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

PATRICK HUMMEL, NAGARAJAN VAIDEHI, WELY B. FLORIANO, SPENCER E. HALL, AND WILLIAM A. GODDARD III

Materials and Process Simulation Center, California Institute of Technology, Pasadena, California 91125, USA

(RECEIVED September 15, 2004; FINAL REVISION November 2, 2004; ACCEPTED November 2, 2004)

Abstract

We tested the Binding Threshold Hypothesis (BTH) for activation of olfactory receptors (ORs): To activate an OR, the odorant must bind to the OR with binding energy above some threshold value. The olfactory receptor (OR) 912-93 is known experimentally to be activated by ketones in mouse, but is inactive to ketones in human, despite an amino acid sequence identity of ~66%. To investigate the origins of this difference, we used the MembStruk first-principles method to predict the tertiary structure of the mouse OR 912-93 (mOR912-93), and the HierDock first-principles method to predict the binding site for ketones to this receptor. We found that the strong binding of ketones to mOR912-93 is dominated by a hydrogen bond of the ketone carbonyl group to Ser105. All ketones predicted to have a binding energy stronger than $E_{BindThresh}$ = 26 kcal/mol were observed experimentally to activate this OR, while the two ketones predicted to bind more weakly do not. In addition, we predict that 2-undecanone and 2-dodecanone both bind sufficiently strongly to activate mOR912-93. A similar binding site for ketones was predicted in hOR912-93, but the binding is much weaker because the human ortholog has a Gly at the position of Ser105. We predict that mutating this Gly to Ser in human should lead to activation of hOR912-93 by these ketones. Experimental substantiations of the above predictions would provide further tests of the validity of the BTH, our predicted 3D structures, and our predicted binding sites for these ORs.

Keywords: mOR912-93; hOR912-93; olfactory receptor; MembStruk; HierDock; odorant binding; structure; protein folding; Binding Threshold Hypothesis

Supplemental material: see www.proteinscience.org

There continues to be uncertainty as to what aspect of an odorant is detected by an olfactory receptor (OR). Early models assumed that the shape of the odorant was most important (Amoore 1952). Experimental results reported by Buck and Axel (1991) showed that each OR recognizes several odors and each odor is recognized by several ORs. This raises the issue of what aspects of the molecule can provide this selectivity. Turin (1996) made the novel proposal that ORs detect the intramolecular vibrations of odors. But recently Keller and Vosshall (2004) provided three distinct experimental tests of the Turin model, all of which gave no evidence to support it. This leaves unresolved exactly what it is about an odorant that activates an OR.

We recently reported first-principles computations that examine details of odorant binding to ORs (Floriano et al. 2000, 2004a; Hall et al. 2004). All results obtained thus far

Reprint requests to: William A. Goddard III, Materials and Process Simulation Center (MC 139-74), California Institute of Technology, Pasadena, CA 91125, USA; e-mail: wag@wag.caltech.edu; fax: (626) 585-0918.

Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.041119705.

are consistent with the hypothesis that the binding strength of each odorant dominates the activation profile of each OR, and that each OR has some energy threshold, $E_{BindThresh}$, below which there is no activation. This leads to *The Binding Threshold Hypothesis (BTH) for activation of olfactory receptors (ORs): To activate an OR, the odorant must bind to the OR with binding energy above some threshold,* $E_{BindThresh}$.

To further test the BTH we report here results for the binding of ketones to the mouse and human orthologs of OR 912-93 (denoted mOR912-93 and hOR912-93, respectively). mOR912-93 is activated by ketones while hOR912-93 is not (Gaillard et al. 2002).

Background on olfaction

In the last two decades there have been tremendous advances in understanding olfaction. The detection of odorants in human and mouse is mediated by hundreds of ORs belonging to the large superfamily of seven-helical transmembrane (TM) G protein-coupled receptors (GPCRs) (Buck and Axel 1991; Mombaerts 1999). There are ~2.5 times as many functional OR genes in mouse (~873) (Godfrey et al. 2004) as in human (~347) (Malnic et al. 2004), as humans possess a significantly higher percentage of OR pseudogenes (Zhang and Firestein 2002), possibly explaining the less refined olfactory acuity in human.

To obtain information about the molecular basis of olfaction, experimental studies of structure recognition of odorants by ORs have been conducted by several research groups (Araneda et al. 2000; Krautwurst et al. 1998; Zhao et al. 1998; Kajiya et al. 2001; Bozza et al. 2002). In particular, systematic experimental studies exposing a variety of small molecules to the mouse and rat I7 ORs provide information on which structural changes in the odorant molecule affect activation of ORs (Krautwurst et al. 1998; Arnaeda et al. 2000; Bozza et al. 2002). The 3D protein structures of these systems were predicted from first principles (MembStruk) by Hall et al. (2004). Hall et al. also used first-principles methods (HierDock) to predict binding sites and binding energies for the 56 odorants studied by Bozza, and obtained activation profiles in excellent agreement with experiment.

