Protein Engineering, Design & Selection vol. 25 no. 8 pp. 377–386, 2012
Published online June 12, 2012 doi:10.1093/protein/gzs037

Automatic modeling of mammalian olfactory receptors and docking of odorants

Guillaume Launay[1], Stéphane Téletchéa[1,4,5], Fallou Wade[2,3], Edith Pajot-Augy[2,3], Jean-François Gibrat[1,6] and Guenhaël Sanz[2,3]

[1]INRA, Mathématique, Informatique et Génome UR1077, 78350 Jouy-en-Josas, France, [2]INRA, Neurobiologie de l'Olfaction et Modélisation en Imagerie UR1197, 78350 Jouy-en-Josas, France, [3]IFR 144 Neuro-Sud Paris, France, [4]INSERM, UMR 957, Nantes, F-44035 France and [5]Université de Nantes, Nantes Atlantique Universités, Laboratoire de Physiopathologie de la Résorption Osseuse et Thérapie des Tumeurs Osseuses Primitives, Nantes, F-44035 France

[6]To whom correspondence should be addressed.
E-mail: jean-francois.gibrat@jouy.inra.fr

Received September 8, 2011; revised May 8, 2012; accepted May 16, 2012

Edited by Peter Tieleman

We present a procedure that (i) automates the homology modeling of mammalian olfactory receptors (ORs) based on the six three-dimensional (3D) structures of G protein-coupled receptors (GPCRs) available so far and (ii) performs the docking of odorants on these models, using the concept of colony energy to score the complexes. ORs exhibit low-sequence similarities with other GPCR and current alignment methods often fail to provide a reliable alignment. Here, we use a fold recognition technique to obtain a robust initial alignment. We then apply our procedure to a human OR that we have previously functionally characterized. The analysis of the resulting *in silico* complexes, supported by receptor mutagenesis and functional assays in a heterologous expression system, suggests that antagonists dock in the upper part of the binding pocket whereas agonists dock in the narrow lower part. We propose that the potency of agonists in activating receptors depends on their ability to establish tight interactions with the floor of the binding pocket. We developed a web site that allows the user to upload a GPCR sequence, choose a ligand in a library and obtain the 3D structure of the free receptor and ligand–receptor complex (http://genome.jouy.inra.fr/GPCRautomodel).
Keywords: G protein-coupled receptors/ligand docking/ olfactory receptors/threading/3D modeling

Introduction

G protein-coupled receptors (GPCRs) form the largest class of membrane proteins and represent the largest gene family, accounting for 3–5% of the mammalian genome (Schlyer and Horuk, 2006). They are seven-transmembrane proteins and key components of several cellular response mechanisms to environmental signals (Gloriam *et al*., 2007). According to the GRAFS classification (Fredriksson *et al*., 2003), GPCRs are divided into five families. Class R family, which corresponds to the 'rhodopsin-like' receptors, is the most populated one, in part because it includes the very large olfactory receptor (OR) sub-family. ORs are present in almost all multicellular organisms. They account for more than half the GPCRs in mammalian species. For instance, the mouse OR repertoire is made of >1000 functional genes. They often exhibit very low-sequence identity.

Class R family is the only one for which structural information is currently available. The first experimental three-dimensional (3D) structure for a receptor, the bovine rhodopsin, was solved in 2000 (Palczewski *et al*., 2000). Since then, our knowledge about GPCR structures has expanded with the release of the 3D structures of the turkey β1-adrenergic receptor (Cherezov *et al*., 2007), squid rhodopsin (Murakami and Kouyama, 2008), human β2-adrenergic receptor (Warne *et al*., 2008), human A_{2A}-adenosine receptor (Jaakola *et al*., 2008), human dopamine D3 receptor (Chien *et al*., 2010) and human CXCR4 chemokine receptor (Wu *et al*., 2010). However, there is still a very important lack of receptor structures as, for instance in Humans, >700 active receptors are identified (half of them being ORs). To narrow the huge gap between known structures and receptors of interest to the biologists, there is a pressing need for high-throughput methods to determine these structures, or, in the meantime, for *in silico* methods able to produce a satisfactory model compatible with known experimental data.

With the availability of high-resolution structures, it became possible to combine traditional homology modeling techniques with ligand-based pharmacophore modeling (Evers and Klabunde, 2005). Most of these approaches have provided useful insights for drug design but are ill-suited for large scale studies. Homology modeling techniques are critically dependent on the initial sequence alignment between the query sequence and the sequence of a receptor whose 3D structure is known (the template). Receptors of the class R family exhibit relatively low-sequence identity, even within the rhodopsin subfamily of GPCRs (Gloriam *et al*., 2007). Automatic sequence alignment methods are prone to errors when aligning such low-identity sequences and human experts often need to correct the alignment manually. Human intervention, of course, makes the process difficult to automate for large scale studies and also less reproducible.

In this paper, we present a fully automated method to perform homology modeling of GPCR structures. It is based on an improved version (to be published) of a threading method that we developed a few years ago, FROST (Marin *et al*., 2002). Threading techniques are intended to detect and align remote homologs, i.e. homologs with low-sequence identities. To the best of our knowledge, only Zhang *et al*. (2006) have applied similar methods to model the 3D structure of GPCRs. In a second step, we carry out the docking of selected ligands with the modeled receptor using VINA (Trott and Olson, 2010) and we propose a new way of scoring the conformations of the resulting complexes. To validate our methodology, we first performed a cross-comparison of all receptors with known 3D structures, i.e. we

modeled the structure of each receptor in turn by aligning its sequence with the 3D structure of other receptors using FROST. X-ray experiments provide the structure of receptor–ligand complexes. We checked the docking of the modeled 3D structures with the corresponding ligand (except for rhodopsin whose ligand, the retinal, is covalently bound to the receptor). These cross-comparisons allowed us to accurately evaluate the performances of the proposed method.

We are chiefly interested in ORs, for which no experimental 3D structure is available. Therefore, to further assess our methodology, we performed two experiments. First, we modeled two ORs (the mouse mOR42-3 (Abaffy *et al.*, 2007) and mOR-EG (Katada *et al.*, 2005) receptors) for which models have previously been manually obtained by experts. Site-directed mutagenesis guided by the models and experimental assays were performed with different ligands to test the functional importance of some residues for these receptors. We checked the consistency of our models with these sets of residues. Second, we also modeled an OR already characterized by one of us (Sanz *et al.*, 2005) for which 3D-Quantitative Structure–Activity Relationship (QSAR) data are available (Sanz *et al.*, 2008): OR1G1 (UniProt P47890, HORDE 42-99). Using this model, we proposed a limited number of mutations and predicted their possible influence on the binding of various odorant molecules. The corresponding mutants were produced and their functional response to different odorant molecules experimentally tested. The results we obtained were in good agreement with the predictions.

We developed a web site (http://genome.jouy.inra.fr/GPCRautomodel) that allows the user to upload a GPCR sequence, choose a ligand in a library (or supply the required information for a new ligand) and obtain the 3D structure of the free receptor and ligand–receptor complex.

Materials and methods

Here we provide a brief description of the methodology used. A more detailed account is given in Supplementary Material.

