Chem. Senses 29: 269–290, 2004 DOI: 10.1093/chemse/bjh030

# Making Sense of Olfaction through Predictions of the 3-D Structure and Function of Olfactory Receptors

**Wely B. Floriano, Nagarajan Vaidehi and William A. Goddard III**

Materials and Process Simulation Center, MC 139-74, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

*Correspondence to be sent to: William A. Goddard III, Materials and Process Simulation Center, MC 139-74, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. e-mail: wag@wag.caltech.edu*

## Abstract

We used the MembStruk first principles computational technique to predict the three-dimensional (3-D) structure of six mouse olfactory receptors (S6, S18, S19, S25, S46 and S50) for which experimental odorant recognition profiles are available for a set of 24 odorants (4–9 carbons aliphatic alcohols, acids, bromo-acids and diacids). We used the HierDock method to scan each predicted OR structure for potential odorant binding site(s) and to calculate binding energies of each odorant in these binding sites. The calculated binding affinity profiles are in good agreement with experimental activation profiles, validating the predicted 3-D structures and the predicted binding sites. For each of the six ORs, the binding site is located between transmembrane domains (TMs) 3–6, with contributions from extracellular loops 2 and 3. In particular, we find six residue positions in TM3 and TM6 to be consistently involved in the binding modes of the odorants. Indeed, the differences in the experimental recognition profiles can be explained on the basis of these critical residues alone. These predictions are also consistent with mutation data on ligand binding for catecholamine receptors and sequence hypervariability studies for ORs. Based on this analysis, we defined amino acid patterns associated with the recognition of short aliphatic alcohols and mono-acids. Using these two sequence fingerprints to probe the alignment of 869 OR sequences from the mouse genome, we identified 34 OR sequences matching the fingerprint for aliphatic mono-acids and 36 corresponding to the recognition pattern for aliphatic alcohols. We suggest that these two sets of ORs might function as basic arrays for uniquely recognizing aliphatic alcohols and acids. We screened a library of 89 additional molecules against the six ORs and found that this set of ORs is likely to respond to aldehydes and esters with longer carbon chains than their currently known agonists. We also find that compounds associated with the flavor in foods are often among the best calculated binding affinities. This suggests that physiologic ligands for these ORs may be found among aldehydes and esters associated with flavor.

**Key words**: biosensors, docking, function prediction, GPCR, G protein coupled receptor, HierDock, mice, olfaction, olfactory receptor, structure prediction, taste, virtual screening



## Introduction

Recent advances in the study of the olfactory system have provided important information about the mechanisms underlying odor recognition. We now know that the early stage in odorant detection involves binding of odor molecules to olfactory receptors (ORs), which are expressed in olfactory sensory neurons (OSNs) in the nose (Buck and Axel, 1991; Lancet and Ben-Airie, 1993). Each OSN expresses only one OR type but a particular OR can respond to multiple odorants, while a particular ligand can elicit response from multiple ORs, leading to unique combinations of ORs for each odorant (Krautwurst *et al.*, 1998; Malnic *et al.*, 1999). ORs are members of the rhodopsin-like class of G protein coupled receptors (GPCR), which also includes catecholamine receptors, and their three-dimensional structure contains the seven helical transmembrane (TM) motif characteristic of GPCRs (Mombaerts *et al.*, 1996). The ORs share sequence fingerprints not seen in other GPCRs and their higher sequence variability is consistent with the ability to recognize structurally diverse odorants (Buck and Axel, 1991). Each OR interacts with a variety of structurally diverse odor molecules, in contrast to other members of the GPCR family such as the receptors for dopamine, histamine and serotonin, which are narrowly tuned to one or a few closely related agonists (Bozza and Kauer, 1998; Krautwurst *et al.*, 1998; Zhao *et al.*, 1998; Duchamp-Viret *et al.*, 1999; Malnic *et al.*, 1999; Mori *et al.*, 1999; Rubin and Katz, 1999).

Bioinformatic analyses of genome sequence databases indicate that mice have ~1200 different ORs (Zhang and Firestein, 2002), while humans have ~350 (Glusman *et al.*, 2000; Zozulya *et al.*, 2001). The combinatorial nature of odor recognition explains how humans can detect and distinguish an immense number and variety of odorants using only 350 distinct ORs. 'Families' of ORs, whose members have >40% protein sequence identity appear largely to be common to both species (Zhang and Firestein, 2002).