Malnic et al. (1999) demonstrated that mouse ORs can be activated by multiple odorants and that each odorant elicits a response from a variety of different ORs. Floriano et al. (2000, 2004a) used MembStruk to predict the 3D protein structures of six of these mouse ORs, and used HierDock to predict binding sites and binding energies for the 24 odorants studied by Malnic. They also found activation profiles in excellent agreement with experiment. These computational studies found that different odorants bind to the same OR in the same binding regions, but sometimes with different binding conformations (Floriano et al. 2000, 2004a). These predictions could be directly tested by mutation experiments on ORs, but no such experiments have been reported.

Gaillard et al. (2002) studied the closely related mouse and human orthologs of OR912-93 to gain an understanding of how protein sequence affects the function of these ORs. hOR912-93 has a single nonsense point mutation in the region corresponding to the N terminus of the protein (Rouquier et al. 1998), but mOR912-93 does not contain such a mutation (Rouquier et al. 1999). More recently it was found that correcting the nonsense mutation in hOR912-93 does *not* restore function (Gaillard et al. 2002). Exposure to straight-chain ketone odorants with 4–10 carbons and a carbonyl group in the second or third position results in a rise in intracellular $[Ca^{+2}]$ for mouse. No such response is observed in human, even after correcting the nonsense mutation. Gaillard et al. (2002) suggested that this indicates that pseudogenes may not be the sole reason for the relatively poor sense of smell in humans. Rather, the presence of other deleterious mutations in the human olfactory subgenome may have weakened the combinatorial code of human odor receptors over the course of evolution. To determine whether this is the case it would be useful to understand which mutations cause hOR912-93 to become inactive to ketones.

To answer such questions we applied the MembStruk first-principles method (Floriano et al. 2000; Vaidehi et al. 2002; Trabanino et al. 2004), designed specifically for predicting the 3D structures of GPCRs, and the HierDock first-principles method (Datta et al. 2002, 2003; Wang et al. 2002; Kekenes-Huskey et al. 2003; Floriano et al. 2004b) for predicting binding sites and energies. These methods have been validated for bovine rhodopsin (Trabanino et al. 2004), where the predicted 3D structure is in good agreement with the crystal structure (2.8 Å RMS deviation in coordinates for the main chain atoms of the helix bundle). These methods were also used to predict the 3D structures for the D2 dopamine receptors (Kalani et al. 2004), the β2 adrenergic receptors (Freddolino et al. 2004), the mouse and rat I7 ORs (Hall et al. 2004), and six other mouse ORs (Floriano et al. 2000, 2004a). Since bovine rhodopsin is the only GPCR for which an experimental crystal structure is known, the structures of the other receptors were validated by predicting the details of the ligand binding sites with HierDock and comparing the results to experimental data on the binding sites for agonists and antagonists to wild-type and mutated versions of dopamine and adrenergic receptors. Excellent agreement was found with all available experimental results on ligand binding. This gives us confidence that the 3D structures of GPCRs predicted by MembStruk are sufficiently accurate to obtain reasonable ligand binding site and energies.

The binding data for ORs are much less complete than that for dopamine and adrenergic receptors, and no mutation studies are available, making validation more difficult. However, we would expect the predictions for ORs to be as

accurate as those for the other receptors, at least for ligands the size of retinal, dopamine, and epinephrine.

We report here the results from first-principles predictions of the 3D structure for mOR912-93 using MembStruk, and we report the results of using the HierDock first-principles method to predict the binding site, conformations, and energies of the various ketones to this receptor. All eight ketones observed to strongly activate mOR912-93 have binding energies calculated to be higher than $E_{BindThresh}$ = 26 kcal/mol, while the two ketones with weaker calculated binding energies do not. Similarly, in hOR912-93, for which ketones are known experimentally not to bind, all ketones are calculated to bind less strongly than $E_{BindThresh}$. To provide data that could guide experiments to probe the mechanism of OR activation, we predict here that 2-undecanone and 2-dodecanone will bind more strongly than the activation threshold. We also predict that a single mutation in hOR912-93 (Gly106→Ser) will allow the mutated human receptor to be activated by the same ketones that activate mOR912-93. Similarly, mutating Ser105 to Gly in mOR912-93 should eliminate activation by these ketones.

Results

Experiments have shown that mOR912-93 is activated by straight alkyl-chain ketones having more than four carbons and a carbonyl group in the second or third position on the alkyl chain (Gaillard et al. 2002). We report here the predicted binding site and energetics for straight-chain ketones having carbon chain length ranging from four to 12 carbons, with the position of the carbonyl group varying from the first to the fourth carbon.

The predicted ketone binding site in mOR912-93 lies near the extracellular region between TMs 3, 5, and 6 (Figs. 1, 2). This predicted binding site is similar to the predicted odorant binding sites in other ORs (Floriano et al. 2000, 2004a; Singer 2000; Hall et al. 2004).

The residues within 4 Å of 2-pentanone and 2-decanone are shown in Figures 3 and 4, respectively. The binding between ketones and mOR912-93 is dominated by a strong hydrogen bond between the carbonyl group on the ketones and the hydroxyl group of Ser105 (O – O distance of 2.9 Å). The alkyl chain of the ketone also experiences favorable interactions with a tight pocket formed primarily by hydrophobic residues in TMs 3, 5, and 6. The residues in mOR912-93 that form favorable van der Waals contacts with each of the ketones in Table 1 are Val108 (TM3), Ser210 (TM5), Ile250 (TM6), and Leu251 (TM6).

