

# Predicted 3D structures of olfactory receptors with details of odorant binding to OR1G1

Soo-Kyung Kim · William A. Goddard III

Received: 20 March 2014 / Accepted: 2 September 2014 / Published online: 16 September 2014  
© Springer International Publishing Switzerland 2014

**Abstract** Olfactory receptors (ORs) are responsible for mediating the sense of smell; they allow humans to recognize an enormous number of odors but the connection between binding and perception is not known. We predict the ensemble of low energy structures for the human OR1G1 (hOR1G1) and also for six other diverse ORs, using the G protein-coupled receptor Ensemble of Structures in Membrane BiLayer Environment complete sampling method that samples 13 trillion different rotations and tilts using four different templates to predict the 24 structures likely to be important in binding and activation. Our predicted most stable structures of hOR1G1 have a salt-bridge between the conserved D3.49 and K6.30 in the D(E)RY region, that we expect to be associated with an inactive form. The hOR1G1 structure also has specific interaction in transmembrane domains (TMD) 3–6 (E3.39 and H6.40), which is likely an important conformational feature for all hORs because of the ~94 to 98 % conservation among all hOR sequences. Of the five ligands studied (nonanal, 9-decen-1-ol, 1-nonal, camphor, and *n*-butanal), we find that the 4 expected to bind lead to similar binding energies with nonanol the strongest.

**Keywords** Olfactory receptors · Protein structure prediction · Modeling · G protein-coupled receptor

## Introduction

The human olfactory system involves ~396 olfactory receptors (ORs) that are responsible for mediating the sense of smell [1]. These ORs provide a unique connection between fundamental signaling at the molecular level and macroscale issues of perception. It is accepted that the perception of many thousands of odors using only a few hundred ORs results from combinatorial coding, in which each OR recognizes multiple odorants and in which each odorant is recognized by multiple ORs [2]. Thus interactions of odors with the ~396 human ORs, lead to complex combinations of activation (and perhaps inhibition) that are decoded by the brain to provide our perception of odors and our response. The human brain analyzes this pattern to recognize particular odorants. Functional NMR studies have shown that the brain responds to some odorants for which the person is not conscious. Although many studies detect only agonists (say by  $\text{Ca}^{2+}$  release), there is a growing evidence that some odorants act both as agonists or antagonists depending on the OR [3]. This dual agonist–antagonist combinatorial coding agrees with behavioral and psychophysical observations on perception of odorant mixtures [4]. The big question is how the patterns of activation of the ~396 ORs lead to particular perceptions. This mapping can be done experimentally, however the task of isolating and studying so many structures is enormous. A major impediment to gaining an atomistic understanding of OR activation and perception is that there are no 3D X-ray structures available for any OR and very little known about which ORs are activated by any particular odor.

---

**Electronic supplementary material** The online version of this article (doi:[10.1007/s10822-014-9793-4](https://doi.org/10.1007/s10822-014-9793-4)) contains supplementary material, which is available to authorized users.

---

S.-K. Kim (✉) · W. A. Goddard III (✉)  
Materials and Process Simulation Center (MC139-74),  
California Institute of Technology, 1200 E. California Blvd.,  
Pasadena, CA 91125, USA  
e-mail: skkim@wag.caltech.edu

W. A. Goddard III  
e-mail: wag@wag.caltech.edu

Our alternative approach is to use new methods of G protein-coupled receptor (GPCR) structure prediction and of ligand docking to predict an atomically accurate structure of each of the ~396 ORs and to determine which odorants bind most strongly to each one. This paper is our first step in this along this pathway. Here we focus on human olfactory receptor 1G1 with some information on several others, because there are available data on activation of OR1G1 by several odorants.

Accurate 3D structures for ORs would be quite useful for understanding the binding mechanism of various exogenous small molecules to the hOR family, which could be used to design odorants that might provide valuable targets for drugs for neurodegenerative disorders due to malfunction of ORs, such as anosmia (inability to detect odors) or hyposmia (decreased ability to detect odors) [5]. It might also be useful for developing new perfumes, making healthful foods more palatable and detecting unpleasant odors in the environment.

To provide accurate structural information we report here the ensemble of stable 3D structures predicted using the GPCR Ensemble of Structures in Membrane BiLayer Environment (GEnSeMBLE) Complete Sample Monte Carlo (CS-MC) method for the hOR1G1. This GEnSeMBLE CS-MC technique has been validated to predict structures with close to X-ray accuracy [6].

We chose the hOR1G1 as a representative in the OR family because it is regarded as prototypical of broadly tuned hORs [7]. It has been shown to have a large recognition spectrum [8], and structure-odor relationships indicate that OR1G1 binds odorants that do not correspond to the same olfactophores [9]. In addition, experimental data [8] is available to validate the predictions for several mutants and the functional response to various odorant molecules. Recently homology modeling and limited applications of our methods have been applied to hOR1G1 to identify the residues controlling odorant recognition [1, 7].

Here we report predicted 3D structures for hOR1G1 and find specific interaction that might provide important conformational constraints for all hORs because of the high sequence conservation among hORs.

## Results and discussions

### GEnSeMBLE predictions of apo-protein structures for hOR1G1

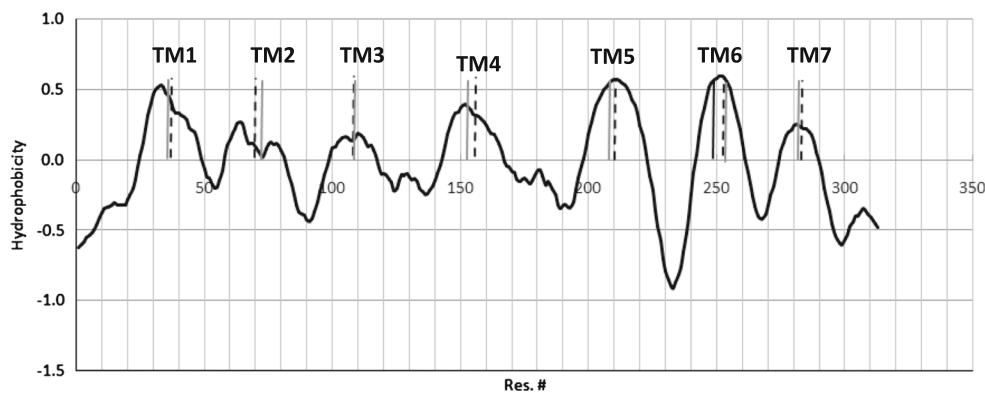
To determine the 7 transmembrane domains (TMD) for hOR1G1 we used PredicTM (see “Methods”), which does a multiple sequence alignment of all 398 hORs out of 1,462 GPCRs from BLAST searches in 2012. PredicTM uses the thermodynamic and biological hydrophobic scales from

White and von Heijne [10, 11] to determine the hydrophobic profile in Fig. 1 which allows the TMDs to be identified readily. Table 1 shows the sequence identities of the hOR1G1 with 13 currently available X-ray structures of GPCRs; In addition, in Table S1 of the Supporting Information we show the sequence identities for the most accurate six previously predicted structures: Based on their high sequence identities and diversities among available templates, we selected four templates for application of GEnSeMBLE having 17.6–22.7 % sequence identities in the TMDs: t $\beta$ 1AR, hS1P1R, hOPRK, and bRho. For families in which more than one GPCR is known, we chose the one with the highest sequence identity. Figure 2 shows the final TM regions and multiple alignments of hORs with the X-ray structures of four templates. All TM regions from PredicTM agree within 1–5 residues at the terminal end except TM4.

Figure 3 displays the multiple sequence alignment of the 10 hORs which sequences are the closest to that of hOR1G1. From the alignment of the family head of seven clusters for 398 hORs out of 1,462 GPCRs using MAFFT program (Fig. S1) we find:

TM7	the conserved residues in NPxxY for which the P is 7.50.
TM1, TM2, TM4	the conserved residues in N at 1.50, D at 2.50, and W at 4.50.
TM3	C for 3.50, which corresponds to R in the normal DRY motif.
TM5	C for 5.50, which corresponds to P in most class A GPCRs. Indeed Y5.58 is also highly conserved, which is consistent with C5.50.
TM6	T for 6.50, which corresponds to P in most other class A GPCRs. Proline shows 29 % conservation at T6.50 in all hORs. In addition, the conserved motif among ORs is FSTCSSH. In our alignment, we consider T254 as 6.50 in hOR1G1, while in the GPCR data base (DB) ( <a href="http://www.gpcr.org/7tm">www.gpcr.org/7tm</a> ) L250 was assigned as 6.50. However if L250 is 6.50, there were no prolines in all hORs. We carried out calculations using both alignments for TM6 and conclude that T254 <sup>6.50</sup> is the correct choice.

The initial step of the GEnSeMBLE method is the Bi-Helix step, in which we sample all rotations ( $\eta$  angle) of each helix using 15° increment up to  $\pm 60^\circ$ . This leads to a total of (9)<sup>7</sup> ~4.7 million combinations for each template. The energies for all 4.7 million packings were estimated using the BiHelix mean field approximation in which each



**Fig. 1** The hydropathy prediction from PredicTM for the human olfactory receptor OR1G1. Here the predicted hydrophobic center (HPC) which is the middle residue in the raw transmembrane (TM) region (RawMid) is indicated with dotted lines. The HPC from the

X-ray structure of turkey  $\beta_1$  adrenergic receptor ( $t\beta_1$ AR) [24] is displayed in the grey line for  $T254 = 6.50$  alignment of TM6 (which we conclude is correct) and a black line for  $L200 = 6.50$  alignment for TM6

**Table 1** The sequence identities of human olfactory receptor OR1G1 (hOR1G1) and the currently available X-ray structures of G protein-coupled receptors (GPCRs), turkey  $\beta_1$  adrenergic receptor ( $t\beta_1$ AR) [24], human  $\beta_2$  adrenergic receptor (h $\beta_2$ AR) [48], kappa opioid receptor (hOPRK) [37], rat M<sub>3</sub> muscarinic receptor (rM<sub>3</sub>MR) [25], human nociceptin/orphanin FQ receptor (hNOP) [49], human H<sub>1</sub>

histamine receptor (hH<sub>1</sub>HR) [50], mouse mu opioid receptor (mOPRM) [51], human M<sub>2</sub> muscarinic receptor (hM<sub>2</sub>MR) [26], mouse delta opioid receptor (mOPRD) [52], human sphingosine 1-phosphate (hS1P1R) [36], human adenosine A<sub>2A</sub> receptor (hAA<sub>2A</sub>R) [27], bovine rhodopsin (bRho) [31–35], and human D<sub>3</sub> dopamine receptor (hD<sub>3</sub>DR) [28]