Modeling of the 3D structure of the receptors

Alignment of the query receptor sequence with the sequences of the structural templates. The first step of the homology modeling procedure consists in aligning the sequence of the query protein with the sequence of the homologous proteins whose 3D structure is available (template proteins). This step is crucial and determines the accuracy of the resulting model. When the sequences of the query and template proteins are not very similar (e.g. $<25\%$ sequence identity), routine sequence alignment methods are prone to make alignment errors. To overcome this problem, it is necessary to use more advanced alignment methods such as profile/profile alignment methods or threading methods. Here we use an in-house threading method, FROST. FROST, unlike most other threading methods, does make use of the 3D structure of the template, i.e. it performs a sequence–structure alignment (Andonov *et al.*, 2008; Collet *et al.*, 2010). Using the 3D structure allows the algorithm to take into account residues that are in close contact in the structure although they may be far apart in the sequence. This additional piece of information is often critical to obtain the proper alignment when the sequence identity between the query and template proteins is low. We use a template library consisting of six of the seven GPCR currently available (listed in Supplementary Table ST1). For this work, we derived a new set of parameters dedicated to all-alpha membrane proteins (to be published). Parameters were computed using multiple sequence alignments of a set of all-alpha membrane proteins. This set was retrieved from the PDBTM database (Tusnady *et al.*, 2005) that comprises 285 proteins sharing $<30\%$ sequence identity. The sequence–structure alignment algorithm used in FROST is exact and it ensures that the alignment with the largest score, for a given score function, of the query sequence with the structural core is obtained. For the GPCR protein family, the structural core, i.e. the 3D structure conserved parts on which the query sequence is aligned, consists of the seven transmembrane helices (TMHs) (see Supplementary Material for further details regarding the FROST procedure).

Receptor 3D structure modeling. FROST alignment is transformed into a PIR alignment which is then used as input for MODELLER 9v6 (Sali and Blundell, 1993). MODELLER optimizes the side-chain rotamers and models the loops joining conserved core elements. We used the default MODELLER protocol to build the models. For each query sequence, template structure alignment, we chose to generate 10 models of the receptor 3D structure.

Ligand docking

Database of ligands. We have compiled a set of 124 ligands in our database. They are defined by a standard set of files obtained from the PRODRG web site (Schüttelkopf and van Aalten, 2004). The server makes use of the PDB coordinates, crystallography and NMR system topology and parameter files. Other ligands can be used for docking provided they comply with the PRODRG format.

Generation of complexes. Ligands are docked with the previously modeled receptors using the VINA software (Trott and Olson, 2010). VINA allows the ligand and the protein residue side chains to be flexible. Although all receptor residue side chains can potentially be treated as flexible, this would generate prohibitive computational times. For a given template, the set of flexible residue side chains is therefore limited to residues potentially involved in ligand recognition. Flexible residues are located in the well and rim (for each template structure, flexible residues are listed in Supplementary Table ST6).

Scoring of complexes. VINA provides its own score function to rank the complexes. However, this score function did not adequately rank the structures of the complexes obtained experimentally (see Supplementary Figs SF1 and SF2). We therefore developed the following scheme to rank the complexes.

All conformations provided by VINA were refined using the Xplor force field (Brunger *et al.*, 1992) using the analytical continuum electrostatics implementation of the generalized Born implicit solvent model and CHARMM param19 parameters. A previous study reported that the relaxation of the complex was important for the correct positioning of ligand in GPCR pocket (Freddolino *et al.*, 2004). We implemented a

two-stage relaxation procedure. First, the complexes produced by VINA were subjected to an all-atom minimization, keeping the receptor fixed but allowing the ligand to move. The receptor was then permitted to accommodate the ligand by keeping its backbone fixed but letting the side chains and the ligand move during a final set of 100 steps of Powell minimization. The final conformation was then checked for the type of ligand–receptor residue contacts. Conformations involving too many contacts with residues that were not part of the binding pocket were discarded. The remaining conformations were then scored for their receptor–ligand shape and electrostatic complementarities, based on the sum of their van der Waals and electrostatic interaction energies corrected for ligand desolvation:

$$E = E_{\mathrm{vdw}} + E_{\mathrm{elec}} + E_{\mathrm{solv}}^{\mathrm{bound}} - E_{\mathrm{solv}}^{\mathrm{free}}$$

where E_{vdw} is the van der Waals energy, E_{elec} is the electrostatic energy and $E_{\mathrm{solv}}^{\mathrm{bound}}$ and $E_{\mathrm{solv}}^{\mathrm{free}}$ are the solvation energies of the bound (respectively free) ligand.

To account for entropic effects, we implemented a modified version of the 'colony energy' statistical re-scoring scheme introduced by Xiang *et al*. (2002) for loop modeling and later used in the context of docking by Lee and Seok (2008):

$$E_{\mathrm{col}}(i) = <E> \ln \sum_j Q_{ij} \ \exp(E(j)/ <E>)$$

where $E(j)$ is the Xplor energy for the jth model, the summation is taken over all models that have a contact map similarity Q_{ij} greater than a threshold (empirically set to 0.5 here) and $<E>$ is the mean Xplor energy for all the models. $<E>$ is used for avoiding underflow/overflow problems with the exponential. Notice that if the Q_{ij} threshold is set to 1, the summation comprises only the ith model and $E_{\mathrm{col}}(i) = E(i)$, i.e. the Xplor energy. To measure the similarity between binding modes we used the receptor–ligand contacts maps (Launay and Simonson, 2011). These maps are sparse matrices in which rows refer to ligand atoms and columns to receptor residues. The contact map, C_i, for a particular complex i contains 1s for matrix elements corresponding to receptor residues and ligand atoms in contact in the 3D structure, 0s otherwise. Using this representation, the similarity Q_{ij} between two binding modes i, j can be computed as

$$Q_{ij} = \frac{N(C_i, C_j)}{\max(C_i, C_j)}$$

where $N(C_i,C_j)$ is the number of common contacts between the contact maps and $\max(C_i,C_j)$ the maximal number of contact found in either maps. Hence, Q_{ij} tends toward 1 as the complexes exhibit identical sets of contacts while being equal to 0 if their sets of contacts are completely different. The colony energy is used to rank the final docking conformations.

To estimate the accuracy of the ligand docking, we computed the interaction root-mean-square deviation (I_{rmsd}) between the predicted conformation and the corresponding GPCR–ligand experimental structure. I_{rmsd} is the heavy atoms rmsd between the ligands after superimposition of the receptors (L_{rms} quality measure in CAPRI contests, see Lensink *et al*., 2007).

Experimental validation of the OR1G1 model

Experimental techniques used in this work are essentially similar to those described in Sanz *et al*. (2005). We only provide a brief description here (see Supplementary Material for more details). OR1G1 or the mutated receptors were stably co-expressed with $G_{\alpha 16}$ protein in HEK293 cells (human embryo kidney cells) as previously described. Single mutations (A112S and A112V) were introduced into the OR1G1 sequence according to the instructions of the QuickChange Site-directed Mutagenesis Kit (Stratagene). Confocal immunofluorescence microscopy was used to check that OR expression at the cell surface was similar for the wild-type and mutant receptors. Functionality of receptors was measured by calcium imaging as previously published, but we applied odorants to cells during 2 min at 50 μl/min in such a way that no air was present above the solution covering the cells (perfused top engineered by BRACER). Odorants were purchased from Sigma-Aldrich, Fluka or Acros Organics at the highest purity available. They were freshly prepared by a first dilution at 100 mM into dimethyl sulfoxide and serial dilutions into Hanks' salt solution (Eurobio) supplemented with 20 mM Hepes, pH 7.2. Calcium responses were expressed as $R_{Ca^{2+}}(F) = (F - F_0)/F_0$, where F is the fluorescence intensity at each point and F_0 the value of emitted fluorescence before the stimulus application. Cell responses to odorants were measured as the mean of $R_{Ca^{2+}}(F)$ values over 100 cells.