Experimentally it is observed that straight-chain ketones with 10 or fewer carbons activate mOR912-93 more strongly as the length of the carbon chain increases (Gaillard et al. 2002). Our results agree with the experimental trends, with each additional carbon in the chain increasing

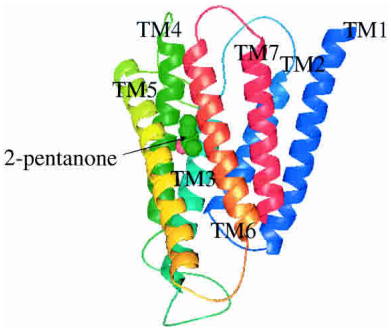

Figure 1. The side view of the predicted structure of mOR912-93, with the extracellular region on *top*. The ketone binding site lies between TM regions 3, 5, and 6.

the binding affinity by ~2 kcal/mol (Table 1). Our theoretically calculated binding energies are consistent with the BTH if $E_{BindThresh}$ is set to 26 kcal/mol for the ketones studied here. Table 1 shows that 2-butanone and 4-heptanone are the only ketones which bind to mOR912-93 less strongly than $E_{BindThresh}$. This correlates well with the experimental observation that the only ketones in Table 1 known not to strongly activate mOR912-93 are 2-butanone and 4-heptanone. It is important to note that the binding energy threshold could be valid for antagonists as well. However, the only experimental data available here are for whether a ketone activates an OR.

We predict that both 2-undecanone and 2-dodecanone will bind to mOR912-93 more strongly than the other ketones. Perturbing the binding conformation of 2-dodecanone to generate the corresponding binding conformation for 2-tridecanone (see Materials and Methods) leads the additional methyl group in 2-tridecanone to interact significantly with extracellular loop 2 in mOR912-93. Because of the difficulty in reliably predicting a unique structure for the loops, we are less confident of the accuracy in the predicted binding energy for 2-tridecanone.

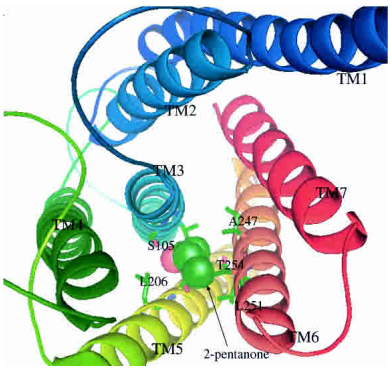

Figure 2. The top view of the predicted structure of mOR912-93. Key residues involved in the binding of 2-pentanone to mOR912-93 are depicted.

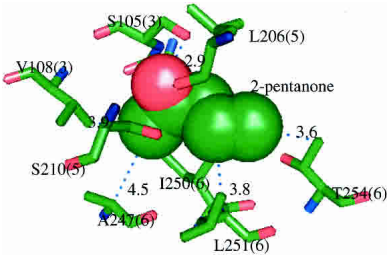

Figure 3. The residues within 4 Å of 2-pentanone bound to mOR912-93. Distances are given in angstroms, and TM domains are given in parentheses.

The experimental studies find that heptanal, 2-heptanone, and 3-heptanone all activate mOR912-93, but that 4-heptanone fails to activate the receptor. This is consistent with our results (Table 1), which indicate a substantially weakened binding for 4-heptanone. In order for 4-heptanone to form a hydrogen bond with Ser105, it must adopt a conformation where its terminal methyl group experiences repulsive van der Waals interactions with Ala247. Though the conformation with this hydrogen bond is still found to be the most energetically favorable, the unfavorable van der Waals interactions cause the binding for 4-heptanone to become weaker than that of 3-heptanone by ~7 kcal/mol.

We found that the best binding site for ketones in hOR912-93 is the region analogous to the binding site for ketones in mOR912-93. However, the calculated binding energies to hOR912-93 are, on average, 13 kcal/mol (~50%) weaker than the binding energies to mOR912-93 (Table 1). Each ketone shown in Table 1 has a calculated binding energy to hOR912-93 weaker than $E_{BindThresh}$. Hence, according to the BTH, none of these ketones should activate hOR912-93, just as observed experimentally. This difference in binding energy between mOR912-93 and hOR912-93 is due to Ser105 in mOR912-93, which makes a strong hydrogen bond with the carbonyl group of the ketones. This residue is mutated to Gly in hOR912-93. If we take our predicted structure for hOR912-93 and change this Gly to Ser while keeping all other structural features of hOR912-93 intact, and then apply HierDock to determine the binding

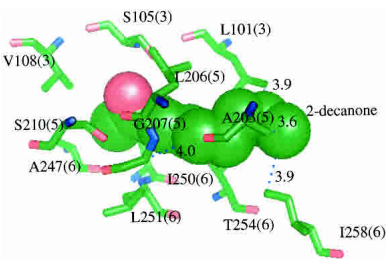

Figure 4. The residues within 4 Å of 2-decanone bound to mOR912-93, with corresponding TM domains given in parentheses, and distances given in angstroms.

site and binding conformations, we find that the predicted ketone binding conformations in the mutated hOR912-93 are analogous to those in mOR912-93 (Fig. 5). Likewise, we find that the ketone binding energies are similar for the mutated hOR912-93 and mOR912-93 (Table 2). Our results can be tested by mutating Ser105 in mOR912-93 to Gly, which should eliminate recognition of ketones in mouse, or by mutating Gly106 in hOR912-93 to Ser, which should enable recognition in human.