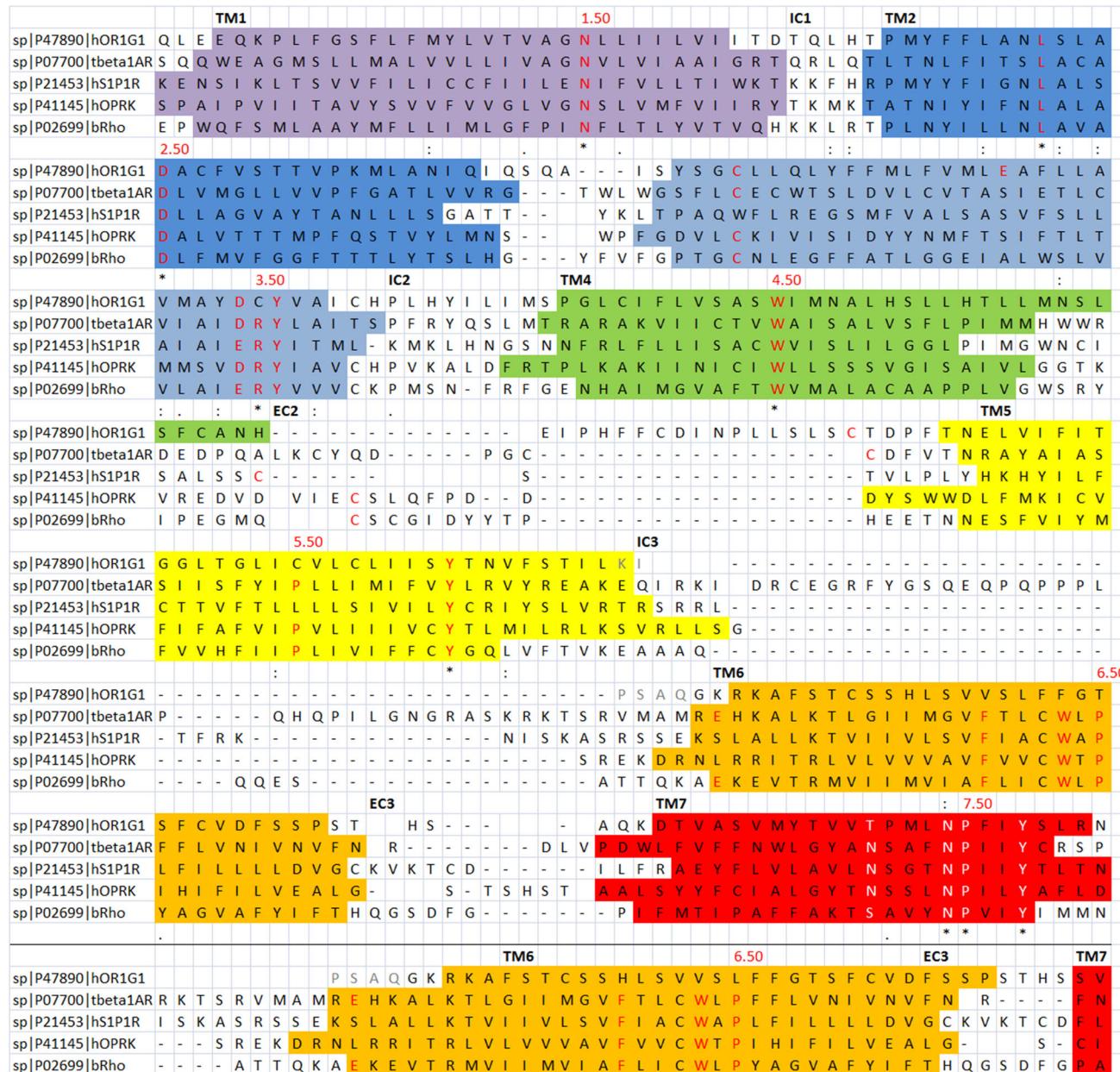
GPCR	All	TMs	TM1	TM2	TM3	TM4	TM5	TM6	TM7
hOR1G1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
<i>tβ<sub>1</sub> AR</i>	<i>18.5</i>	<i>22.7</i>	<i>33.3</i>	<i>21.4</i>	<i>26.9</i>	<i>14.7</i>	<i>15.6</i>	<i>25.0</i>	<i>21.7</i>
h $\beta_2$ AR	17.3	21.9	25.9	28.6	26.9	14.7	9.4	21.4	26.1
<i>hOPRK</i>	<i>15.7</i>	<i>16.5</i>	<i>22.2</i>	<i>28.6</i>	<i>19.2</i>	<i>5.9</i>	<i>18.8</i>	<i>3.6</i>	<i>17.4</i>
rM <sub>3</sub> MR	14.7	18.2	29.6	17.9	23.1	5.9	15.6	17.9	17.4
hNOP	14.7	17.0	25.9	28.6	19.2	5.9	9.4	3.6	26.1
hH <sub>1</sub> HR	14.7	18.0	33.3	17.9	19.2	5.9	9.4	14.3	26.1
mOPRM	14.4	17.1	22.2	35.7	19.2	8.8	12.5	3.6	17.4
hM <sub>2</sub> MR	14.4	17.0	29.6	17.9	19.2	8.8	15.6	10.7	17.4
mOPRD	14.4	15.5	18.5	35.7	23.1	2.9	6.3	0.0	21.7
<i>hS1P1R</i>	<i>14.1</i>	<i>19.1</i>	<i>14.8</i>	<i>32.1</i>	<i>15.4</i>	<i>14.7</i>	<i>12.5</i>	<i>17.9</i>	<i>26.1</i>
hAA <sub>2A</sub> R	14.1	16.9	14.8	17.9	30.8	8.8	12.5	7.1	26.1
<i>bRho</i>	<i>13.7</i>	<i>17.6</i>	<i>14.8</i>	<i>28.6</i>	<i>30.8</i>	<i>11.8</i>	<i>12.5</i>	<i>7.1</i>	<i>17.4</i>
hD <sub>3</sub> DR	13.7	15.0	22.2	14.3	11.5	8.8	3.1	14.3	30.4

The sequence was ordered by the sequence identity in all sequences. Each sequence identity of the transmembrane (TM) region was compared with hOR1G1. The selected structures for the homology templates are shown in italic

pair of helices has its side chains optimized independently with the interaction energies added. Then we select the best 2,000 structures from the BiHelix list to build into full 7-helix bundles, with optimized side chains for which we evaluate the energy of the full 7 TMD. For cases in which we start with a known X-ray structure as the template this essentially always predicts correctly the X-ray structure, indicating that the energy scoring is reliable.

We carried out BiHelix calculations for 14 cases [(1)  $t\beta_1$ AR-T-Hom-60, (2)  $t\beta_1$ AR-T-Hom-360, (3)  $t\beta_1$ AR-L-

Hom-60, (4)  $t\beta_1$ AR-L-Hom-360, (5)  $t\beta_1$ AR-T-Gen-60, (6)  $t\beta_1$ AR-T-Gen-360, (7)  $t\beta_1$ AR-L-Gen-60, (8)  $t\beta_1$ AR-T-Gen-360, (9) hS1P1R-T-Hom-60, (10) hS1P1R-L-Hom-60, (11) hOPRK-T-Hom-60, (12) hOPRK-L-Hom-60, (13) bRho-T-Hom-60, (14) bRho-L-Hom-60] based on the 4 X-ray templates, the two alignments of TM6 [6.50 (T or L)], both helix shapes (Hom vs. Gen) and both sampling ranges (fine sampling up to  $\pm 60^\circ$  vs. coarse sampling up to  $\pm 360^\circ$  for the best case of the  $t\beta_1$ AR). We selected the best 2,000 structures for evaluation of the full 7 TMD for each of the 14



**Fig. 2** The final transmembrane (TM) regions from the PredicTM method and multiple alignments of human olfactory 1G1 receptor (hOR1G1) with the X-ray structures for four templates. The predicted TM regions from PredicTM are displayed in colored boxes (TM1 in purple, TM2 in blue, TM3 in cyan, TM4 in green, TM5 in yellow, TM6 in orange, TM7 in red). The TM region from the L6.50

alignment in TM6 is displayed at the bottom. Highly conserved residues in family A G protein-coupled receptors (GPCRs) are shown in red in TM 1–6 and white in TM7. We use Ballesteros-Weinstein numbering consisting of the TM helix number followed by residue number relative to the highly conserved residue in the helix, numbered as 50

BiHelix runs (see “Methods”) and compared all 28,000 structures to select the top 20 shown in Table S2, based on neutral interhelical E between TMs. Here we see that the most stable and 16 others of the top 20 belong to the t $\beta$ 1AR homology template with the T254 = 6.50 alignment. The other three are from the t $\beta$ 1AR homology template with the L200 = 6.50 alignment. The most stable case (the best neutral interhelical energy) has eta angles of (0°, 15°, -15°,

-30°, 0°, 0°, 0° for TM 1–7). (A homology model would have 0 for all 7). We also see that the same structure is obtained using the alternative energy scoring in which all four ways to evaluate the energy (charge interhelical, charge total, neutral interhelical and neutral total).

Since the t $\beta$ 1AR template shows the best energy (E) in BiHelix, we also considered a coarse sampling of  $\eta$  angles up to ±180° by 30° increment, leading to a total of (12)<sup>7</sup>

	<b>TM3</b>	<b>3.25</b>													<b>3.50</b>
sp P47890 OR1G1	S Y S G C L L Q L Y F F M L F V M L E	A F L L A V M A Y D C Y V A													
sp O76100 OR7AA	T Y A G C I T Q M C F F L L F V G L D	N F L L T V M A Y D R F V A													
sp Q15622 OR7A5	T Y I A C L M Q M Y F F I L F A G F E	N F L L S V M A Y D R F V A													
sp O43749 OR1F1	S F C G C L T Q M Y F V F M F V D M D	N F L L A V M A Y D H F V A													
sp O14581 OR7AH	T Y A G C I T Q M C F F V L F G G L D	S L L L A V M A Y D R F V A													
sp Q8NHA8 OR1FC	S Y P G C L A Q M Y F C M M F A N M D	N F L L T V M A Y D R Y V A													
sp Q96RA2 OR7D2	S Y M D C L T Q V Y F S M F F P I L D	T L L L T V M A Y D R F V A													
sp Q8NG98 OR7D4	S Y M G C L T Q V Y F L M M F A G M D	T F L L A V M A Y D R F V A													
sp P30953 OR1E1	P Y A D C L T Q M Y F F L L F G D L E	S F L L V A M A Y D R Y V A													
sp Q15619 OR1C1	S F A G C L T Q L F F F V S F V N M D	S L L L C V M A Y D R Y V A													
sp Q8NGA1 OR1M1	S Y P C C L I Q M Y F F H F F G I V D	S V I I A M M A Y D R F V A													
	<b>TM6</b>														<b>6.50</b>
sp P47890 OR1G1	G K R K A F S T C S S H	L S V V S L F F G T S F C V D F S S P S T													
sp O76100 OR7AA	G K Y K A F S T C A S H	L S V V S L F Y G T C L G V Y L S S A A T													
sp Q15622 OR7A5	G K Y K A F S T C A S H	L S V V S L F Y G A I L G V Y L S S A A T													
sp O43749 OR1F1	G R W K A F S T C G S H	L A V V L L F Y S T I I A V Y F N P L S S													
sp O14581 OR7AH	G K Y K A F S T C A S H	L S V V S L F C C T G L G V Y L T S A A T													
sp Q8NHA8 OR1FC	G K L K A F S T C G S H	L A L V I L F Y G A I T G V Y M S P L S N													
sp Q96RA2 OR7D2	G K Q K A L S T C G S H	L S V V S L F Y G T G I G V H F T S A V T													
sp Q8NG98 OR7D4	G K Y K A F S T C G S H	L C V V S L F Y G T G L G V Y L S S A V T													
sp P30953 OR1E1	G I C K A F S T C G S H	L S V V S L F Y G T V I G L Y L C S S A N													
sp Q15619 OR1C1	G K Q R A V S T C S C H	L S V V V L F Y G T A I A V Y F S P S S P													
sp Q8NGA1 OR1M1	G R K K A F S T C S S H	L S V V V A L F Y G T T I G V Y L C P S S V													

**Fig. 3** Multiple alignments of human olfactory 1G1 receptor (hOR1G1) with 10 closest human olfactory receptors (hORs) to hOR1G1 out of 398 hORs from PredicTM in TM3 and TM6

~35 million combinations. This is to ensure that we do not miss some case with dramatically different angles. For all 35 million packings, the 7-TMD interaction is partitioned into 12 sets of pairwise interactions, which are added together (mean field approximation). For each of the 1,728 pairwise interactions we use SCREAM [12] to optimize the side-chains and minimized by 10 steps.

