket of prostacyclin receptor. BMC Biochem 2005; 6:23.
86. Stillman BA, Audoly L, Breyer RM. A conserved threonine in the second extracellular loop of the human EP2 and EP4 receptors is required for ligand binding. Eur J Pharmacol 1998;357:73–82.
87. Audoly L, Breyer RM. The second extracellular loop of the prostaglandin EP3 receptor is an essential determinant of ligand selectivity. J Biol Chem 1997;272:13475–13478.
88. Scholl DJ, Wells JN. Serine and alanine mutagenesis of the nine native cysteine residues of the human A(1) adenosine receptor. Biochem Pharmacol 2000;60:1647–1654.
89. D'Angelo DD, Eubank JJ, Davis MG, Dorn GW II. Mutagenic analysis of platelet thromboxane receptor cysteines. Roles in ligand binding and receptor-effector coupling. J Biol Chem 1996;271: 6233–6240.
90. Ezumi K, Yamakawa M, Narisada M. Computer-aided molecular modeling of a thromboxane receptor antagonist S-145 and its related compounds. J Med Chem 1990;33:1117–1122.
91. Bissantz C, Folkers G, Rognan D. Protein-based virtual screening of chemical databases. I. Evaluation of different docking/scoring combinations. J Med Chem 2000;43:4759–4767.
92. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. A critical assessment of docking programs and scoring functions. J Med Chem 2006;49: 5912–5931.
93. Kellenberger E, Rodrigo J, Müller P, Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins 2004;57:225–242.
94. Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP. Comparison of automated docking programs as virtual screening tools. J Med Chem 2005;48:962–976.
95. Moitessier N, Therrien E, Hanessian S. A method for induced-fit docking, scoring, and ranking of flexible ligands. Application to peptidic and pseudopeptidic β-secretase (BACE 1) inhibitors. J Med Chem 2006;49:5885–5894.
96. Kramer B, Rarey M, Lengauer T. Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. Proteins 1999;37:228–241.
97. Kämper A, Rognan D, Lengauer T. Lead identification in virtual screening. In: Lengauer T, editor. Bioinformatics–from gene to therapies. Weinheim: Wiley-VCH; 2007, pp 651–693.
98. Evers A, Klabunde T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the α1A adrenergic receptor. J Med Chem 2005;48:1088–1097.