Discussion

Our predicted binding site is consistent with the results of a recent study aimed at finding residues conserved across orthologous ORs but not in paralogous ORs and using this to predict the amino acids involved in the binding of odors to ORs (Man et al. 2004). Man et al. found 22 residues, all in TMs 2–7 or extracellular loop 2, which could potentially be involved in the binding site of ORs. Ser105 is located at the position of one of these 22 residues.

Since the ketone carbonyl group cannot form a hydrogen bond with Gly106 in hOR912-93, there is no reason for it to tend to the same place in hOR912-93 as in mOR912-93. In fact, the optimal position of the ketone carbonyl group is predicted to be closer to the extracellular region for hOR912-93 than for mOR912-93 or the G(106)S mutant of hOR912-93. As a result, for hOR912-93 we found that the optimal configuration of 2-dodecanone interacts significantly with the extracellular loops, but not in mOR912-93 or the mutated hOR912-93. Likewise, hOR912-93 has more space in the binding site than mOR912-93 for the additional methyl groups in 4-heptanone. The effects of these differences are reflected in the data presented in Table 1.

The binding of the ketones to mOR912-93 is dominated by hydrogen bonding and van der Waals interactions. For example, for 2-pentanone bound to mOR912-93, 40% of the binding energy comes from hydrogen bonding and 52% from van der Waals interactions. There is no ketone for which the contribution of electrostatic interactions to the binding energy exceeds 10%.

This validation of the MembStruk and HierDock first-principles methods for predicting the structural differences between the mouse and human orthologs of OR912-93 demonstrated here, along with the previous studies on the I7 and Malnic ORs, suggests the possibility that such predictions of the structures and activation profiles for the other sensory receptors for humans, mice, and other species could provide the basis for a detailed and complete molecular level understanding of olfaction.

We should point out that the first-principles MembStruk method used here is specific for GPCRs. Although it has been successful for several classes of GPCRs, we have not yet tested it on any other membrane receptor, and it is not likely to be adequate for nonmembrane proteins.

Table 1. *Calculated binding energies (kcal/mol, minimized at 0K without corrections for zero-point energy or dynamics) of ketones with hOR912-93 and mOR912-93*

Ligand	Activates hOR912-93 experimentally?	Calculated binding energy to hOR912-93	Activates mOR912-93 experimentally?	Calculated binding energy to mOR912-93
2-butanone	Not tested	11.3	Very weakly	23.8
2-pentanone	Not tested	13.2	Yes	26.7
2-hexanone	Not tested	16.4	Yes	29.4
Heptanal	Not tested	18.9	Yes	30.1
2-heptanone	No	17.8	Yes	31.8
3-heptanone	Not tested	19.5	Yes	31.8
4-heptanone	Not tested	22.1	No	24.5
2-octanone	Not tested	19.7	Yes	34.0
2-nonanone	Not tested	21.2	Yes	35.6
2-decanone	No	21.6	Yes	37.1
2-undecanone	Not tested	22.6	Not tested	39.0
2-dodecanone	Not tested	19.3[a]	Not tested	41.2

Experimental activation profiles are due to Gaillard et al. 2002.
[a] Significant interactions with loops.

We should also emphasize that the mOR912-93 structure was from MembStruk but the hOR912-93 structure was built from a homology model to mOR912-93. This is likely to be accurate since there is a 66% sequence identity between these structures. Indeed it would probably tend to bias the results toward similar binding, whereas we find quite distinct differences in binding.

In addition we should note that our method of determining the optimum structure without ligand (the apo-protein) and then using this to predict binding of various ligands might not find binding sites that change dramatically upon binding, even though we optimize the whole ligand-protein complex after binding. The good agreement with observed *activation* suggests that the critical step in binding of agonists is to bind to the lowest-energy structure of the apo-protein. This supports a model of activation in which it is the bound ligand-OR complex that transforms to activate the G-protein (rather than a second apo-protein structure that is stabilized only when the agonist is bound).

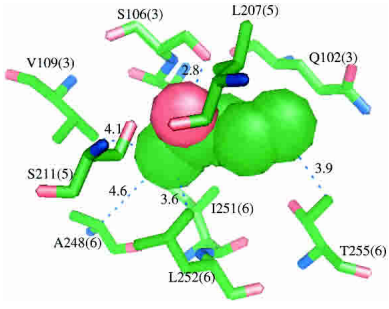

Figure 5. The residues within 4 Å of 2-pentanone bound to hOR912-93 with Gly106 mutated to Ser. Distances are given in angstroms, and TM domains are given in parentheses.

We should also point out that nothing in our calculation indicates whether the strongly bound ligand is an agonist or antagonist. Thus it is conceivable that there might be other ligands that could bind more strongly than $E_{BindThresh}$, but which might not activate the OR. However, we are not aware of any such ligands for OR.