Starting with the best BiHelix structure ( $0^\circ$ ,  $15^\circ$ ,  $-15^\circ$ ,  $-30^\circ$ ,  $0^\circ$ ,  $0^\circ$ ) and referencing these angles now all as 0, we carried out the SuperBiHelix procedure to find the best combination of tilts and rotations. In the SuperBiHelix procedure we allow each helix to tilt (theta) with respect to the z axis (perpendicular to the membrane) by 0 or  $\pm 10^\circ$  while allowing its azimuthal angle (phi) and the rotation angle (eta) to vary up to  $\pm 30^\circ$  in increments of  $15^\circ$ . This leads to  $(3*7*5)^7 = 10$  trillion combinations, all of which are estimated using the BiHelix energies by combining 4 helices at a time as discussed in Methods.

The best 20 SuperBiHelix structures are listed in Table S3. The final best structure from SuperBiHelix has angles (for TMD 1–7) of  $10^\circ$ ,  $0^\circ$ ,  $-10^\circ$ ,  $10^\circ$ ,  $-10^\circ$ ,  $10^\circ$ ,  $10^\circ$  for theta,  $-30^\circ$ ,  $-30^\circ$ ,  $-30^\circ$ ,  $0^\circ$ ,  $0^\circ$ ,  $30^\circ$ ,  $15^\circ$  for phi, and  $0^\circ$ ,  $-15^\circ$ ,  $0^\circ$ ,  $0^\circ$ ,  $30^\circ$ ,  $0^\circ$ ,  $15^\circ$ . The starting structure (all 0 angles for theta, phi, and eta from BiHelix best) was ranked as number 6.

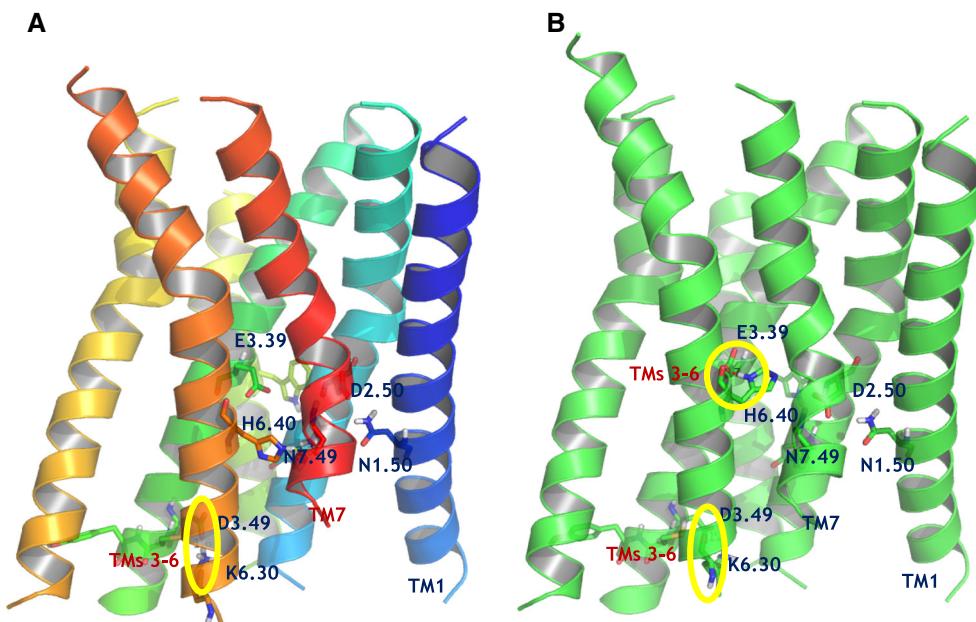
#### Analysis of the predicted best conformation

##### *The 3–6 ionic lock*

From Fig. 3, we see that D/E3.39 and H6.40 are highly conserved, occurring with 94–98 % conservation among the hORs. Thus we expected that they might form an ionic lock. Indeed we find that the best structure in Fig. 4 has this ionic lock.

However for the highly conserved E111<sup>3.39</sup> and H244<sup>6.40</sup> we find no interaction between them for the best 20 conformations. E111<sup>3.39</sup> is exposed to the inside of the protein and has no special interaction while H244<sup>6.40</sup> forms an H-bond with the backbone carbonyl atom of N286<sup>7.49</sup>. D121<sup>3.49</sup> and K234<sup>6.30</sup> are also conserved in hORs with 99 and 73 %, respectively, and indeed the best structure from SuperBiHelix leads to a salt-bridge interaction in the cytoplasmic end between K234<sup>6.30</sup> and D121<sup>3.49</sup> in the D/ERY region. This occurs for 21 of the 24 conformations.

We also examined the best structure based on the L200 = L6.50 alignment and found no salt bridge between D121<sup>3.49</sup> and K234<sup>6.30</sup> in the D/ERY region because the side chain of K234<sup>6.30</sup> was exposed to the membrane. However in this structure we find that the highly conserved E111<sup>3.39</sup> and H244<sup>6.40</sup> make a salt bridge.



**Fig. 4** Predicted best structures of human olfactory 1G1 receptor (hOR1G1) from SuperBiHelix analysis. **a** Sampling of  $\theta$ ,  $\phi$ , and  $\eta$ , **b** Sampling of  $\theta$ ,  $\phi$ ,  $\eta$ , and hydrophobic center of TM6. Highly conserved residues in family A G protein-coupled receptors (GPCRs) are shown with stick models. The best structure of SuperBiHelix has the conserved salt-bridge between D3.49 and K6.30 in the D(E)RY

region and the OR family specific H-bonding between E3.39 and H6.40. The salt-bridge is shown in yellow circle. We use Ballesteros-Weinstein numbering consisting of the TM helix number followed by residue number relative to the highly conserved residue in the helix, numbered as 50. H-bonding is indicated by the dotted lines

Based on these results we concluded that we must have the wrong hydrophobic center (HPC) for TM6. In order to optimize the HPC of TM6, we carried out a SuperBiHelix calculation in which we sampled the translation of TM6 up to  $\pm 4$  residues by 2 residue increments, while simultaneously sampling the other angles by  $\pm 10^\circ$  for  $\theta$ ,  $\pm 30^\circ$  for  $\phi$ , and  $\pm 30^\circ$  for  $\eta$  by  $15^\circ$  increments for a total of  $(3 \times 5 \times 5)^6(5 \times 3 \times 5 \times 5) = 6.6$  trillion configurations. Re-indexing the best SuperBiHelix angles from the first round as zero, we find that the final best structure from the 2nd SuperBiHelix has  $10^\circ, 0^\circ, -10^\circ, 10^\circ, -10^\circ, 10^\circ, 0^\circ$  for theta,  $-30^\circ, -30^\circ, -30^\circ, 0^\circ, 15^\circ, 30^\circ, 0^\circ$  for phi,  $0^\circ, -15^\circ, 0^\circ, 0^\circ, 30^\circ, 15^\circ, 0^\circ$  for eta of each TM, and 249.7 for HPC of TM6 compared to 251.7 (i.e. the helix moved up by two residues from the starting structure). The initial SuperBiHelix structure (all 0 angles for theta, phi, and eta and 251.7 for HPC of TM6) is ranked at number 24, as shown in Table S4. This observation that the HPC should have been 249.7 suggests that we should reconsider how the HPC was determined from the hydrophobicity analysts. Thus, to obtain the best structure requires both optimization of translation of TM6 as well as simultaneous tilting/rotation.

The final best structure now has the expected conserved salt-bridge between D3.49 and K6.30 in the D(E)RY region, as shown in Fig. 4. In Fig. 4b, we see that hOR forms favorable specific interaction between the conserved

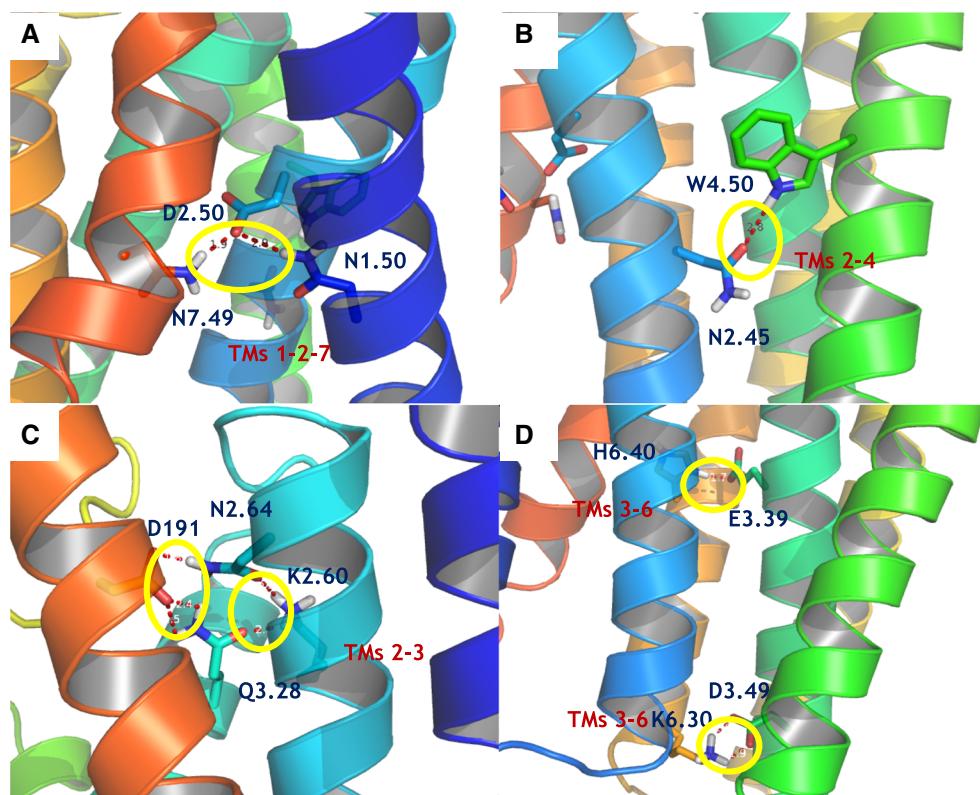
E111<sup>3.39</sup> and H244<sup>6.40</sup>, giving further credence to this final predicted structure. This is likely an important conformational feature for all hORs because of the ~94–98 % conservation among all hOR sequences.