Results

Evaluation of the homology modeling procedure

We carried out a cross-comparison of the six receptor 3D structures experimentally determined (these structures correspond to the inactive form of the receptors, see Supplementary Table ST1). The upper triangular part of Supplementary Table ST2 displays the rmsd of the Cαs (alpha carbons of the residues) after optimal superimposition of the seven TMHs using VAST (Madej *et al*., 1995; Gibrat *et al*., 1996). The lower triangular part displays the sequence identity in the seven TMHs after the structural alignment. Figure 1 presents the boundaries and sequences of the helical segments that were aligned in the six templates.

We then aligned, in turn, the sequence of each of the six receptors of known structure with FROST, using the 3D structure of the remaining five receptors as templates. From the resulting alignment, we built 10 models for each query sequence–template structure pair with MODELLER. Table I shows the minimum and maximum rmsd obtained when comparing the 10 models with the corresponding X-ray structure. The rmsd is computed on the set of Cαs used in Supplementary Table ST2. In all cases, FROST provided the same alignment as the one obtained with VAST (see Fig. 1). Rmsds between the models and the X-ray structure are similar to those of their template structures with the latter.

MODELLER provides a 3D-conformation for the loops joining the TMHs. As there is no structural template available for the loops, the software has to resort to *de novo* modeling. It is usually very difficult to accurately model loops

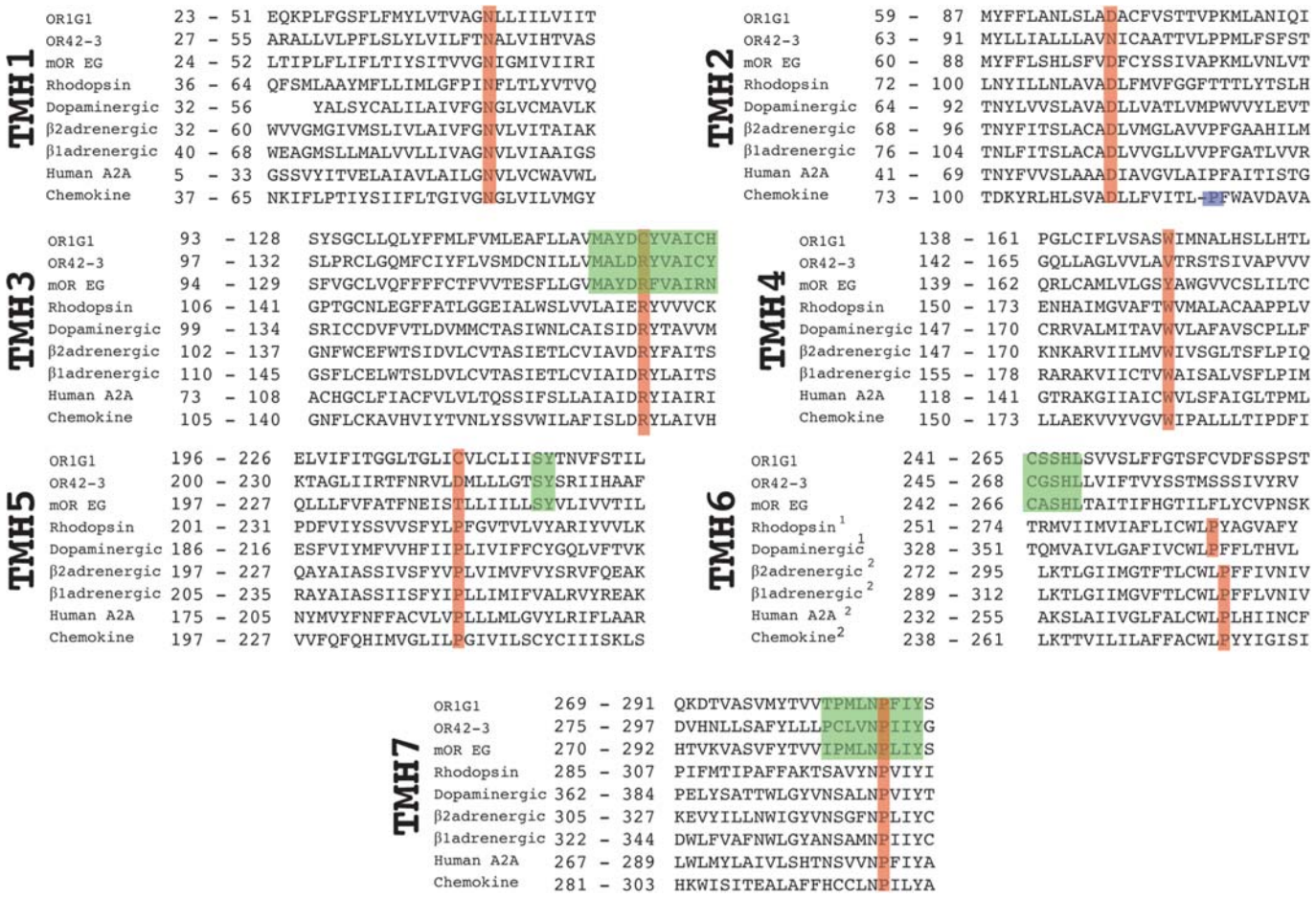

Fig. 1. GPCR templates and olfactory receptor multiple sequence alignment. Starting and ending positions of TMHs are indicated. Green boxes highlight conserved positions previously reported in the literature. Position 50 of the Ballesteros–Weinstein numbering scheme is shown in orange. Pro II.58 is shown in blue. 1, 'C-start' alignment; 2, 'x-start'.

Table I. Structural comparisons between modeled and experimental receptors

Queries (sequences)	Targets (3D templates)					
	2rh1A	1u19A	2vt4A	3emlA	3oduA	3pblA
2rh1A		2.17–2.47	0.80–0.91	2.14–2.37	2.17–2.42	1.51–1.82
1u19A	2.04–2.32		2.11–2.36	2.50–2.84	2.14–2.50	1.75–2.19
2vt4A	1.17–1.28	2.26–2.45		1.96–2.18	2.44–2.67	1.68–1.96
3emlA	2.08–2.27	2.36–2.76	1.95–2.24		2.68–3.07	2.25–2.61
3oduA	2.42–2.72	2.16–2.35	2.29–2.53	2.60–3.10		1.94–2.48
3pblA	1.38–1.54	1.68–1.96	1.56–1.80	2.22–2.70	1.82–2.18	

Modeled and experimental receptor structures are compared using Cα rmsd. Positions taken into account for this calculation are identical to those in Table II.

that have more than 12 residues (Gibrat *et al.*, 1992). Indeed, the results provided by MODELLER for the loops of the receptors were not very convincing (data not shown). This is true, in particular, for the second extra-cellular loop (ECL2) joining TMH4 to TMH5. This long loop acts as a sort of lid on top of the binding pocket and some of its residues are involved in interactions with the antagonists. In the following, we omit the extra-cellular loops in the ligand docking simulations since we found that they were more a nuisance than a help. The omission of ECL2 certainly affects the docking of ligands that interact with it, principally the antagonists but also some of the largest agonists (see below).