Materials and methods

Structure prediction for mOR912-93

The MembStruk first-principles method for predicting the tertiary structures of GPCRs is described in Trabanino et al. (2004). Here we indicate the relevant details for modeling mOR912-93 and

Table 2. *Calculated binding energies (kcal/mol, minimized at 0K without corrections for zero-point energy or dynamics) of ketones for mOR912-93 and hOR912-93 with Gly106 mutated to Ser*

Ligand	Calculated binding energy to mOR912-93	Calculated binding energy to mutated hOR912-93
2-butanone	23.8	24.6
2-pentanone	26.7	26.5
2-hexanone	29.4	29.4
Heptanal	30.1	28.6
2-heptanone	31.8	30.7
3-heptanone	31.8	29.8
4-heptanone	24.5	24.3
2-octanone	34.0	31.7
2-nonanone	35.6	35.0
2-decanone	37.1	36.6
2-undecanone	39.0	38.7
2-dodecanone	41.2	40.6

hOR912-93. All calculations used the DREIDING forcefield (FF) for the protein and ligands (Mayo et al. 1990) with CHARMM22 charges (MacKerrell et al. 1998) for the protein and Gasteiger charges (Gasteiger and Marsili 1980) for the ligands or odorants. The MembStruk method (version 3.5) for predicting the structure of seven-TM-helix GPCRs consists of the following steps (Vaidehi et al. 2002; Trabanino et al. 2004):

1. TM prediction. First we predict the seven TM domains using hydrophobicity analysis combined with information from sequence alignments. The 862 complete mouse OR sequences from the Trembl database were aligned using the ClustalW multiple sequence alignment program (Altschul et al. 1990). The TM helices were predicted using TM2ndS (Trabanino et al. 2004) with multiple sequence alignment input obtained by aligning mOR912-93 with nine other closely related mouse ORs chosen from the phylogenetic tree analysis performed with ClustalW (Table S1; see Electronic Supplemental Material). In brief, TM2ndS calculates the average hydrophobicity for every residue position for all the sequences in the multiple sequence alignment while averaging over a window size from 12 to 30 residues. This leads to the profile in Figure S1 (see Electronic Supplemental Material), where the baseline serves as the threshold value for determining the TM regions.

2. Position of maximum hydrophobicity. For each TM region, we determine the maximum peak in hydrophobicity of the sequence (Table S1) to identify lipid-accessible residues from the less-conserved residues of the sequence alignments. This position of maximum hydrophobicity is used to estimate the relative translational orientation of the helices. These calculations use the Eisenberg scale for hydrophobicity (Eisenberg et al. 1984).

3. Assembly of the helix bundle and optimization of the translational orientation of the helices. The helical axes are oriented according to the 7.5 Å electron density map of frog rhodopsin (Schertler 1998), and the initial relative translational orientation of each helix is determined by placing all hydrophobic centers obtained from step 2 in the same fitting plane.

4. Optimization of helical kinks. We construct canonical helices for the predicted TM segments and optimize the structures of the individual helices using energy minimization followed by torsional molecular dynamics (NEIMO MD) (Jain et al. 1993; Mathiowetz et al. 1994; Vaidehi et al. 1996). This optimizes the bends and kinks in each helix.

5. Monte Carlo optimization of rotational orientation of the helices. The initial rotational orientation of each helix about its axis is determined by computing the direction of the net hydrophobic moment of the middle one-third of each helix about its hydrophobic center, and then pointing this net hydrophobic moment vector away from the center of mass of the whole TM bundle. Since MD is not likely to surmount the barriers that might separate one good rotational state from another, we optimize the rotational degrees of freedom of each helix in the presence of other helices as follows: We carry out a systematic search in which each helix is rotated through a grid of rotational angles. For each value, the side chain conformations in all seven helices are re-optimized with SCWRL (Bower et al. 1997), and the other six helices are optimized sequentially by minimizing energy using conjugate gradients with the DREIDING FF. This allows the system to surmount rotational energy barriers and sample many possible rotational minima for each helix. This rotational optimization is carried out for each helix in turn over the same grid until there is no longer a change.

6. Optimization of the assembled helical bundle with explicit lipid layers. The optimized helical bundle is further equilibrated by immersing the protein into a lipid bilayer and performing rigid-body MD (Ding et al. 1992; Lim et al. 1997). The helix bundle surrounded by lipid bilayers is optimized using rigid-body dynamics with the DREIDING FF (Mayo et al. 1990) and CHARMM22 (MacKerrell et al. 1998) charges for the protein.

7. Optimization of the final model. The loops are added to the helices using the loop builder in the MODELLER homology modeling program (John and Sali 2003). Disulfide linkages were created between Cys97–Cys179 and Cys169–Cys189, the pairs of cysteines conserved in ORs. We then perform side chain replacement with SCWRL (Bower et al. 1997) and optimize the positions of the atoms in the loop region with MPSim MD (Lim et al. 1997). Finally, a full-atom optimization of the structure using MPSim leads to the final predicted structure for the mouse OR.

It should be noted that MembStruk can only be used to predict the structures of GPCRs, not other types of proteins.