We also carried out the BiHelix/SuperBiHelix procedure to predict for the family heads of seven clusters formed from the 398 hORs as shown in Fig. S1 (1-OR1B1, 2-OR8S1, 3-ORV1, 4-O11A1, 5-O13H1, 6-O10AD, 7-O51J1). Here we found strong interactions between the highly conserved E111<sup>3.39</sup> and H244<sup>6.40</sup> among most of the Top10 low energy conformations. However for cluster number 7, the family head 7-O51J1 does not have E3.39. Here we found that the ionic lock between TM 3 and 6 was replaced by a H-bonding network among TM 2, 3 and 4, R75 (2.52): D113 (3.38), D113 (3.38): S165 (4.50) and R75 (2.52): S165 (4.50).

#### Analysis of the interhelical interactions

As shown in Fig. 5, we find the following important interhelical interactions:

- TMs 1–2–7: We find strong interactions among N42<sup>1.50</sup>–D70<sup>2.50</sup>–N286<sup>7.49</sup>. This is also observed in 21 of the 23 other conformations. This interaction is observed in many class A GPCRs, but not in human Chemokine CXCR4 receptor (hCXCR4) which instead has an interaction between R77<sup>2.43</sup> and N298<sup>7.49</sup> [13].



**Fig. 5** H-bonding networks and salt-bridges in the predicted best structures of human olfactory 1G1 receptor (hOR1G1) in the transmembrane (TM) regions. **a** TMs 1–2–7 among N42<sup>1.50</sup>–D70<sup>2.50</sup>–N286<sup>7.49</sup> which is observed in many class A GPCRs, **b** TMs 2–4 between N65<sup>2.45</sup>–W149<sup>4.50</sup>, **c** TMs 2–3, **d** TMs 3–6

between D3.49 and K6.30 in the D(E)RY region and the OR family specific H-bonding between E3.39 and H6.40. We use Ballesteros-Weinstein numbering consisting of the TM helix number followed by residue number relative to the highly conserved residue in the helix, numbered as 50. H-bonding is indicated by the dotted lines

- TMs 2–4: We find a coupling N65<sup>2.45</sup>–W149<sup>4.50</sup> as in hCXCR4 (between H79<sup>2.45</sup> and W161<sup>4.50</sup>). However in other GPCRs, there is an extra interaction between S/N2.45 and T/S/N3.42. We do not find this TMs 2–3 interaction because the corresponding amino acid of T/S/N3.42 in other GPCRs is replaced by the hydrophobic residue, L114<sup>3.42</sup>, as in hCXCR4 (I126<sup>3.42</sup>). This is also observed in 6 of the 23 other conformations.
- TMs 2–3: We find an H-bonding network between a highly conserved Q100<sup>3.28</sup> and a semi-conserved (~77 %) K80<sup>2.60</sup>, which is also stabilized through intrahelical interaction of Q100<sup>3.28</sup> with a conserved N84<sup>2.64</sup>. This is also observed in 21 of the 23 other conformations. This constraint might be important in many ORs, but is not observed in class A GPCRs.

#### H-bonding networks in the low-lying ensemble of structures

As shown in Table 2, we find the following important H-bonding and/or salt-bridge interactions in top 24 low energy structures:

- TMs 1–2–7: We find strong couplings between N42<sup>1.50</sup>–S67<sup>2.47</sup> (except #20, 24) and between D70<sup>2.50</sup>–N286<sup>7.49</sup> (except #12, 18, 19). N42<sup>1.50</sup>, D70<sup>2.50</sup>, and N286<sup>7.49</sup> are in close proximity for all cases.
- TMs 2–3: For the upper TM close to the binding site, we find H-bonds for N84<sup>2.64</sup>–Q100<sup>3.28</sup> (except #20, 24) and for K80<sup>2.60</sup>–Q100<sup>3.28</sup> (except #2, 17, 20).
- TMs 2–4: We find strong couplings between N65<sup>2.45</sup>–W149<sup>4.50</sup> for #2, 10, 11, 14, 17, 20, and 24.
- TMs 1–7: Only structures #12 and 18 showed a salt-bridge between E23<sup>1.31</sup> and K270<sup>7.33</sup>.
- TMs 3–6: We find a salt-bridge between D121<sup>3.49</sup>–K234<sup>6.30</sup> (except #15, 18, and 24) in the D/ERY region which corresponds to the conserved 3–6 salt bridge in family A GPCRs. We also find E111<sup>3.39</sup>–H244<sup>6.40</sup> for 8 cases out of total 24 (#1, 2, 6–8, 10, 17 and 20).
- Top12, 18 and 24 in bold in Table 2 do not have TMs 3–6 interaction in the middle and at the bottom, which might be the agonist preferred conformations.

**Table 2** The interhelical H-bond or salt-bridge distance ( $\text{\AA}$ ) between the closest hetero atoms in top 24 predicted structures of the human olfactory 1G1 receptors (hOR1G1) from the SuperBiHelix analysis in Table S4

#	1.50–2.47 N42_O-S67	2.50–7.49 D70_O1-N286_N	2.64–3.28 N84_N-Q100_N	2.60–3.28 K80-Q100_O	2.31–7.33 E23_O-K270	3.49–6.30 D121_O2-K234	3.39–6.40 E111_O1-H244
1	2.79	3.35	3.09	3.30	8.99	2.59	2.69
2	2.69	2.65	<i>4.95</i>	<i>10.01</i>	9.78	2.85	2.74
3	2.78	3.24	<i>4.09</i>	2.83	8.23	2.76	<i>5.06</i>
4	2.78	2.74	<i>4.09</i>	2.83	8.23	2.74	<i>6.11</i>
5	2.78	2.76	3.78	2.82	8.23	2.76	<i>4.59</i>
6	2.78	3.24	<i>4.09</i>	2.83	8.23	2.76	<i>3.74</i>
7	2.79	3.22	3.09	3.30	8.99	2.65	2.73
8	2.79	3.22	3.09	3.30	8.99	3.03	2.73
9	2.79	3.64	<i>4.83</i>	2.82	<i>7.11</i>	2.66	<i>7.18</i>
10	2.79	3.22	3.09	3.30	8.99	2.64	2.73
11	2.79	2.77	<i>4.09</i>	2.83	8.23	2.74	<i>5.11</i>
<b>12</b>	<b>2.72</b>	<b>9.68</b>	<b>4.59</b>	2.96	2.69	<b>7.23</b>	<b>7.14</b>
13	2.79	3.24	<i>4.09</i>	2.83	8.23	2.77	<i>5.48</i>
14	2.79	3.27	<i>4.09</i>	2.83	8.99	2.74	<i>4.95</i>
15	2.79	3.14	<i>4.09</i>	2.83	<i>7.11</i>	4.67	<i>4.91</i>
16	2.78	2.80	<i>4.09</i>	2.83	8.23	2.76	<i>4.34</i>
17	2.86	2.65	<i>4.80</i>	<i>10.21</i>	6.24	3.31	2.77
<b>18</b>	<b>2.72</b>	<b>9.68</b>	<b>4.59</b>	<b>2.96</b>	<b>2.69</b>	<b>6.83</b>	<b>7.14</b>
19	2.79	5.25	<i>4.09</i>	2.83	<i>10.09</i>	2.57	<i>4.84</i>
20	<i>4.60</i>	2.75	<i>7.61</i>	<i>6.56</i>	<i>13.83</i>	2.87	2.72
21	2.79	3.38	<i>4.14</i>	3.26	<i>11.33</i>	2.57	<i>6.25</i>
22	2.79	3.26	<i>4.09</i>	2.83	8.99	2.64	<i>5.12</i>
23	2.79	3.27	<i>4.09</i>	2.83	8.99	2.58	<i>4.72</i>
<b>24</b>	<b>4.03</b>	<b>3.34</b>	<b>9.16</b>	<b>2.92</b>	<b>17.54</b>	<b>6.83</b>	<b>10.67</b>

The structures were ordered by the average rank of charge interhelical, charge total, neutral interhelical, and neutral total E (kcal/mol). The cases with longer than 5  $\text{\AA}$  are shown in italic

## Docking studies

### Binding sites of five odorants

Using DarwinDock (see “DarwinDock”), we considered 20–30 conformations for each of five odorants.

- nonanal (fatty-roselike), 27.3 nM
- 9-decen-1-ol (fresh, dewy, rose), 25.1 nM
- 1-nonanol (fresh, clean, fatty, floral, rose, orange, dusty, wet), 24.4 nM
- camphor (aldehydic, green, camphor, pine), 22.6 nM
- *n*-butanal a non-binder of hOR1G1

and predicted the best binding pose to each of the 24 low energy conformations of OR1G1 structures.

We docked all five ligands into each of the best 24 predicted structures from SuperBiHelix. Table 3 shows the summary of the docking result for five odorants. Each ligand shows a different preference of protein conformation (Top1, Top12, Top18 and Top24) leading to an H-bond at Q100<sup>3,32</sup>, T202<sup>5,42</sup>, T206<sup>5,46</sup>, and T279<sup>7,42</sup>, depending on the odorants. For example, 9-decen-1-ol

prefers the Top1 conformation, but interacting with T279<sup>7,42</sup>, while 1-nonanol prefers to binds to the Top24 conformation through H-bonding at Q100<sup>3,32</sup>. Camphor and *n*-butanal form H-bonds at T206<sup>5,46</sup> but with the Top18 and Top12 conformation, respectively. The docking results of nonanal to all 24 structures are shown in Table S5 ordered by the UCav.