Despite the low-sequence identity observed between some TMHs in different receptors (sometimes, down to 10%), the models provided by our procedure were all reasonably confirmed by the structure alignments.

Evaluation of the ligand docking procedure using experimental complexes

We then addressed the evaluation of the ligand-docking step. We used as test set all the GPCR–ligand complexes with non-covalent binding mode, i.e. we omitted the rhodopsin/retinal couple. We also discarded the β1-adrenergic receptor–cyanopindolol complex, because cyanopindolol and

Table II. Docking results for known ligand-receptors

PDB code	Complex[a]	Ligand type	Predicted binding mode		No. of contacts with ECL2
			I_{rmsd}	M_{dist}	
2rh1	b2AR/carazolol	Inverse agonist	1.20	0.84	4
2y04	b1AR/salbutamol	Partial agonist	1.66	0.99	0
2y01	b1AR/dobutamine	Partial agonist	2.89	1.84	0
2y03	b1AR/isoprenaline	Agonist	3.20	1.88	0
2y02	b1AR/carmoterol	Agonist	4.95	3.78	4
3pbl	D3DR/eticlopride	Antagonist	3.59	1.44	4
2ydv	A2AR/NECA	Agonist	6.52	3.66	14
3eml	A2AR/ZM241385	Antagonist	9.95	5.71	13
3odu	CXCR/it1t	Antagonist	8.81	3.35	16

The predicted docking mode is the one with the smallest colony energy (columns 4–5). The last column indicates the number of atomic contacts between the ligand and the second extra-cellular loop. A contact corresponds to a distance <3.6 Å between an atom of the ligand and an atom of ECL2.
[a]See Supplementary Table ST1 for receptor acronyms.

carazolol are close analogs and the β1 and β2-adrenergic receptors are also very similar. The list of ligand–receptor complexes that were used is shown in Supplementary Table ST1. References for the A_{2A} adenosine receptor–NECA complex and for the complexes of β1-adrenergic receptor with dobutamine, salbutamol, isoprenaline and carmoterol are, respectively, Lebon *et al*. (2011) and Warne *et al*. (2011).

The results of the nine trials are summarized in Table II, where the predicted binding modes (the docking conformations of lowest colony energy) are characterized by their I_{rmsd} and their center of mass distance (M_{dist}) with the experimental ligand conformation. The colony energy weights the Xplor energy by the number of ligand conformations that establish the same types of interaction with the receptor. Thus, it tends to favor ligand conformations that are clustered in the binding pocket (see the energy–I_{rmsd} scatter plots on Supplementary Fig. SF1 and SF2). As illustrated on these figures, the use of the colony energy scheme does result in a better prediction of the ligand-docking conformation. A small value of M_{dist} indicates that the predicted ligand position is close to the experimentally observed one in the binding pocket. A small value of I_{rmsd} shows that the predicted and experimental positions have, in addition, the same orientation in the binding pocket. Supplementary Figs SF3 and SF4 show the docking position of the nine ligands in the corresponding receptors, both for the experimental (red) and predicted (green) conformations.

The exclusion of ECL2 may have two undesirable effects on the docking of the ligands. First, some of them, e.g. NECA, ZM241385, it1t, do form extensive interactions with ECL2 that cannot be correctly modeled (Table II). For instance, ligands of the A_{2A} adenosine receptor (NECA, ZM241385) are stabilized by an aromatic stacking interaction with Phe168 in ECL2. Second, the absence of ECL2 creates an artificial cavity where the ligand can dock without steric hindrance. Thus, the predicted ligand conformation (lowest colony energy) for the D3 dopaminergic receptor–eticlopride complex would actually result in a severe steric clash with Ile183 of ECL2. Overall, it appears that the absence of ECL2 can be especially detrimental for ligands establishing many contacts with ECL2 as, as shown in Table II, we cannot position the center of mass of these ligands within 3 Å of their true location, not to mention their orientation.

Results of the docking of the different ligands with the adreno-receptors (β1 and β2) are more satisfactory than those for the A_{2A} adenosine receptor (with the exception of carmoterol that makes contacts with ECL2). The docking of ligands on models of the adreno-receptors is certainly a favorable case for two reasons. First, the interaction of the ligands with the receptors involves no, or few, atomic contacts with residues located in ECL2. Second, and we believe more importantly, some of the models are very accurate (0.8–0.9 Å, see Table I) since when modeling an adreno-receptor we use the other adreno-receptor among the templates. This makes the docking procedure more likely to succeed. In less favorable cases, such as the modeling of ORs described below, we expect models to be in the range of 2–3 Å rmsd with respect to the genuine structure.

Thus, quite obviously, it appears difficult to predict accurately ligand-binding conformations when the complexes involve extensive interactions with residues belonging to extra-cellular loops that are not present in the model. As will be discussed hereafter, we believe this type of interactions to be mostly characteristic of the binding of antagonists and some of the largest agonists (see Fig. 2c). In contrast, when the ligand–receptor interactions mainly involve residues in the TMHs, the methodology we propose provides satisfactory results.

Application to olfactory receptors

To verify the applicability of the above modeling/docking procedure to ORs and further evaluate it, we carried out the modeling of human OR1G1. For this OR, functional characterization (Sanz *et al*., 2005) and 3D-QSAR analyses (Sanz *et al*., 2008) have already been performed by one of us.

Homology modeling of olfactory receptors. We carried out the modeling of OR1G1 using the six templates mentioned in Methods. Alignments of the OR1G1 sequence with the structural templates are presented in Fig. 1 (Supplementary Table ST3 displays the sequence identity of the TMHs for these alignments). TMH alignments are consistent except the one for TMH6. For this TMH, two different alignments exist, one for the rhodopsin and dopaminergic templates and another one, shifted by one position, for the four remaining templates. TMH6 in ORs is characterized by the 'TCxSHL' motif, located at the beginning of the TMH. Thr is not