Structure prediction for hOR912-93

The sequence identity between mOR912-93 and hOR912-93 is 66% (Table S1). Thus we expect that homology modeling (with MODELLER) using our predicted structure for mOR912-93 as the template will lead to a reasonably accurate 3D structure of the seven-helix TM bundle of hOR912-93. Side chain replacement for hOR912-93 was performed with SCWRL followed by optimization of the positions of the atoms in the TM by minimizing the potential energy with conjugate gradients to a RMS force of 0.1 kcal/mol/Å. The loops were then added to hOR912-93 as previously outlined for mOR912-93 with disulfide bonds between Cys98–Cys180, Cys170–Cys190, and Cys123–Cys128 to close intracellular loop 2. Finally, the full structure of hOR912-93 was optimized as described above for mOR912-93.

We prefer to generate the 3D structure for hOR912-93 by homology modeling, rather than by independently predicting the 3D structure with MembStruk, because we want to eliminate noise arising from random differences due to independent constructions. Avoiding such random differences is important for comparing relative binding energies.

Prediction of odorant binding sites

HierDock

The HierDock method (version 4.0) for predicting binding sites and energies follows a hierarchical strategy for examining ligand binding conformations and calculating their binding energies. We proceeded according to the following steps:

1. Within each 10 Å cube containing regions with sufficient voids to accommodate a ligand (see section below on generating the void regions), we carry out a coarse grain docking procedure to generate a set of trial conformations for ligand binding in the receptor. We use Dock 4.0 (Ewing and Kuntz 1997) to generate and score 1000 conformations, of which 50 are selected for further analysis. This selection is based on the number of hydrogen bonds between ligand and protein, a buried surface area cutoff of 90%, and the energy scoring from Dock 4.0. In Dock 4.0, we use flexible ligand docking and allow for 10 bumps. Since the protein structure was predicted without ligand, we expect the ligand binding site to be contracted. Consequently, in the docking, we reduce the DREIDING van der Waals radii of the ligand atoms by 25%.
2. The 50 best conformations selected for each ligand from step 1 are minimized while keeping protein coordinates fixed. The

solvation energy of each of these 50 minimized conformations is calculated using the Analytical Volume Generalized Born (AVGB) continuum solvation method (Zamanakos 2001). The binding energies (BE) are calculated using

$$BE = PE\ (\text{ligand in solvent}) - PE\ (\text{ligand in protein}) \quad (1)$$

where PE is the potential energy calculated using the methods of Trabanino et al. (2004). From these 50 conformations, the five best-scoring conformations based on binding energies are selected for further optimization.

3. For each of the five structures selected in step 2, we optimize the structure of the receptor/ligand complex allowing all atoms of the receptor and ligand to relax. Using these optimized structures, we calculate the binding energy as the difference between the energy of the ligand in the protein and the energy of the ligand in water. The energy of the ligand in water is calculated using the DREIDING FF and the AVGB continuum solvation method (Zamanakos 2001).
4. From these five optimized ligand-protein complexes, we select the conformation with the most favorable binding energy. We then optimize this structure using the SCREAM side chain replacement program (V. Kam, N. Vaidehi, and W.A. Goddard III, unpubl.) to reassign all side chains for the residues within 4 Å of the ligand. SCREAM utilizes a side-chain rotamer library consisting of 1478 rotamers, each of which differs by at least 1.0 Å CRMS from the others, and uses the DREIDING all-atom energy function to evaluate the energy for the ligand-protein complex. The binding energy of each optimized complex is calculated as above.

Scanning the entire receptor to locate the binding region. To locate the binding site we scanned the entire receptor structure, making no assumptions about the nature or location of the site. To do this the molecular surfaces of the predicted structures for mOR912-93 and hOR912-93 were each mapped using the SPHGEN program (Kuntz et al. 1982) in Dock 4.0 to obtain spheres representing void regions in the receptor. The PASS program (Brady and Stouten 2000) was then used to determine 11 centers of potential binding regions in the receptor.

Subsequently, the receptor spheres located within 4 Å of each of the 11 centers identified by PASS were used to scan for the putative binding site by applying the first two steps of the HierDock protocol described above to each putative binding region. For this scanning procedure we used 10 conformations in step 2 instead of 50.

Optimization of the putative binding site and determination of the binding site of ketones. After determining the putative binding site, we used the SCREAM side chain placement method to optimize the side chain conformations of all the residues within 4 Å of the spheres generated in the putative binding site. After selecting the most favorable side chain orientations, we minimized the energy of the whole ligand-receptor structure (conjugate gradients to a RMS force of 0.1 kcal/mol-Å).

Determination of binding conformation and binding energy. Starting with the optimized ligand binding site, we used SPHGEN to regenerate the spheres in the binding region. This was necessary since the side chain optimization of the receptor changed the void regions in the receptor structure. We then performed all four steps detailed in the HierDock protocol to determine the best binding conformation for each ketone (Table 1) to mOR912-93 and hOR912-93. These calculations were done independently for mOR912-93 and hOR912-93. We calculated the binding energies for these ketone odorants using equation 1 and an internal dielectric constant of 2.5.

Since we want to compare binding energies for various odorants, we followed the Monte Carlo docking and minimization with a perturbation approach. Here the binding conformation of each ketone was perturbed to generate the corresponding binding conformation for 2-butanone by removing all carbons except those closest to the carbonyl group. The most energetically favorable conformation of 2-butanone was then perturbed to generate the corresponding binding conformation for each of the other ketones in Table 1. To perturb from 2-butanone to 2-pentanone, each of the three terminal hydrogens was replaced in turn by a methyl group, thus generating three possible conformations of 2-pentanone. Each of these three conformations was optimized using minimization in MPSim, first holding protein coordinates fixed, and then allowing both ligand and protein atoms to be moved. The final conformation selected was the conformation with the most favorable energy. The conformations for the other ketones were calculated in a similar manner. This procedure is expected to increase the likelihood that the optimal conformation is selected.