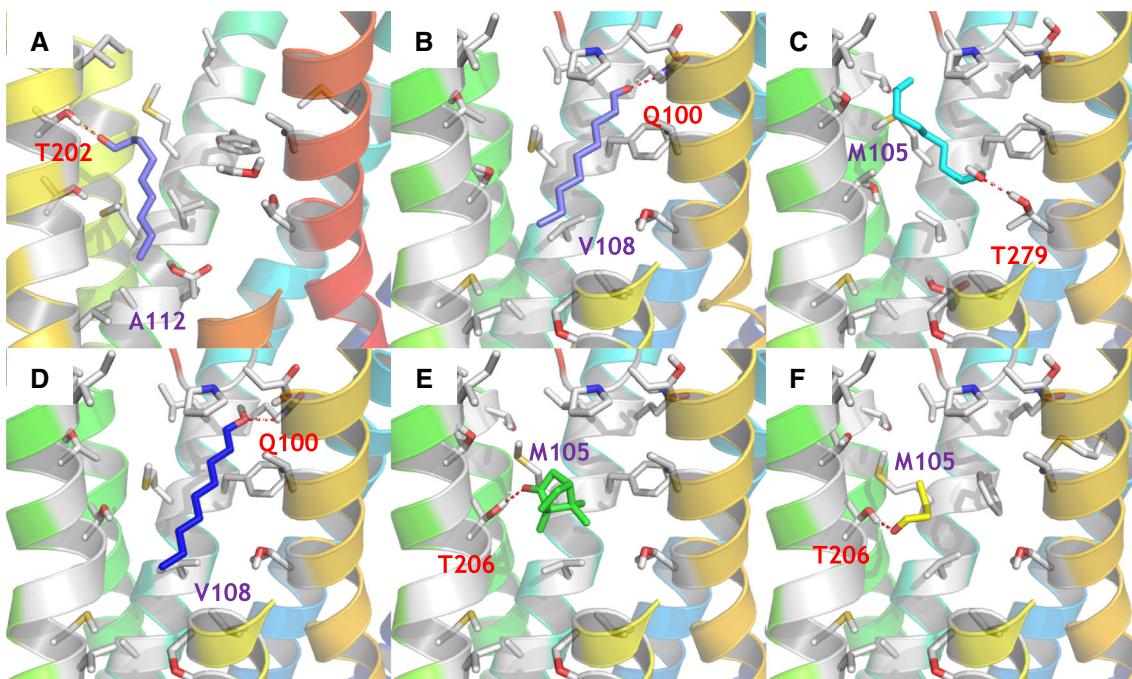
Here the binding affinities are based on intracellular calcium imaging experiments [7], representing functional activation, not necessarily binding.

The final best docking pose of nonanal is to Top24 of the ensemble (24th low energy conformation from SuperBi-Helix) leading to an H-bond to T202<sup>5,42</sup> in Fig. 6a and Table 4 (top1). The nonanal is vertically positioned in upper TMs 3, 5, 6, and 7. The aliphatic tail of the ligand is directed toward the hydrophobic residues, M105<sup>3,33</sup>, V108<sup>3,36</sup>, A112<sup>3,50</sup>, V248<sup>6,44</sup> and F252<sup>6,48</sup>, at the middle of TMDs (Table 4 top1). This binding pose is consistent with experiments on the A112S mutant; which shows weaker response of the receptor for a variety of ligands having a long aliphatic tail, compared with the wild type [46]. The second best docking pose for nonanal [based on the unified

**Table 3** The docking summary of 5 odorants at Top 24 predicted structures of the human olfactory 1G1 receptors (hOR1G1) from the SuperBiHelix analysis

Odorants	Ca <sup>2+</sup> Conc. (nM)	UCav (kcal/mol)	Selected OR1G1 conformation	H-bonding residues
Nonanal	27.3	−24.69	Top24	T202
9-Decen-1-ol	25.1	−23.02	Top1	T279
1-Nonanol	24.4	−23.32	Top24	Q100
camphor	22.6	−17.10	Top18	T206
n-Butanal	—	−11.02	Top12	T206

The label Top number indicates the rank order within the top 24 structures from the SuperBiHelix sampling. UCav (Unified cavity) energy is the sum of van der Waals (vdW), Coulomb, and H-bond energy (kcal/mol) in the unified cavity



**Fig. 6** Predicted best binding mode of odorants bound to the human OR1G1 olfactory receptor (hOR1G1) **a** nonanal, **b** the second best pose of nonanal (purple), **c** 9-decen-1-ol (cyan), **d** 1-nonanol (blue), **e** camphor (green), **f** *n*-butanal (yellow)

cavity energy (UCav) in Table S5] prefers the Top1 conformation from SuperBiHelix, leading to a H-bond of the nonanal hydroxyl to Q100<sup>3,28</sup> in Table 4 (top2). Figure 6b shows the predicted binding mode. The ligand is vertically positioned between the upper parts of TMs 3, 5, 6, and 7. The aliphatic tail of nonanal is directed toward the same hydrophobic residues, M105<sup>3,33</sup>, V108<sup>3,36</sup>, A112<sup>3,50</sup>, V248<sup>6,44</sup> and F252<sup>6,48</sup>, near the middle of the TMDs (Table 4 top2).

The predicted structures for the other four ligands to their best conformation are shown in Fig. 6c–f. The best binding pose of 9-decenol shows the opposite orientation of the aliphatic tail toward the hydrophobic residues of upper TM3, M105<sup>3,33</sup> and L101<sup>3,29</sup>. The hydroxyl group forms

the H-bonding at T279<sup>7,42</sup>. For the best pose of nonanol is the same as the second best pose of nonanal, forming an H-bond to Q100<sup>3,28</sup>. The bulky ring of camphor interacts with the hydrophobic residues, M105<sup>3,33</sup>, V108<sup>3,36</sup>, and F252<sup>6,48</sup>. Camphor and butanal form H-bonds at T206<sup>5,46</sup>. However the butanal forms a weak hydrophobic interaction with those hydrophobic residues because of short chain length.

This predicted binding E is in good agreement with results obtained by a previous study of the OR1G1 odorant [7]. We also find the H-bonding interactions at T206<sup>5,46</sup> and T279<sup>7,42</sup> as an alternative binding. Since the binding affinity is not particularly high, we consider that there may be multiple binding poses.

**Table 4** Cavity energy of nonanal bound to the human olfactory 1G1 receptors (hOR1G1)

Top1					Top2						
Residue #	VdW	Coulomb	H-bond	NonBond	Residue #	VdW	Coulomb	H-bond	NonBond		
THR	202	3.01	-3.09	-4.69	-4.77	GLN	100	-0.62	-1.77	-2.15	-4.55
PHE	252	-3.05	-0.02	0.00	-3.07	PHE	252	-3.89	0.05	0.00	-3.84
VAL	108	-2.81	0.11	0.00	-2.70	PHE	104	-2.85	-0.22	0.00	-3.07
PHE	256	-2.39	0.05	0.00	-2.34	MET	105	-2.19	0.02	0.00	-2.18
MET	105	-1.82	0.01	0.00	-1.81	VAL	276	-1.74	0.01	0.00	-1.74
MET	109	-1.75	0.15	0.00	-1.61	LEU	101	-1.70	-0.02	0.00	-1.72
GLY	203	-1.61	0.13	0.00	-1.48	VAL	108	-1.58	0.00	0.00	-1.58
GLY	207	-1.15	0.01	0.00	-1.14	THR	279	-0.95	0.00	0.00	-0.95
LEU	208	-0.92	-0.06	0.00	-0.98	VAL	248	-0.77	-0.02	0.00	-0.79
PHE	104	-0.74	0.00	0.00	-0.73	THR	206	-0.50	-0.05	0.00	-0.56
APP	259	-0.41	-0.01	0.00	-0.42	GLY	207	-0.55	0.07	0.00	-0.48
THR	206	-0.67	0.30	0.00	-0.38	PHE	256	-0.35	0.03	0.00	-0.32
ILE	199	-0.80	0.46	0.00	-0.35	SER	249	-0.27	0.02	0.00	-0.26
ALA	112	-0.40	0.07	0.00	-0.34	MET	81	-0.35	0.10	0.00	-0.25
GLY	204	-0.18	-0.11	0.00	-0.29	PRO	192	-0.41	0.18	0.00	-0.23
GLP	111	-0.20	0.05	0.00	-0.15	GLY	203	-0.19	-0.01	0.00	-0.21
GLY	253	-0.13	-0.01	0.00	-0.14	LEU	208	-0.18	0.01	0.00	-0.18
SER	255	-0.19	0.06	0.00	-0.13	PHE	107	-0.15	-0.02	0.00	-0.17
VAL	248	-0.12	-0.01	0.00	-0.13	THR	202	-0.17	0.03	0.00	-0.14
SER	156	-0.22	0.09	0.00	-0.13	VAL	280	-0.12	0.00	0.00	-0.12

Top 20 residues are ordered by total non-bond energy, which is the sum of van der Waals (vdW), Coulomb, and H-bond energy (kcal/mol) in the unified cavity

#### Previous predictions for the binding site to OR1G1

We recently published a docking study based on our best structure with the alignment at T254 = 6.50 [7]. This indicated that residues of F104<sup>3.32</sup>, M105<sup>3.33</sup>, V108<sup>3.36</sup>, T202<sup>5.42</sup>, T206<sup>5.46</sup>, F256<sup>6.52</sup>, F260<sup>6.56</sup> and T279<sup>7.42</sup> play an important role in binding of five odorants, decen-1-ol, nonanol, nonanal, camphor, and butanal. All poses had the hydrophilic part of the odorant directed towards T202<sup>5.42</sup>. Focusing on the binding cavity, we find that our new structure has these binding residues in nearly the same place compared to our early homology model, differing by only 1.6 Å. The shift of the TM6 helix discussed above has little effect on the binding site, so the ligand still recognizes the same site. These hot spots may serve as important odorant binding residues for other ORs. This suggests that residues in positions close to 3.32, 3.33, 3.36, 5.42, 5.46, 6.52, 6.56 and 7.42 might be involved in broad recognition of their various odorants in the binding sites of hORs.

These hot spot residues agree with previous docking studies performed by other groups, indicating a generalized odorant binding pocket in ORs [14–16]. The homology model of the rhodopsin-based hOR2AG1 model with MD simulation revealed a binding cavity of 1-amylbutyrate at

A104<sup>3.32</sup>, V260<sup>6.48</sup>, S263<sup>6.51</sup>, S264<sup>6.52</sup>, and T279<sup>7.42</sup>, supported by site-directed mutagenesis study [6]. The mutation results for the homology model of the rhodopsin-based mouse olfactory receptor MOR42-3 demonstrate that I112<sup>3.33</sup>, V113<sup>3.37</sup>, N117<sup>3.41</sup>, V202<sup>5.38</sup>, V206<sup>5.43</sup>, R207<sup>5.43</sup>, N210<sup>5.46</sup>, and T259<sup>6.52</sup> are involved in determining the specificity of a variety of dicarboxylic acids for MOR42-3 [7]. The combined site-directed mutagenesis and rhodopsin-based homology model found that the binding site of S-(−)-citronellal and S-(−)-citronellol at human OR1A1 includes the residues of G108<sup>3.36</sup>, N109<sup>3.37</sup>, S112<sup>3.40</sup>, N155<sup>4.56</sup>, and I205<sup>5.46</sup>, while the binding site of the binding site of S-(−)-citronellal and S-(−)-citronellol at human OR1A2 includes the residues of A108<sup>3.36</sup>, K109<sup>3.37</sup>, S112<sup>3.40</sup>, S155<sup>4.56</sup>, and V205<sup>5.46</sup> [8].