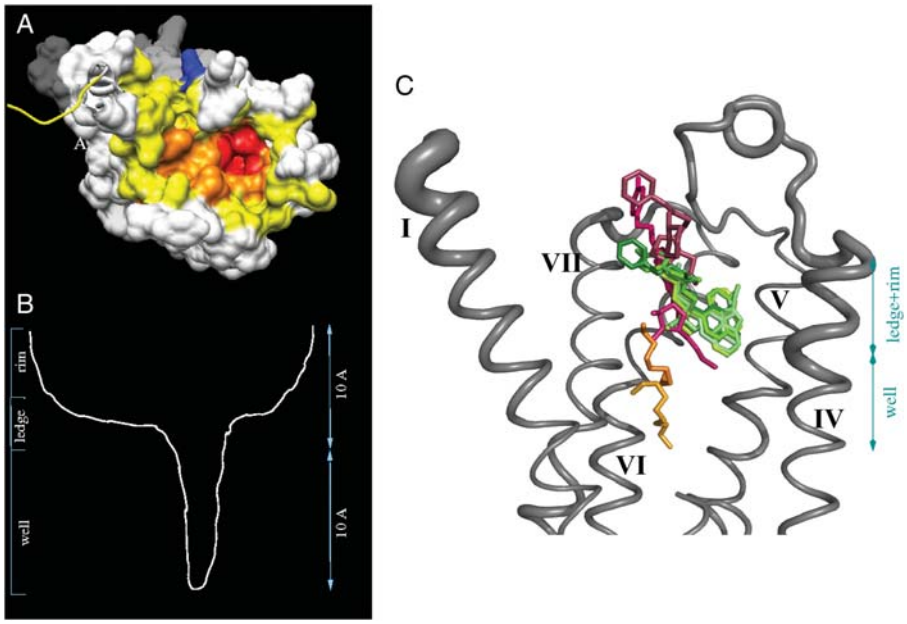

Fig. 2. OR1G1-binding pocket. Inset **A** represents the molecular surface of the OR1G1-binding pocket for the model built with 2rh1A as structural template. The binding pocket is viewed from the extra-cellular side. The extra-cellular loops have been removed for clarity. Inset **B** shows a schematic representation of the pocket (viewed laterally). Three zones are defined, the well, the ledge and the rim shown, respectively, in red, orange and yellow in inset A. The well is ~10 Å deep. The cumulative depth of the rim and ledge zones is similar (10 Å). Inset **C** shows the binding mode of OR1G1 ligands (1-nonanol: light orange, octanal: darker orange), β1 and β 2-adrenergic receptors (carazolol, dobutamine, salbutamol, carmoterol, isoprenaline in different shades of green), CXCR4 chemokine ligand (it1i in dark pink) and A2A adenosine receptor ligand (ZM241385 in light pink). *Note*: D3 dopaminergic receptor ligand, eticlopride, docks at the same location as the ligands of the adreno-receptors and is not represented to avoid further cluttering the figure. Some residues of TMH2, TMH3 and ECL1 have been removed to allow a better view of the ligands.

shown on Fig. 1 and 'x' represents a moderately conserved residue (Ser in OR1G1, Gly in mOR42-3 and Ala in mOR-EG). The first aforementioned alignment corresponds to the alignment of the Cys in the motif and the second one to the alignment of the 'x' residue with the first position of the block representing TMH6 in the structural core. Hereafter, we will refer to these alignments, respectively, as the 'C-start' and 'x-start' alignments. Both models have very close threading scores. TMH6 is particularly difficult to align with the available templates as there is very little sequence conservation in this TMH between the olfactory and the template receptors. TMH6 is believed to play an important role in the signal transduction mechanism. The Trp of the 'CWxP' motif ('x' is a Leu for all templates in Fig. 1) has been described as a toggle switch (Nygaard *et al.*, 2009). Notice that this motif is not conserved in OR sequences.

To decide between the two above models, we carried out the modeling of mOR-EG (Katada *et al.*, 2005) and mOR42-3 (Abaffy *et al.*, 2007) for which previously published models had been manually crafted. Using these models and ligand docking studies to guide the experiments, the authors investigated the functionally important residues by site-directed mutagenesis and functional assays. Unfortunately, they did not explain how the alignments were obtained (other than 'the alignment was refined manually' and 'Sequence alignment [...] was manually improved') and they did not supply the resulting alignments. However, they provided a list of functionally important residues and figures showing the interaction between the ligands and these residues. For mOR-EG, residues Phe252, Thr255, Ile256 and Leu259 of TMH6 are interacting with the odorant and for mOR42-3 only Thr259 in TMH6 is involved. We used these pieces of information to find out which of the above two alignments, 'C-start' or 'x-start' was the most likely correct, if any. Figure 1 also displays the alignments of mOR-EG and mOR42-3 with the six templates. Alignments of the three ORs are consistent, and exhibit two alternative alignments for TMH6. In the 'x-start' alignment, the side chains of the residues interacting with the ligand point to the neighboring TMH. In contrast, for the models based on the 'C-start' alignment of TMH6, the side chains of these residues are positioned toward the 'lumen' of the binding pocket and are thus able to establish interactions with the ligand. We also checked that other residues involved in interactions with the odorants and found in other TMHs (TMH3 and TMH5 for mOR-EG and TMH3, TMH4 and TMH5 for mOR42-3) were also correctly oriented relative to the binding pocket. In conclusion, it appears that the 'C-start' alignment of TMH6 is the one that best explains experimental data.

Docking of odorants on the OR1G1 model. The modeling of the OR1G1 3D structure was performed, forcing the 'C-start' alignment for the A_2-adenosine, chemokine, and the two β-adrenergic templates. Figure 2A shows the molecular surface of the binding pockets for one of the OR1G1 models based on the structural template 2rh1A and Supplementary Table ST4 displays the 47 residues contributing to this molecular surface. To quantify the effect of the modeling procedure on the definition of the binding pocket, we determined the accessible surface area of these residues in different models, providing us with as many 47 element vectors. Using these vectors we computed the correlation coefficient. For the 10 models built from the same template, the correlation coefficient is 0.903. For models built with different templates, the correlation coefficient is 0.751. Thus, different models lead to reasonably similar binding pockets.

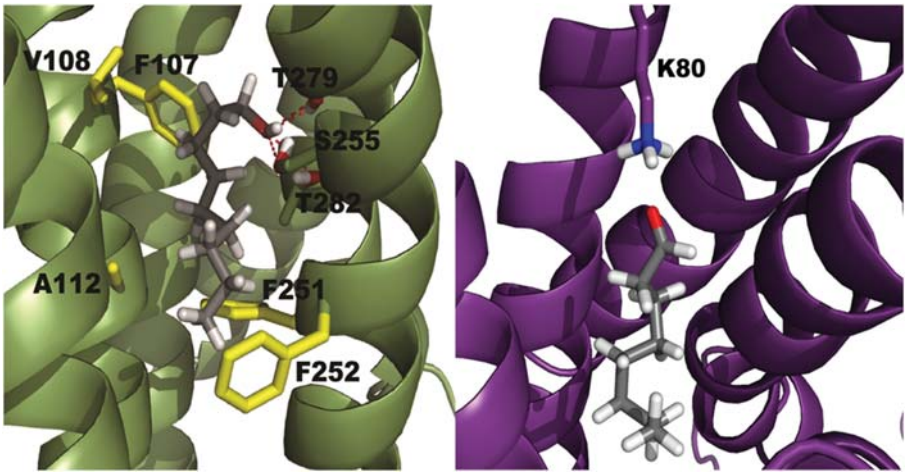

Fig. 3. (**A**) docking mode of 1-nonanol in OR1G1 (green helices). The network of hydrogen bonds is shown with broken lines. (**B**) docking mode of octanal in OR1G1 (magenta helices). Lys80 that interacts with the aldehyde function is shown on top of the figure (its nitrogen is shown in blue)

Using the multiple sequence alignment of OR1G1, we also computed the entropy of the different positions. We found 32 positions with an entropy >0.7 (we use a base 20 log thus the value of the entropy is between 0 and 1), eight of them (25%) belonging to the binding pocket.

Two zones can be identified, the lower part corresponds to a relatively narrow and deep 'well' (shown in red in Fig. 2A), and the upper part to a more open area that consists of a kind of 'ledge' around the well and the 'rim' of the binding pocket (shown in orange and yellow). Nygaard *et al.* (2009) described the binding pocket as 'a funnel with a partial lid', the lid consisting of the extra-cellular loops, in particular ECL2. They also noticed that antagonists, in all experimentally determined complexes, are located in the upper part of the binding pocket and establish contacts with residues in extra-cellular loops. In contrast, retinal forms direct interactions with the Trp of 'CWxP' motif in TMH6 located in the 'well'. This Trp is one of the identified micro-switches. Therefore, they proposed that 'a common mode of action for all antagonists known today would be that they prevent TMH6 from tilting inward in the main ligand-binding pocket'.