Acknowledgments

We thank the Caltech Summer Undergraduate Research Fellowship (SURF) program and the NIH for providing support for this research. The computational facilities of the MSC used in this research have been supported by grants from ARO-DURIP, ONR-DURIP, NSF (MRI, CHE), and IBM-SUR. In addition, the MSC is supported by grants from DOE ASCI, ARO-MURI, ARO-DARPA, ONR-MURI, NIH, NSF, ONR, Aventis Biopharma, General Motors, ChevronTexaco, Berlex Pharma, Seiko Epson, Beckman Institute, and Asahi Kasei.

Electronic supplemental material

The supplementary material contains Table S1, a Clustal sequence alignment of mOR912-93, hOR912-93, and nine other mouse ORs sharing sequences identities from 49% to 95% with mOR912-93. It also contains Figure S1, a plot of the average hydrophobicity of the amino acids on mOR912-93 versus the position of the residues.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Amoore, J.E. 1952. The stereochemical specificities of human olfactory receptors. *Perfumery and Essential Oil Record* **43:** 321–330.

Araneda, R.C., Kini, A.D., and Firestein, S. 2000. The molecular receptive range of an odorant receptor. *Nat. Neurosci.* **3:** 1248–1255.

Bower, M., Cohen, F.E., and Dunbrack Jr., R.L. 1997. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. *J. Mol. Biol.* **267:** 1268–1282.

Bozza, T., Feinstein, P., Zheng, C., and Mombaerts, P. 2002. Odorant receptor expression defines functional units in the mouse olfactory system. *J. Neurosci.* **22:** 3033–3043.

Brady Jr., P.G. and Stouten, P.F.W. 2000. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Design* **14:** 383–401.

Buck, L. and Axel, R. 1991. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* **65:** 175–187.

Datta, D., Vaidehi, N., Xu, X., and Goddard III, W.A. 2002. Mechanism for antibody catalysis of the oxidation of water by singlet dioxygen. *Proc. Natl. Acad. Sci.* **99:** 2636–2641.

Datta, D., Vaidehi, N., Floriano, W.B., Kim, K.S., Prasadarao, N.V., and Goddard III, W.A. 2003. Interaction of *E. coli* outer-membrane protein A with

sugars on the receptors of the brain microvascular endothelial cells. *Proteins* **50:** 213–221.

Ding, H.Q., Karasawa, N., and Goddard III, W.A. 1992. Atomic level simulations on a million particles: The cell multipole method for Coulomb and London nonbond interactions. *J. Chem. Phys.* **97:** 4309–4315.

Eisenberg, D., Weiss, R.M., and Terwilliger, T.C. 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci.* **8:** 140–144.

Ewing, T.A. and Kuntz, I.D. 1997. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **18:** 1175–1189.

Floriano, W.B., Vaidehi, N., Goddard III, W.A., Singer, M.S., and Shepherd, G.M. 2000. Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor. *Proc. Natl. Acad. Sci.* **97:** 10712–10716.

Floriano, W.B., Vaidehi, N., and Goddard III, W.A. 2004a. Making sense of olfaction through molecular structure and function prediction of olfactory receptors. *Chem. Senses* **29:** 269–290.

Floriano, W.B., Vaidehi, N., Zamanakos, G., and Goddard III, W.A. 2004b. Virtual ligand screening of large molecule databases using a hierarchical docking protocol (HierVLS). *J. Med. Chem.* **47:** 56–71.

Freddolino, P.L., Kalani, M.Y., Vaidehi, N., Floriano, W.B., Hall, S.E., Trabanino, R.J., Kam, V., and Goddard III, W.A. 2004. 3D structure for human β2 adrenergic receptor and the binding site for agonists and antagonist. *Proc. Natl. Acad. Sci.* **101:** 2736–2741.

Gaillard, I., Rouquier, S., Pin, J., Mollard, P., Richard, S., Barnabé, C., Demaille, J., and Giorgi, D. 2002. A single olfactory receptor specifically binds a set of odorant molecules. *Eur. J. Neurosci.* **15:** 409–418.

Gasteiger, J. and Marsili, M. 1980. Iterative partial equalization of orbiatal electronegativity—a rapid access to atomic charges. *Tetrahedron* **36:** 3219–3228.

Godfrey, P.A., Malnic, B., and Buck, L.B. 2004. The mouse olfactory receptor gene family. *Proc. Natl. Acad. Sci.* **101:** 2156–2161.

Hall, S.E., Floriano, W.B., Vaidehi, N., and Goddard III, W.A. 2004. 3D structures for mouse I7 and rat I7 olfactory receptors from theory and odor recognition profiles from theory and experiment. *Chem. Senses* **29:** 595–616.

Jain, A., Vaidehi, N., and Rodriguez, G. 1993. A fast recursive algorithm for molecular dynamics simulation. *J. Comp. Phys.* **106:** 258–268.