In 2000–2004 our group used the MembStruk [17] method for predicting the protein structure of ORs and used the HierDock [18] method for predicting the binding sites. These studies were carried out before X-ray structures were available for GPCRs used structure prediction methods that were very primitive compared to our current methods, which have benefited from comparison to many X-ray structures of ligand-GPCR complexes. These early studies including predicting the structure of mouse OR S25

interacting with 24 odorants [19], the binding site of 56 odorants including heptanol and octanol to mouse and rat i7 ORs with validation through experiments [20, 17], and the differential binding of ketones to the mouse and human orthologs of OR 912-93 [21]. They used only the criterion of binding energy to predict which ORs were activated by which ligands.

Several groups have attempted to understand the binding of odorants using homology modeling to X-ray structures of other GPCRs combined with site-directed mutagenesis of ORs [14–16]. Because the ORs have rather low homology (less than ~20 %) with known crystal structures of GPCRs, we expect that homology methods would not be adequate to determine a sufficiently accurate structure to determine the binding site.

#### Molecular dynamics (MD)

For the best predicted ligand–protein complex for each ligand, we constructed the loops using homology to  $\text{t}\beta_1\text{AR}$ , built the N and C termini using Maestro program and then optimized the loops and the N/C termini through 10 cycles of simulated annealing from 50–600 K (100 ps per cycle). We then carried out two independent simulations.

- Case 1 included just the one disulfide coupling between C97 and C189 which is conserved across the family A GPCRs.
- Case 2 included the second disulfide coupling between C169 and C179 in the second extracellular loop (EC2) which is conserved in 93 and 98 % of the hOR family, respectively.

After 10 cycles of simulated annealing, we observed that case 2 with two sets of disulfide bonds formed an  $\alpha$ -helix within EC2 after 8 cycles as shown in Fig. S3. Indeed SSPro predicts a helical secondary structure for this part of EC2 [22]. This, we conclude that both disulfide bonds should be present. Supporting this, hOR1D2 (alternately known as OR17-4) was confirmed the presence of the intra-EC2 disulfide bond (Cys-169–Cys-189) by mass spectrometry [23].

Other observations from the MD are that D191 in the EC2 forms an additional H-bond with Q100<sup>3.28</sup> and N84<sup>2.64</sup>, which are involved in the H-bond networks of TMs 2–3. In addition, we found new intracellular H-bonds among S165 and N195, D180 and N182, and N171 and E173 in the EC2.

To study the effect of membrane and water on the structures of the ligand-GPCR complex, we inserted the predicted protein–ligand complexes into a periodic infinite membrane fully solvated with water (37,841 total atoms) and carried out 40 ns of MD at 300 K.

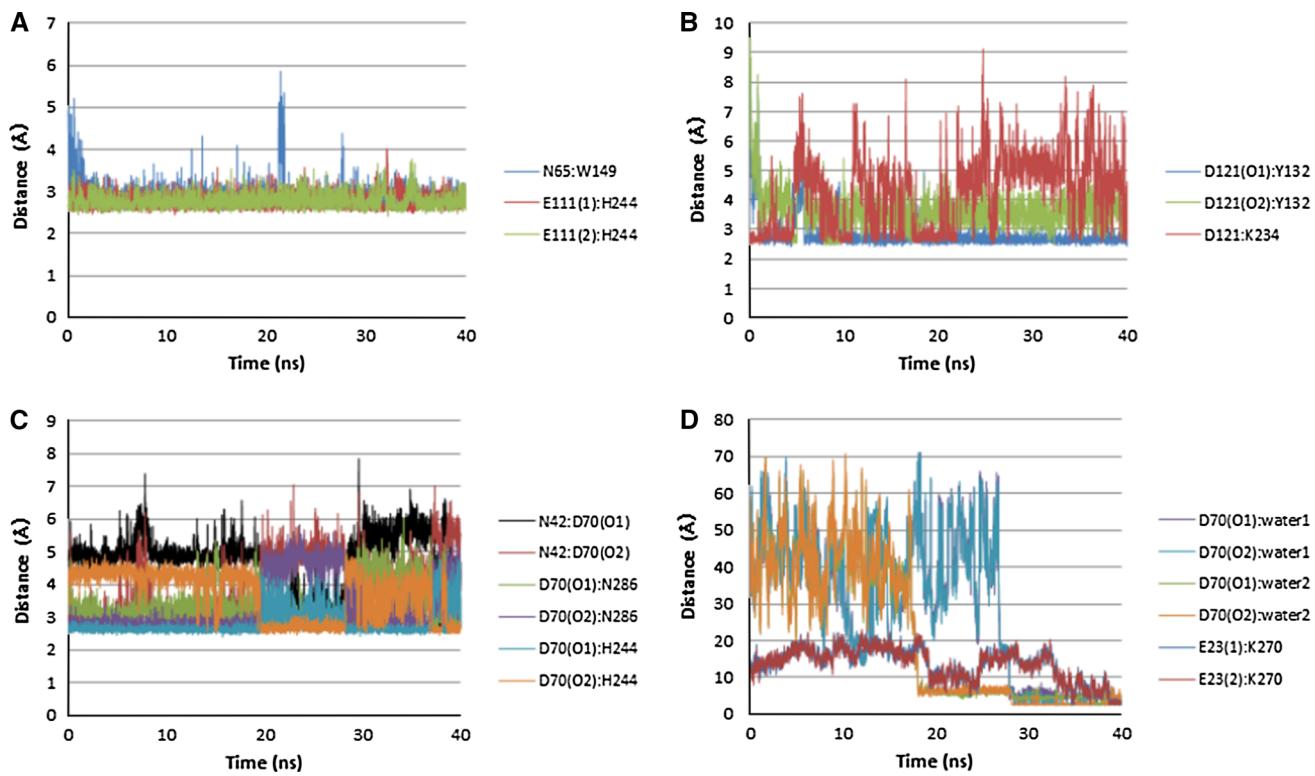
We found some modest changes in the helix–helix interactions after the MD.

1. After 40 ns of MD, the largest total change from beginning to end is observed in TM1 with  $-72.1^\circ$  rotation, because of lack of interhelical constraints compared with other helices. These TM1 changes lead to formation of a new salt-bridge between E23<sup>1.31</sup> and K270<sup>7.33</sup>.
2. TM2 also leads to a large total change of  $\pm 34.9^\circ$  rotation. These changes in TM2 result from formation of a new HB in the upper TM region between T77<sup>2.57</sup> and Q100<sup>3.28</sup> that replace the interactions among K80<sup>2.60</sup>, N84<sup>2.64</sup>, and Q100<sup>3.28</sup> found in our SuperBi-Helix structure prediction.
3. We also observed new water mediated interactions within the NPxxY motif of TM7, also often reported in crystal structures. We found that at 28 ns a water inserts between D70<sup>2.50</sup> and N42<sup>1.50</sup> to form a water mediated H-bond. At 37 ns we found another water inserted between D70<sup>2.50</sup> and N286<sup>7.49</sup> leading to two new H-bonds in place of the original direct interaction between D70 and N286 in our predicted structure. We expect that this water is protonated to balance the charge on D70.
4. We find that the strong TMs 1–2–7 and TMs 3–6 interactions were maintained during the MD. Also the x, y, and HPC have small changes of less than  $\pm 5 \text{ \AA}$  from the beginning to the end, as shown in Fig. S4. However the azimuthal angle  $\varphi$  for tilting of TM6 ( $-19.7^\circ$ ) and the rotations  $\eta$  of TMs 1 ( $-72.1^\circ$ ), 2 ( $-34.9^\circ$ ), 6 ( $-59.0^\circ$ ) exhibit large changes Table S6.

These simulations found the following intra-TM HB networks to be stable for the Apo-hOR1G1 dynamics, as shown in Fig. 7.

N42<sup>1.50</sup> and D70<sup>2.50</sup>; with the average heteroatom distance of  $4.6 \text{ \AA}$  through water,  
D70<sup>2.50</sup> and N286<sup>7.49</sup>; the average heteroatom distance of  $3.4 \text{ \AA}$ ,  
D70<sup>2.50</sup> and H244<sup>6.40</sup>; the average heteroatom distance of  $2.8 \text{ \AA}$ ,  
N65<sup>2.45</sup> and W149<sup>4.50</sup>; the average heteroatom distance of  $2.9 \text{ \AA}$ ,  
E111<sup>3.39</sup> and H244<sup>6.40</sup>; the average heteroatom distance of  $2.8 \text{ \AA}$ ,  
D121<sup>3.49</sup> and K234<sup>6.30</sup>; the average heteroatom distance of  $4.2 \text{ \AA}$ ,  
D121<sup>3.49</sup> and Y132<sup>1C2</sup>; the average heteroatom distance of  $2.8 \text{ \AA}$ .

In addition we found that several new HBs form during the dynamics:



**Fig. 7** Trajectory analysis of H-bonding and salt-bridge interactions among transmembrane (TM) helix. The distance was measured between two heavy atoms of the H-bond donor and acceptor. **a** TMs 2–4, N65 (2.45): W149 (4.50) with distances ranging from 2.51 to 5.78 Å, and TMs 3–6, E111 (3.39): H244 (6.40) with distances ranging from 2.40 to 4.01 Å, **b** D(E)RY, D121 (3.49): Y132 (IC2) with distances ranging from 2.40 to 8.78 Å, D121 (3.49): K234 (6.30)

- The interaction between E23<sup>1.31</sup> and K270<sup>7.33</sup> was formed in the upper TM region; the heteroatom distance changes from 13.3 to 2.8 Å.
- D121<sup>3.49</sup> and Y132<sup>IC2</sup> in the D(E)RY region; the heteroatom distance changes from 8.3 to 2.8 Å. This additional interaction is also observed in the several crystal structures such as tβ<sub>1</sub>AR
- [24], rM<sub>3</sub>MR [25], hM<sub>2</sub>MR [26], hAA<sub>2</sub>AR [27], and hD<sub>3</sub>DR [28].

Thus, we conclude that the hOR family specific H-bonding network between highly conserved E3.39 and H6.40 and the disulfide constraint in the EC2 are an important structure constraints in the hOR family.

## Conclusions

We predicted the structure for hOR1G1 using GEnSeMBLE methods which sample trillions of potential packings to find the ensemble of the best 24 packings (helix rotations

with distances ranging from 2.47 to 9.12 Å, **c** TMs 1–2–7, N42 (1.50): D70 (2.50) with distances ranging from 2.52 to 7.85 Å, D70 (2.50): N286 (7.49) with distances ranging from 2.49 to 6.05 Å, D70 (2.50): H244 (6.40) with distances ranging from 2.40 to 5.04 Å, **d** TMs 1–7, E23 (1.31): K270 (7.33) with distances ranging from 2.55 to 23.06 Å and water-mediated interaction with D70 (2.50) with distances ranging from 2.40 to 71.14 Å

and tilts) based on predicted interhelical and total E. Our predictions of the best conformations of the hOR1G1 lead to several conclusions:

- Out of all 28,000 structures (2,000 from each 14 BiHelix runs, four different templates, tβ<sub>1</sub>AR, hS1P1R, hOPRK, and bRho, T6.50 vs. L6.50 alignment, fine ±60° vs. coarse ±360° rotation), the tβ<sub>1</sub>AR homology template with T6.50 alignment from fine sampling (0°, 15°, -15°, -30°, 0°, 0° for eta of TM 1–7) showed the best energy
- After a ±10° sampling of θ tilt angle by 10° increment and a ±30° sampling by 15° increment of both φ tilt and η rotation angles, leading to a total of (3 × 5 × 5)<sup>7</sup> ~ 10 trillion combinations, the final best structure has the classical interhelical H-bonding networks in TMs 1–2–7 (N1.50, D2.50 and N7.49) and TMs 2–4 (N2.45 and W4.50), which are known in many members of the rhodopsin family A GPCRs. We also observed the conserved salt-bridge between D3.49 and K6.30 in the D(E)RY region. We find the H-bonding network between a highly conserved

Q100<sup>3,28</sup> and a semi-conserved K80<sup>2,60</sup> which is also stabilized through intrahelical interaction with a conserved N84<sup>2,64</sup>. This constraint might be important in many ORs.