We carried out the docking of 1-nonanol, a full agonist of OR1G1, using the above OR1G1 models. Figure 2C shows the predicted binding mode in the full-length receptor. Figure 3A shows a zoom of the binding mode of 1-nonanol. The ligand is vertically positioned in the 'well'. It interacts with residues located in TMH3, TMH6, TMH7 and, to a lesser extent, TMH5. The aliphatic tail of the ligand establishes close contacts with the hydrophobic floor of the well, consisting of the aromatic rings of Phe251 and Phe252 and the methyl group of Ala112. The hydroxyl group of the ligand, at the opposite side, takes part in a hydrogen bond network with the backbone nitrogen of Thr279 and the hydroxyl groups of Ser255, Thr279 and Thr282 side chains. The ligand also makes hydrophobic contacts with residues of TMH3 (Phe107, Val108). This predicted binding mode is in good agreement with results obtained by a previous study of the OR1G1 odorant repertoire (Sanz *et al.*, 2008). This study performed a 3D-QSAR analysis of OR1G1 agonists. This led to two 'hypothesis' models, one of them ('A36') consisting of a set of hydrophobic, hydrophobic aliphatic and hydrogen bond donor features. These features and the distances between them can be approximately mapped onto the binding conformation of 1-nonanol (data not shown).

Table III. Response of OR1G1 wild type and A112S mutant to odorants

Odorants	Wild type $R_{Ca^{2+}}(F_{max})$	A112S $R_{Ca^{2+}}(F_{max})$
1-Nonanol	40	7↓
Nonanal	24	9↓
γ-Decalactone	31	13↓
Tridecanal	38	8↓
9-Decen-1-ol	30	7↓

Results are presented for odorant doses inducing the largest variation of $R_{Ca^{2+}}(F_{max})$ (mean for 100 cells, %) between mutants and OR1G1 (SD: between 5 and 10% of the values, three independent experiments). Arrows indicate a decrease of the mutant response compared with the wild-type one.

We performed a limited number of site-directed mutations with heterologous expression of the mutant and wild-type receptors and functional Ca^{2+} imaging assays. On the basis of the docking prediction, introducing a polar residue in the hydrophobic floor of the well should result in less favorable interactions with the hydrophobic tail of the ligand and to a concomitant decrease of the mutant receptor functional response. As shown in Table III (see Supplementary Fig. SF5), the A112S mutant results in a weaker response of the receptor for a variety of ligands having a long aliphatic tail, compared with the wild type.

Functional response analysis of OR1G1 ranks octanal as a weaker agonist than 1-nonanol (Sanz *et al.*, 2005). To study this difference, we docked octanal on the OR1G1 models. Figure 2C presents the predicted binding mode in the full-length receptor. Figure 3B presents a zoom of the binding mode. It differs significantly from the conformation observed for the 1-nonanol–OR1G1complex. The carbonyl oxygen of octanal forms a strong electrostatic interaction with Lys80. Singer (2000) proposed that the formation of a Schiff base was the mechanism underlying the activation of the rat rORI7 by aldehydes. He also mentioned the critical role of a nearby Asp, acting as a counter ion that would prevent the ligand 'from binding the receptor too avidly'. Indeed, a Schiff base results in a covalent double bond between the

nitrogen of a Lys side chain and the carbon of an aldehyde carboxyl (as observed with the retinal in rhodopsin). Such a strong bond would tether the ligand to the receptor in a more or less irreversible way. We believe it is much more likely that the interaction of rOR17 or OR1G1 with aldehydes involves a Lys via non-covalent electrostatic interactions between the positively charged nitrogen of the Lys side chain and the partial, negative, charge on the carboxyl oxygen than through a Schiff base formation.

According to the scoring scheme of GPCRautomodel, the octanal–OR1G1 complex is energetically favored over the 1-nonanol–OR1G1 complex. In other words, octanal binds tighter to OR1G1 than 1-nonanol does, yet it is a weaker agonist. A possible explanation for this behavior is that although the binding energy is important to explain the receptor activation, the position of the ligand in the binding pocket plays an even more critical role. For instance, an antagonist might bind to the receptor even better than some agonists, but, as discussed above, at a different location in the pocket. In the binding mode shown in Fig. 3, octanal, unlike 1-nonanol, does not come into contact with the bottom of the well (residue Ala112). The A112V mutant, by increasing the size of the side chain at this position, i.e. raising the floor of the well, results in a slightly enhanced response of the receptor to octanal (see Supplementary Fig. SF6). Likewise, as shown in Table III, tridecanal, with its longer aliphatic tail, is as good an agonist as 1-nonanol.

Discussion

One of the main sources of error regarding homology modeling that has been reported many times in CASP contests is due to inaccurate alignments between the query and the template sequences. Usually, it is impossible to recover from this type of error and the resulting model is irremediably wrong. GPCRs have low sequence similarity, even within families. Indeed, this is the case of the R family (rhodopsin-like) we consider in this work. In particular, in this family, members of the OR sub-family seem to be among the most divergent receptors. Automatically aligning R family receptors is thus quite challenging. In most published works, it has proven necessary to resort to human expertise to improve manually the initial sequence alignment. Note that in some markedly difficult cases, such as the one discussed above for TMH6, it is likely that most experts would fail to find the right alignment without clues from experimental data.

In this work, we developed a completely automated procedure for modeling OR 3D structures. The thorny step of this procedure, i.e. the alignment of the query sequence with the template sequences, relies on a fold recognition methodology. The results we obtained, both with available GPCR 3D structures and OR models validated with experimental data, indicate that our methodology is robust.

We used a standard homology modeling technique (MODELLER) to create OR 3D structure models from the threading alignment. With this method, it is difficult to accurately model long loops. Extra-cellular loops are important for the docking of antagonists, in particular ECL2. However, we noticed that including them in subsequent docking simulations caused more harm than benefit. Thus, we decided to omit extra-cellular loops from our models. Yet, Goldfeld *et al.* (2011) have recently published a paper showing that it is possible to predict accurately GPCR loops, including ECL2, starting from the X-ray structures. It remains to be seen whether the same accuracy can be reached starting from models of the receptors.

As often with homology modeling methods, the resulting model is closer, in terms of rmsd, to the template structure than to the actual 3D structure. Some programs, such as MEDELLER (Kelm *et al.*, 2010), are specifically dedicated to membrane protein modeling. Its authors claim that it provides better results than MODELLER in terms of rmsd of the 3D structure models with respect to the native structures. They obtained results of 2.1, 2.3 and 2.5 Å rmsd for three models of the 3D structure of the human A_2-adenosine receptor. Our results for this receptor were comparable to these values, with a minimum rmsd value of 1.95 Å (see Table I). To improve the quality of the models, Monte Carlo or Molecular Dynamics simulations could be performed permitting a thorough exploration of the local conformational space. However, we cannot afford to run such computer-intensive simulations on the server for the GPCRautomodel users.