John, B. and Sali, A. 2003. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Res.* **31:** 3982–3992.

Kajiya, K., Inaki, K., Tanaka, M., Haga, T., Kataoka, H., and Touhara, K. 2001. Molecular bases of odor discrimination: Reconstitution of olfactory receptors that recognize overlapping sets of odorants. *J. Neurosci.* **21:** 6018–6025.

Kalani, M.Y., Vaidehi, N., Freddolino, P.L., Floriano, W.B., Hall, S.E., Trabanino, R.J., Kam, V., and Goddard III, W.A. 2004. Structure and function of human D_{2L} receptor. *Proc. Natl. Acad. Sci.* **101:** 3815–3820.

Kekens-Huskey, P.M., Vaidehi, N., Floriano, W.B., and Goddard III, W.A. 2003. Fidelity of phenyl alanyl tRNA synthetase. *J. Phys. Chem.* **107:** 11549–11557.

Keller, A. and Vosshall, L.B. 2004. A psychophysical test of the vibration theory of olfaction. *Nat. Neurosci.* **4:** 337–338.

Krautwurst, D., Yau, K.W., and Reed, R.R. 1998. Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell* **95:** 917–926.

Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., and Ferrin, T.E. 1982. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161:** 269–288.

Lim, K.T., Brunett, S., Iotov, M., McClurg, R.B., Vaidehi, N., Dasgupta, S., Taylor, S., and Goddard III, W.A. 1997. Molecular dynamics for very large systems on massively parallel computers: The MPSim program. *J. Comput. Chem.* **18:** 501–521.

MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem.* **102:** 3586–3616.

Malnic, B., Hirono, J., Sato, T., and Buck, L. 1999. Combinatorial receptor codes for odors. *Cell* **96:** 713–723.

Malnic, B., Godfrey, P.A., and Buck, L.B. 2004. The human olfactory receptor gene family. *Proc. Natl. Acad. Sci.* **101:** 2584–2589.

Man, O., Gilad, Y., and Lancet, D. 2004. Prediction of the odorant binding site of olfactory receptors by human-mouse comparisons. *Protein Sci.* **13:** 240–254.

Mathiowetz, A.M., Jain, A., Karasawa, N., and Goddard III, W.A. 1994. Protein simulations using techniques suitable for very large systems: The cell multipole method for nonbond interactions and the Newton-Euler inverse mass operator method for internal coordinate dynamics. *Proteins* **20:** 227–247.

Mayo, S.L., Olafson, B.D., and Goddard III, W.A. 1990. DREIDING: A generic forcefield for molecular simulations. *J. Phys. Chem.* **94:** 8897–8909.

Mombaerts, P. 1999. Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* **286:** 707–711.

Rouquier, S., Friedman, C., Delettre, C., van den Engh, G., Blancher, A., Crouau-Roy, B., Trask, B., and Giorgi, D. 1998. A gene recently inactivated in human defines a new olfactory receptor family in mammals. *Hum. Mol. Genet.* **7:** 1337–1345.

Rouquier, S., Stubbs, L., Gaillard-Sanchez, I., and Giorgi, D. 1999. Sequence and chromosomal localization of the mouse ortholog of the human olfactory receptor gene 912–913. *Mamm. Genome* **10:** 1172–1174.

Singer, M.S. 2000. Analysis of the molecular basis for octanal interactions in the expressed rat I7 olfactory receptor. *Chem. Senses* **25:** 155–165.

Schertler, G.F.X. 1998. Structure of rhodopsin. *Eye* **12:** 504–510.

Trabanino, R.J., Hall, S.E., Vaidehi, N., Floriano, W.B., and Goddard III, W.A. 2004. First principle predictions of the structure and function of G-protein coupled receptors: Validation for bovine rhodopsin. *Biophys. J.* **86:** 1904–1921.

Turin, L. 1996. A spectroscopic mechanism for primary olfactory reception. *Chem. Senses* **21:** 773–791.

Vaidehi, N., Jain, A., and Goddard III, W.A. 1996. Constant temperature constrained molecular dynamics: The Newton-Euler inverse mass operator method. *J. Phys. Chem.* **100:** 10508–10517.

Vaidehi, N., Floriano, W.B., Trabanino, R., Hall, S.E., Freddolino, P.L., Choi, E., Zamanakos, G., and Goddard III, W.A. 2002. Prediction of structure and function of G protein-coupled receptors. *Proc. Natl. Acad. Sci.* **99:** 12622–12627.

Wang, P., Vaidehi, N., Tirrell, D.A., and Goddard III, W.A. 2002. Virtual screening for binding of phenylalanine analogs to phenylalanyl-tRNA synthetase. *J. Am. Chem. Soc.* **124:** 14442–14449.

Zamanakos, G. 2001. "A fast and accurate analytic method for the computation of solvent effects in molecular simulations." Ph.D. thesis, California Institute of Technology, Pasadena, CA.

Zhang, X. and Firestein, S. 2002. The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* **5:** 124–133.

Zhao, H.Q., Ivic, L., Otaki, J.M., Hashimoto, M., Mikoshiba, K., and Firestein, S. 1998. Functional expression of a mammalian odorant receptor. *Science* **279:** 237–242.