3. From further optimizing the HPC of TM6, we observed hOR specific interaction between the conserved D/E3.39 and H6.40 in hORs, which might be an important conformational strain in all hORs because of high conservation in the sequence of hORs.
4. Two disulfide bridge constraints between C97<sup>3,25</sup> and C189, which is conserved in family A GPCRs, and between C169 and C179 in the EC2, which is conserved in most of the hOR family stabilize the  $\alpha$ -helix of EC2. This is consistent with secondary structure predictions.

We expect our predicted 3D structures for hOR1G1 will be useful to understand the important structure constraint in all the hOR family and thus design better odorants.

## Methods

### Protein structures for the hOR1G1

We used the GEnSeMBLE method [6] to generate the 3D structures for the ensemble conformations of hOR1G1. GEnSeMBLE provides a very complete sampling (millions to trillions) over possible rotations and tilts, leading to an ensemble of energetically low lying structures expected to include those conformations energetically accessible for binding of ligands. This replaces our earlier MembStruk method [17].

The structure prediction methodology has been described previously [6], and is summarized here:

1. *PredicTM*: Uses multiple sequence alignments of over 1,462 GPCR sequences using the MAFFT [29] program to predict the TM domains (TMDs) for membrane protein. These TMDs were extended by capping rules and the secondary structure prediction using PORTER [30] and SSPro [22], as shown in Fig. S5 of Supporting Information.
2. *OptHelix*: OptHelix uses molecular dynamics to predict the structure for each TMD, which is usually helical except for kinks that may be caused by Prolines.
3. *Homologize Helices*: When closely related X-ray structures are available, we also try using homology helices rather than OptHelix. For hOR1G1, we considered the following templates: the X-ray structures for t $\beta_1$ AR [24], bRho [31–35], hS1P1R [36], and hOPRK [37]. The sequence identities of the TM regions for these systems are 22.7 % for t $\beta_1$ AR,

20.7 % for hS1P1R, 17.9 % for hOPRK, and 17.4 % for bRho. The other X-ray and predicted structures all had lower sequence identity.

4. *BiHelix*: Samples all possible packings by allowing 30° increments in the rotation about each TMD axis ( $\eta$  angle), leading to  $(12)^7 \sim 35$  million packings. To make it practical to evaluate the energies for all 35 million packings we use the BiHelix method in which the 7-TMD interaction problem is partitioned into 12 sets of pairwise interactions, which are added together (mean field approximation). For each of the 1,728 pairwise interactions we use SCREAM [12] to optimize the side-chains. The BiHelix mean field energies for all 35,000,000 packings are used to select the best 2,000. We then build the full-helix bundle for each of these 2,000 structures, optimize the side-chains for each using SCREAM, and neutralize the charged residues for more accurate energy scoring. The Dreiding D3 force field (D3FF) [38] was used throughout wherever energies were evaluated.
5. *SuperBiHelix*: Starting with the experimental tilt angles we validated in the reported data [6] that the BiHelix step always identifies the experimental rotation angles correctly. However we have shown that even for closely related GPCRs, we must optimize the tilt angles: theta  $\theta$  (tilting away from the z-axis) and phi  $\phi$  (the azimuthal angle of the tilting from the xy plane) angle [39]. Here we have found that starting with the best angles from the BiHelix optimization, it is sufficient to sample  $\pm 10^\circ$  for  $\theta$  tilt angle with simultaneously sampling by  $\pm 30^\circ$  for both  $\phi$  and  $\eta$  angles, leading to a total of  $(3 \times 5 \times 5)^7 \sim 10$  trillion combinations. If the optimum angle is at the boundary of these variations, we often do a 2nd round starting from the new best conformation until it is consistent. Generally the HPC for each TMD from the PredicTM analysis is placed at  $z = 0$ , but we can also use SuperBiHelix to optimize translation along the TMD axis simultaneously with optimizing rotations. The BiHelix mean field energies for all 10 trillion combinations packings are used to select the best 2,000, which are then examined much more accurately by building the 7 helix bundles and optimizing side chains. Our experience is that the mean field energies are sufficiently accurate that the best 7-helix bundle energies are all in the first 700 of the 2,000. This indicates that 2,000 was sufficient.
6. *Validation*: To validate the GEnSeMBLE method, we previously applied this procedure to six of the available GPCR crystal structures: bovine rhodopsin (PDB ID: 1u19), h $\beta_2$ AR (2rh1), hAA<sub>2</sub>AR (3eml), t $\beta_1$ AR (2vt4), bovine opsin (3cap), squid rhodopsin (2z73) [6]. First we applied only BiHelix, and compared the energies

obtained for three-helix bundles extracted from h $\beta_2$ AR and hAA<sub>2A</sub>R using the BiHelix approximation. The crystal conformation is ranked “first” for all cases in 308 sampling and all but one case in 158 sampling. The exception in the 158 sampling mode is for the t $\beta_1$ AR case, where the crystal conformation is ranked second.

These results show that the energies used are reliable enough that, upon complete sampling of the conformational space they can still identify the lowest energy conformation assumed to be the crystal conformation. Since the crystal ligand is absent during sampling (except for Bovine Opsin case which is the ligand-free form of Rhodopsin), these results indicate that the predicted low-energy GPCR conformations in the absence of ligands are identical to or close to those observed experimentally at least in the helix rotation angle  $\eta$  space.

For another validations for our methods, we also predicted the structure of hAA<sub>2A</sub>R starting with the h $\beta_2$ AR as a template. This allows direct comparison of homology based models with the predicted de novo methods. The starting structure for hAA<sub>2A</sub>R was built by placing crystal hAA<sub>2A</sub>R TM helices in the h $\beta_2$ AR template. The starting structure can be considered the homology structure and has a Ca RMSD of 2.1 Å relative to the hAA<sub>2A</sub>R crystal TM bundle. After the SuperBiHelix steps, the lowest energy conformation had an improved Ca RMSD of 1.4 Å relative to the crystal structure [40].

7. *Loop generation:* The loops and the N/C terminus including helix 8 for a new structure are modeled through homology with the template structure, and then relaxed through 10 cycles of simulated annealing between 50 K and 600 K each for 0.1 ps, followed by minimization. The disulfide bridge between C97<sup>3,25</sup> and C189 for OR1G1, which is conserved among Class A GPCRs, was constructed by homology. The OR family has a second disulfide bond between conserved C169–C179 in the EC 2 so we built this using modeling. Then keeping the 7 TMD fixed, we then minimized the loops (up to 1,000 steps or down to 0.5 RMS force) followed by annealing between 50 K and 600 K for 10 cycles.
8. *Full solvent MD:* The full protein was inserted into a fully equilibrated hydrated POPC lipid bilayer having cell size in the XY plane of 75 Å × 75 Å (102 POPC molecules) and solvated with 6,502 water molecules (37,841 total atoms) using the Membrane builder in the VMD program. The disulfide bridges between C169 & C179 and between C97 & C189 were constrained. H244<sup>6,40</sup> was treated as a protonated HIS, HSP. The procedure here was identical to that in Kim et al. [41]

The lipid tails were almost fully extended, allowing for easy insertion of the protein into the membrane. The distance

between the layers was set at  $c = 72.3$  Å to fit the actual membrane thickness plus the solvent thickness. The lateral supercell was set to fit the actual surface density of lipid molecules,  $a = 69.9$  and  $b = 68.4$  Å. We introduced disorder into the lipid bilayer patches to better resemble a membrane at 300 K by allowing random orientation of each lipid about its axis with a short (1 ps) equilibration at 300 K in vacuum. This eliminated steric contacts between the lipid atoms but left the lipid tails mostly extended. These steps in our procedure reduce the time required for equilibration of the lipid/protein complex. After inserting the protein into the lipid-water cell, lipids overlapped within 1 Å of protein and waters overlapped within 5 Å were removed.

For the particle mesh Ewald (PME) in the electrostatics calculation [42], the charge of system was balanced through replacing waters into 1 Cl<sup>−</sup> ions. After inserting the 7 helix bundle including loops into box containing the periodic membrane, water, and ions, we fixed the protein and minimized the lipid, water, ion atoms for 1,000 steps, and then equilibrated with NPT dynamics for 500 ps while continuing to keep the protein fixed. This allows the lipid and water to readjust to the protein. Then we minimized the full system for 1,000 steps. After this we did 40 ns of NPT dynamics. This predicted structure was then equilibrated at 300 K for 40 ns using the NAMD 2.9 (Nanoscale Molecular Dynamics) program [43]. We used the CHARMM22 force field parameters for the protein, the TIP3 model for water [44], and the CHARMM27 force field parameters for the lipids [45].

#### DarwinDock

Our earlier studies used the HierDock [18] and MSCDock [46] which have now been replaced by DarwinDock. For each ligand conformation, DarwinDock iteratively generates ~50,000 poses into the putative binding regions of the bulky-residue-alanized protein. This is followed by the energy scoring of family heads to select the top 10 % ordered by total energy. The top 100 conformations are chosen for further optimization. For each of these we dealanize the protein side-chains (using SCREAM) to find the optimum side chains for each of the best 100 poses. Then we neutralize the protein and ligand by transferring protons appropriately within salt bridges and protonating or deprotonating exterior ligands, followed by further full geometry minimization. We consider that use of these neutral residue charges improves the accuracy for comparing different docked structures. The result is that small changes in geometries of charged ligands far from the binding site can lead to large differential binding energies of 10–30 kcal/mol. We find that neutralizing these exposed residues removes the sensitivity to details of the distances of charged residues (and counter ions) remote from the

active site. This neutralization leads to differential binding energies that are dominated by the local cavity interactions and leads to much smaller solvation energies.

This same procedure was followed for each of seven ligand conformations generated as follows. Starting from the minimized structure of the ACT, SB, GSK1, and GSK2 ligands, we performed the conformational search of Mixed torsional/Low-mode sampling (1000 steps, 100 steps per rotatable bond, 5 kcal/mol of energy window, 0.5 Å of maximum atom deviation cutoff) using the Maestro software [47]. The low energy conformations were re-minimized by D3FF and clustered by 2.5 and 1.5 Å of RMSD in two steps. For docking, the lowest 7–10 ligand conformations (within 10 kcal/mol of the best energy) were selected out of 227 for ACT-058362, 360 for SB-706375, and 158 for GSK-1562590. The final docked structure with the best binding energy from all ligand conformations was selected.