For the sake of comparisons with existing modeling methods, we submitted the sequences of mOR-EG and mOR42-3 to the I-TASSER web site (Roy *et al.*, 2010). I-TASSER is a platform for modeling protein structures that has been ranked as the best method in recent CASP experiments. It integrates updated version of the programs that were used by Zhang *et al.* (2006) to model human GPCRs. I-TASSER provides five models and the alignments of the first (best) one with the closest proteins of the PDB, in terms of 3D structures, as well as other pieces of information that do not concern us here. For mOR42-3, the list of closest proteins includes, in this order, the A_{2A}-adenosine receptor, the β2 and β1-adrenergic receptors, the squid and bovine rhodopsins. For mOR-EG, the corresponding list comprises the A_{2A}-adenosine receptor, the β2-adrenergic receptor and the bovine rhodopsin. We checked the positions of the TMHs in these alignments. In brief, alignments of all TMHs except TMH6 are identical to the ones we obtained. For mOR-EG the alignment of TMH6 shows a shift of one residue (corresponding to our 'x-start' alignment) and for mOR42-3 the alignment of TMH6 exhibit a shift of six or four residues, depending on the template. We then mapped the residues that have been found by experimentalists as being critical for the function of these two ORs. For mOR-EG, except for residue Ile256, all other functionally important residues of TMH6 (Phe252, Thr255, Leu259) are either facing another TMH or the membrane, indicating a problem with the alignment provided by I-TASSER. For mOR42-3, there is only one functionally important residue located in TMH6 (Thr259) and as the alignment is shifted by four positions, i.e. one helix turn, it also points to the lumen of the binding pocket, albeit higher compared with our model. It is thus difficult to decide, based on this evidence alone, which TMH6 alignment is the best for this OR. However, we note that in our models, the conserved motif 'CxSHL', at the beginning of TMH6 in ORs, is aligned in a consistent manner for the three ORs we modeled. This is not the case for I-TASSER models.

We performed, with VINA, the docking of ligands on the previously obtained receptor models. We introduced a new way for computing the docking energy of the ligand with the

receptor, based on the concept of statistical rescoring (colony energy). A similar scheme has previously been used to score loop conformations (Xiang *et al.*, 2002) and in the context of ligand docking (Lee and Seok, 2008). As we consider a large number of receptor models that differ significantly in terms of rmsd, we cannot trust the I_{rmsd} to cluster ligand conformations. I_{rmsd} is too sensitive to small variations in the ligand conformation and receptor 3D structure. As described in Methods, our scheme favors ligand conformations that have the same interaction 'fingerprint' with the receptor (based on the receptor–ligand contact map overlap). Loosely speaking, this statistical rescoring scheme allows us to take some sort of entropic effect into account. We tested this scoring scheme on available experimental complexes. Unfortunately, a number of these complexes involve antagonists that have extensive interactions with extra-cellular loops, which we chose to omit in our docking simulations. Therefore, the modeling of such complexes is beyond the capacity of our method. We succeeded in modeling the carazolol–β2-adrenergic receptor and salbutamol–β1-adrenergic receptor complexes very accurately. We were also able to model reasonably well the dobutamine–β1-adrenergic receptor and isoprenaline–β1-adrenergic receptor complexes. As the β2-adrenergic receptor and the β1-adrenergic receptor are close homolog sharing >70% sequence identity within the TMHs, we obtained very good models (in the range 0.8–0.9 Å rmsd, see Table I) using one as the template and the other as the query. This docking experiment shows that our protocol is capable of locating the ligand precisely in the receptor 3D structure, when the receptor model is accurate and the ligand does not interact too closely with the extra-cellular loops. More generally, when the query sequence shares ~20% sequence identity with the structural template sequences, we expect the models to be in the range of 2–3 Å rmsd. It is the case for ORs and the results we obtained were consistent with our experiments, which is quite encouraging. Our protocol for docking ligands on the receptor 3D models provides only a 'static' view of the process and a relatively coarse estimate of the corresponding energy terms. Calculations based on free energy perturbation methods (see for instance Kurland *et al.*, 2010) might prove more effective to carry out a more comprehensive analysis of the energetics of ligand docking.

The binding free energy of a ligand to a GPCR is of course an important parameter for receptor activation. However, ligand binding is not a sufficient condition for receptor activation. Indeed, antagonists bind tightly to the receptor too, but they do not trigger the signal transduction. A crucial parameter consists of the ligand-docking position within the binding pocket. Nygaard *et al.* (2009) have proposed that antagonists dock in the upper part of the binding pocket (corresponding to the ledge and rim in our definition of the pocket). Odorants are often low molecular weight organic compounds. We suggest that OR agonists, which are thus relatively small compared with other GPCR ligands, bind in the narrow lower part of the pocket (the well). Our results with OR1G1 tend to show that the efficacy of the ligand in activating the receptor depends on its capacity to interact with the bottom of the well. The most potent ligands are able to establish tight van der Waals contacts with side chains of residues forming the well floor. It is interesting to notice that, although all OR agonists appear to bind within the well, they do so in very different ways according to their functional groups. For OR1G1, the aliphatic tail of alcohol ligands interacts with the floor of the well while the hydroxyl group participates in a hydrogen bond network involving Ser255, Thr279 and Thr282. For aldehyde ligands, the carbonyl group forms a strong electrostatic interaction with Lys80 while their aliphatic tail, if long enough, also interacts with the well floor. Although using completely different ligands (a benzene ring substituted at three different positions with various functional groups), Katada *et al.* (2005) defined a very similar binding site in the well. Supplementary Table ST5 shows the 10 residues that they predicted to interact with eugenol (mOR-EG was isolated from a eugenol-responsive olfactory neuron) and the corresponding residues in OR1G1 according to the alignment shown in Fig. 1. Residues that we found interacting with 1-nonanol are indicated in bold. Ala112 in OR1G1 corresponds to the bottom of the well. It is predicted to interact with the aliphatic tail of the ligands. Indeed, its substitution by a Ser results in a decrease of the receptor response to all strong agonist tested (see Table III). Incidentally, in mOR-EG, this position in the well corresponds to Ser113, but Ser113 is predicted to interact with the hydroxyl group at position R_1 of eugenol. Conversely, residue Ser255 in OR1G1, which participates in the hydrogen bond network also involving the hydroxyl group of 1-nonanol, corresponds to Ile256 in mOR-EG. Ile256 is involved in van der Waals interactions with the ligand and 'is responsible for defining the spatial configuration of the binding pocket' according to Katada *et al.* (2005) Recently, Baud *et al.* (2011) showed that mOR-EG is capable of being activated by a variety of ligands and that the residues involved in the interaction receptor–ligand differ with the type of ligand.

Conclusions

To further dissect the mechanisms resulting in signal transduction, it would be interesting, in the future, to go beyond the approach we presented here. From the theoretical standpoint, using more powerful (but more costly) simulation techniques such as those based on free-energy perturbation methodologies could provide a more dynamical view of the ligand-docking process and activation mechanisms. In the absence of experimental structures, molecular modeling and computer simulations can suggest a number of hypotheses to be validated experimentally, for instance by mutagenesis, heterologous expression of receptors and functional assays. This interplay between simulations and experimental validations is crucial to progress in our understanding of the signal transduction by ORs or, more generally, GPCRs.