DarwinDock has been validated for a number of X-ray co-crystals including 3 crystal structures of ligand/GPCR complexes: h $\beta_2$ AR (0.4 Å RMSD) [48], hA<sub>2A</sub>AR (0.8 Å RMSD) [27], and turkey  $\beta_1$ -adrenergic receptor (0.1 Å RMSD) [24]. This shows that we can accurately identify ligand binding sites in proteins, which can then be used to optimize the ligands with desirable properties.

**Acknowledgments** Funding for this project was provided by gifts to the Materials and Process Simulation Center (MSC) at California Institute of Technology, Pasadena, CA. In addition some funding was provided by NIH (R01NS071112 and 1R01NS073115).

## References

- de March CA, Golebiowski J (2014) A computational microscope focused on the sense of smell. *Biochimie*. doi:[10.1016/j.biochi.2014.06.006](https://doi.org/10.1016/j.biochi.2014.06.006)
- Malnic B, Hirono J, Sato T, Buck LB (1999) Combinatorial receptor codes for odors. *Cell* 96:713–723
- Duchamp-Viret P, Duchamp A, Chaput MA (2003) Single olfactory sensory neurons simultaneously integrate the components of an odour mixture. *Eur J Neurosci* 18:2690–2696
- Jacquier V, Pick H, Vogel H (2006) Characterization of an extended receptive ligand repertoire of the human olfactory receptor OR17-40 comprising structurally related compounds. *J Neurochem* 97:537–544
- Lai PC, Crasto CJ (2012) Beyond modeling: all-atom olfactory receptor model simulations. *Frontiers in Genetics* 3:1–10
- Abrol R, Bray JK, Goddard WA III (2011) Bihelix: towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. *Proteins* 80:505–518
- Charlier L, Topin J, Ronin C, Kim S-K, Goddard WA III, Efremov R, Golebiowski J (2012) How broadly tuned olfactory receptors equally recognize their agonists. Human OR1G1 as a test case. *Cell Mol Life Sci* 69:4205–4213
- Sanz G, Schlegel C, Pernollet JC, Briand L (2005) Comparison of odorant specificity of two human olfactory receptors from different phylogenetic classes and evidence for antagonism. *Chem Senses* 30:69–80
- Sanz G, Thomas-Danguin T, Hamdani EH, Poupon CL, Briand L, Pernollet J-C, Guichard E, Tromelin A (2008) Relationships between molecular structure and perceived odor quality of ligands for a human olfactory receptor. *Chem Senses* 33:639–653
- Wimley WC, Creamer TP, White SH (1996) Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochem* 35:5109–5124
- Hessa TM-B, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH, von Heijne G (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 450:1026–1030
- Kam VWT, Goddard WA III (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. *J Chem Theor Comput* 4:2160–2169
- Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330:1066–1071
- Gelis L, Wolf S, Hatt H, Neuhaus EM, Gerwert K (2012) Prediction of a ligand-binding niche within a human olfactory receptor by combining site-directed mutagenesis with dynamic homology modeling. *Angew Chem Int Ed Engl* 51:1274–1278
- Abaffy T, Malhotra A, Luetje CW (2007) The molecular basis for ligand specificity in a mouse olfactory receptor: a network of functionally important residues. *J Biol Chem* 282:1216–1224
- Schmiedeberg K, Shirokova E, Weber HP, Schilling B, Meyerhof W, Krautwurst D (2007) Structural determinants of odorant recognition by the human olfactory receptors OR1A1 and OR1A2. *J Struct Biol* 159:400–412
- Vaidehi N, Floriano WB, Trabanino R, Hall SE, Freddolino P, Choi EJ, Zamanakos G, Goddard WA III (2002) Prediction of structure and function of G protein-coupled receptors. *Proc Natl Acad Sci USA* 99:12622–12627
- Floriano WB, Vaidehi N, Zamanakos G, Goddard WA III (2004) HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J Med Chem* 47:56–71
- Floriano WB, Vaidehi N, Goddard WA III, Singer MS, Shepherd GM (2000) Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor. *Proc Natl Acad Sci USA* 97:10712–10716
- Hall SE, Floriano WB, Vaidehi N, Goddard WA III (2004) Predicted 3-D structures for mouse I7 and rat I7 olfactory receptors and comparison of predicted odor recognition profiles with experiment. *Chem Senses* 29:595–616
- Hummel P, Vaidehi N, Floriano WB, Hall SE, Goddard WA III (2005) Test of the binding threshold hypothesis for olfactory receptors: explanation of the differential binding of ketones to the mouse and human orthologs of olfactory receptor 912-93. *Protein Sci* 14:703–710
- Pollastri G, Przybylski D, Rost P, Baldi B (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* 47:228–235
- Cook BL, Steuerwald D, Kaiser L, Graveland-Bikker J, Vandenberghe M, Berke AP, Herlihy K, Pick H, Vogel H, Zhang S (2009) Large-scale production and study of a synthetic G protein-coupled receptor: human olfactory receptor 17-4. *Proc Natl Acad Sci USA* 106(29):11925–11930
- Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG, Schertler GF (2008) Structure of a  $\beta_1$ -adrenergic G-protein coupled receptor. *Nature* 454:486–491
- Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO, Shaw DE, Weis WI, Wess J, Kobilka BK (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482:552–556

26. Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T, Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482:547–551
27. Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP, Stevens RC (2008) The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322:1211–1277
28. Chien EY, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, Stevens RC (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330:1091–1095
29. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acid Res* 33:511–518
30. Pollastri G, McLysaght A (2005) Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* 21:1719–1720
31. Palczewski K, Kumada T, Hori T, Behnke CA, Motoshima H (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 289:739–745
32. Teller DC, Okada T, Behnke CA, Palczewski K, Stenkamp RE (2001) Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). *Biochemical* 40:7761–7762
33. Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y (2002) Fundamental role of internal water molecules in rhodopsin revealed by X-ray crystallography. *Proc Natl Acad Sci USA* 99:5982–5987
34. Li J, Edwards PC, Burghammer M, Villa C, Schertler GF (2004) Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol* 343:1409–1438
35. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J Mol Biol* 342(2):571–583
36. Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, Desale H, Clemons B, Cahalan SM, Schuerer SC, Sanna MG, Han GW, Kuhn P, Rosen H, Stevens RC (2012) Crystal structure of a lipid G protein-coupled receptor. *Science* 335:851–855
37. Wu H, Wacker D, Mileni M, Katritch V, Han G, Vardy E, Liu W, Thompson AA, Huang X-P, Carroll FI, Mascarella SW, Westkaemper RB, Mosier PD, Roth BL, Cherezov V, Stevens RC (2012) Structure of the human κ-opioid receptor in complex with JDTic. *Nature* 485:327–332
38. Mayo SL, Olafson BD, Goddard WA III (1990) DREIDING—a generic force field for molecular simulations. *J Phys Chem* 94:8897–8909
39. Abrol R, Griffith AR, Bray JK, Goddard WA III (2012) Structure prediction of G protein-coupled receptors and their ensemble of functionally important conformations. *Methods Mol Biol* 914:237–254
40. Abrol R, Kim SK, Bray JK, Griffith AR, Goddard WA (2011) Characterizing and predicting the functional and conformational diversity of seven-transmembrane proteins. *Methods* 55:405–414
41. Kim S-K, Li Y, Park C, Abrol R, Goddard WA III (2010) Prediction of the 3D structure for the rat urotensin II receptor and comparison of the antagonist binding sites and binding selectivity between human and rat from atomistic simulations. *ChemMedChem* 5:1594–1608
42. Bhandarkar M, Brunner R, Chipot C, Dalke A, Dixit S, Grayson P, Gullingsrud J, Gursoy A, Hardy D, Hénin J, Humphrey W, Hurwitz D, Krawetz N, Kumar S, Nelson M, Phillips J, Schinozaki A, Zheng G, Zhu F (2008) NAMD User's Guide (Theoretical Biophysics Group, University of Illinois at Urbana-Champaign and Beckman Institute, Urbana) Version 2.6
43. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802
44. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616
45. Feller SE, Yin D, Pastor RW, MacKerell J, MacKerell AD Jr (1997) Molecular dynamics simulation of unsaturated lipids at low hydration: parametrization and comparison with diffraction studies. *Biophys J* 73:2269–2279
46. Cho AE, Wendel JA, Vaidehi N, Kekenes-Huskey PM, Floriano WB, Maiti PK, Goddard WA III (2005) The MPSim-dock hierarchical docking algorithm: application to the eight trypsin Inhibitor co-crystals. *J Comput Chem* 26:48–71
47. Watts KS, Dalal P, Murphy RB, Sherman W, Friesner RA, Shelley JC (2010) ConfGen: a conformational search method for efficient generation of bioactive conformers. *J Chem Inf Model* 50:534–546
48. Cherezov V, Rosenbaum D, Hanson M, Rasmussen S, Thian FS, Kobilka TS, Choi H, Kuhn P, Weis W, Kobilka B, Stevens R (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318:1258–1265
49. Thompson AA, Liu W, Chun E, Katritch V, Wu H, Vardy E, Huang XP, Trapella C, Guerrini R, Calo G, Roth BL, Cherezov V, Stevens RC (2012) Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 485:395–399
50. Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC, Iwata S (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475:65–70
51. Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK, Granier S (2012) Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. *Nature* 485:321–326
52. Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, Kobilka BK (2012) Structure of the delta-opioid receptor bound to naltrindole. *Nature* 485:400–404
53. Peng JY, Vaidehi N, Hall SE, Goddard WA III (2006) The predicted 3D structures of the human M1 muscarinic acetylcholine receptor with agonist or antagonist bound. *Chem Med Chem* 1:878–890
54. Vaidehi N, Schlyer S, Trabanino RJ, Floriano WB, Abrol R, Sharma S, Kochanny M, Koovakat S, Dunning L, Liang M, Fox JM, de Mendonca FL, Pease JE, Goddard WA III, Horuk R (2006) Predictions of CCR1 chemokine receptor structure and BX 471 antagonist binding followed by experimental validation. *J Biol Chem* 281:27613–27620
55. Bray JK, Goddard WA III (2008) The structure of human serotonin 2c G-protein-coupled receptor bound to agonists and antagonists. *J Mol Graph Model* 27:66–81
56. Kim S-K, Li Y, Abrol R, Heo J, Goddard WA III (2011) Predicted structures and dynamics for agonists and antagonists bound to serotonin 5-HT2B and 5-HT2C receptors. *J Chem Inf Model* 51:420–433
57. Kim S-K, Riley L, Abrol R, Jacobson KA, Goddard WA III (2011) Predicted structures of agonist and antagonist bound complexes of adenosine A3 receptor. *Proteins* 79:1878–1897
58. Kim S-K, Fristrup P, Abrol R, Goddard WA III (2011) Structure-based prediction of subtype selectivity of histamine H3 receptor selective antagonists in clinical trials. *J Chem Inf Model* 51:3262–3274