The method developed in this work, GPCRautomodel, should constitute a useful starting point for this endeavor. Its main purpose was to automate the construction of GPCR 3D structure models, based on all the structural templates presently available, and to obviate the need for a human expert to manually curate the initial alignment, thus allowing for high throughput, reproducible modeling of GPCR structures. The results presented in this paper indicate that our procedure, in this regard, is robust and accurate. In addition, the web site allows the user to dock ligands (chosen from a predefined library or user-defined) with the generated models. The colony energy scheme we proposed to rank ligand–

receptor complexes provides encouraging results when the ligands do not interact too intimately with extra-cellular loops, in particular ECL2. Thus, GPCRautomodel allows biologists to explore, *in silico*, the very large repertoire of functional fingerprints that permit ORs to recognize a seemingly infinite variety of chemical structures. Hopefully, thorough investigations of the complex models coupled with experimental validations will allow them to decipher the biochemical mechanisms used by evolution to tune the mammalian olfactory repertoire.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

The authors are grateful to the INRA MIGALE bioinformatics platform (http://migale.jouy.inra.fr) for providing computational resources, and the MIMA2 platform for the use of confocal microscopes. They thank Anne Tromelin for explanations regarding 3D-QSAR methodology and Pascal Bento for his help in designing the GPCRautomodel web site.

Funding

G.L. was financially supported by the 'Nanoscience, nanotechnologies, materials and new production technologies' program under the project BOND from the 7th Research Framework Programme of the European Union (European project code 228685-2). S.T. was financially supported by the French Institut National de la Recherche Agronomique (INRA). Developments performed on FROST have been supported by the ANR-06-CIS6-008 Grant.

References

Abaffy,T., Malhotra,A. and Luetje,C. (2007) *J. Biol. Chem.*, **282**, 1216–1224.
Andonov,R., Collet,G., Gibrat,J.F., Marin,A., Poirriez,V. and Yanev,N. (2008) In Talbi,E.-G. and Zomaya,A. (eds), *Grids for bioinformatics and computational biology*, Chapter 14. Wiley-Interscience, Hoboken, New Jersey, pp. 325–56.
Baud,O., Etter,S., Spreafico,M., Bordoli,L., Schwede,T., Vogel,H. and Pick,H. (2011) *Biochemistry*, **50**, 843–853.
Brunger,A.T. (1992) *XPLOR 3.1: A system for X-ray crystallography and NMR*. Yale University Press, New Haven, CT.
Cherezov,V., Rosenbaum,D., Hanson,M., *et al*. (2007) *Science*, **318**, 1258–1265.
Chien,E., Liu,W., Zhao,Q., *et al*. (2010) *Science*, **330**, 1091–1095.
Collet,G., Andonov,R., Yanev,N. and Gibrat,J.-F. (2010) *Discrete Appl. Math.*, **159**, 1707–1716.
Evers,A. and Klabunde,T. (2005) *J. Med. Chem.*, **48**, 1088–1097.
Freddolino,P., Kalani,M., Vaidehi,N., Floriano,W., Hall,S., Trabanino,R., Kam,V.W. and Goddard,W.A., III (2004) *Proc. Natl. Acad. Sci. U S A*, **101**, 2736–2741.
Fredriksson,R., Lagerstrm,M., Lundin,L. and Schith,H. (2003) *Mol. Pharmacol.*, **63**, 1256–1272.
Gibrat,J.-F., Higo,J., Collura,V. and Garnier,J. (1992) *Immunomethods*, **1**, 107–125.
Gibrat,J.-F., Madej,T. and Bryant,S. (1996) *Curr. Opin. Struct. Biol.*, **6**, 377–385.
Gloriam,D.E., Fredriksson,R. and Schioth,H.B. (2007) *BMC Genom.*, **8**, 338.
Goldfeld,D.A., Zhu,K., Beuming,T. and Friesner,R.A. (2011) *Proc. Natl. Acad. Sci. U S A*, **108**, 8275–8280.
Jaakola,V., Griffith,M., Hanson,M., Cherezov,V., Chien,E., Lane,J., Ijzerman,A. and Stevens,R. (2008) *Science*, **322**, 1211–1217.
Katada,S., Hirokawa,T., Oka,Y., Suwa,M. and Touhara,K. (2005) *J. Neurosci.*, **25**, 1806–1815.
Kelm,S., Shi,J. and Deane,C. (2010) *Bioinformatics*, **26**, 2833–2840.
Kurland,M., Newcomer,M., Peterlin,Z., Ryan,K., Firestein,S. and Batista,V. (2010) *Biochemistry*, **49**, 6302–6304.
Launay,G. and Simonson,T. (2011) *J. Comput. Chem.*, **32**, 106–120.
Lebon,G., Warne,T., Edwards,P.C., Bennett,K., Langmead,C.J., Leslie,A.G. and Tate,C.G. (2011) *Nature*, **474**, 521–525.
Lee,J. and Seok,C. (2008) *Proteins*, **70**, 1074–1083.
Lensink,M., Mendez,R. and Wodak,S. (2007) *Proteins*, **69**, 704–718.
Madej,T., Gibrat,J.-F. and Bryant,S. (1995) *Proteins*, **23**, 356–369.
Marin,A., Pothier,J., Zimmermann,K. and Gibrat,J.-F. (2002) *Proteins*, **49**, 493–509.
Murakami,M. and Kouyama,T. (2008) *Nature*, **453**, 363–367.
Nygaard,R., Frimurer,T., Holst,B., Rosenkilde,M. and Schwartz,T. (2009) *Trends Pharmacol. Sci.*, **30**, 249–259.
Palczewski,K., Kumasaka,T., Hori,T., *et al*. (2000) *Science*, **289**, 739–745.
Roy,A., Kucukural,A. and Zhang,Y. (2010) *Nat. Protocol*, **5**, 725–738.
Sali,A. and Blundell,T. (1993) *J. Mol. Biol.*, **234**, 779–815.
Sanz,G., Schlegel,C., Pernollet,J. and Briand,L. (2005) *Chem. Senses*, **30**, 69–80.
Sanz,G., Thomas-Danguin,T., Hamdani el,H., Le Poupon,C., Briand,L., Pernollet,J., Guichard,E. and Tromelin,A. (2008) *Chem. Senses*, **33**, 639–653.
Schlyer,S. and Horuk,R. (2006) *Drug Discov. Today*, **11**, 481–493.
Schüttelkopf,A. and van Aalten,D. (2004) *Acta Crystallogr. D*, **60**, 1355–1363.
Singer,M. (2000) *Chem. Senses*, **25**, 155–166.
Trott,O. and Olson,A. (2010) *J. Comput. Chem.*, **31**, 455–461.
Tusnady,G., Dosztanyi,Z. and Simon,I. (2005) *Nucleic Acids Res., Database Issue*, **33**, D275–D278.
Warne,T., Serrano-Vega,M., Baker,J., Moukhamet-Zianov,R., Edwards,P., Henderson,R., Leslie,A., Tate,C. and Schertler,G. (2008) *Nature*, **454**, 486–491.
Warne,T., Moukhametzianov,R., Baker,J.G., Nehmé,R., Edwards,P.C., Leslie,A.G., Schertler,G.F. and Tate,C.G. (2011) *Nature*, **469**, 241–244.
Wu,B., Chien,E., Mol,C., *et al*. (2010) *Science*, **330**, 1066–1071.
Xiang,Z., Soto,C. and Honig,B. (2002) *Proc. Natl. Acad. Sci. U S A*, **99**, 7432–7437.
Zhang,Y., Devries,M. and Skolnick,J. (2006) *PLoS Comput Biol.*, **2**, e13.