In order to learn more about the function of GPCRs, we developed the MembStruk computational technique to predict the three-dimensional (3-D) structure of these proteins and we developed the HierDock computational technique to predict the binding site and binding energy of ligands to GPCRs (Floriano *et al.*, 2000; Vaidehi *et al.*, 2002). These methods have previously been applied to predict the binding site for epinephrine to adrenergic receptor (Vaidehi *et al.*, 2002), for dopamine to dopamine receptor (Vaidehi *et al.*, 2002) and for alcohols to the S25 mouse OR (Floriano *et al.*, 2000). Particularly for epinephrine and dopamine, detailed mutation and binding studies provide strong evidence for the accuracy of the predicted 3-D structure.

In an earlier work (Floriano *et al.*, 2000), we used computational modeling to build a 3-D structure for OR S25 shown experimentally by Malnic *et al.* (1999) to bind hexanol and heptanol. The calculated binding affinity profile for a C4–C9 linear alcohol series suggests that it is the pattern of structural features rather than the specific molecular structure that triggers recognition. We found that the binding site of hexanol and heptanol in OR S25 involves the helical domains TM3, TM4, TM5 and TM6, which is consistent with earlier computational work on rat I7 receptor (Singer, 2000) and with mutation studies for binding of dopamine and epinephrine in biogenic amine receptors (Strader *et al.*, 1994; Shi and Javitch, 2002; Bissantz *et al.*, 2003). Residues we found to be involved in binding in S25 (Floriano *et al.*, 2000) were previously proposed on the basis of high variability among ORs to be involved in odorant binding (Malnic *et al.*, 1999; Pilpel and Lancet, 1999).

In this paper we report the predicted structures of six ORs studied by Malnic *et al.* (1999). We also report the predicted binding sites of the 24 aliphatic odorants they studied experimentally and calculated their relative binding energies to each OR. We find quite good agreement between the predicted binding profiles and the experimental activation profiles.

We also screened a library of 89 additional molecules against the six ORs and found that compounds associated with the flavor in mouse foods appear often among the best binders. This suggests that ORs function partly to identify nutrients-rich foods.

## Materials and methods

We built 3-D structures for six mice (S6, S18, S19, S25, S46 and S50) olfactory receptors (ORs) studied experimentally by Malnic *et al.* (1999) who measured recognition profiles for a systematic library of 24 aliphatic alcohols and acids (each with 4–9 carbons). Of the 14 ORs studied by Malnic *et al.* (1999), we selected six with complete amino acid sequences.

After predicting each OR structure, we used the HierDock procedure to first scan the receptors for potential binding site(s) and then to dock the ligands into the putative binding site and predict their binding energies. The steps of 3-D structure prediction, scanning for possible binding sites and binding of ligands were performed independently to each OR, with no assumptions or constraints being carried from one OR to another. We compared the predicted affinity profiles for the set of 24 odorants to the known experimental response and found good correlation. We then screened a library of 89 additional molecules against the six ORs.

### Ligand/odorant structure building

We considered the two libraries of ligands shown in Table 1. List A consists of the 24 aliphatic acids, alcohols, diacids and bromoacids (each with 4–9 carbons) tested experimentally by Malnic *et al.* (1999). List B contains an additional 89 molecules including common odorants, molecules that are odorless to humans, tastants, glucose and some molecules of potential biodefense interest. We included tastants in this list to examine how tastants might affect ORs. Comparing the binding affinities of such tastants as sugars (which are natural agonists of the functionally related taste receptors, also members of the GPCR family) against ORs could provide insights into how function diversity evolved among the chemosensory receptors and more generally in the GPCR family.

Three-dimensional structures for all molecules in Table 1 were built as follows.

1. Chemical structures were drawn using the program ISISDRAW (MDL Information Systems Inc., 2001) and saved in the two-dimensional (2-D) 'mol' format.
2. The program Stereoplex 1.2 (Tripos Inc., 2001) was used to generate all stereoisomers, whenever applicable.
3. We used the 2-D structures from (1) and (2) and the program Concord version 1.2 (Tripos Inc. 2001) to generate 3-D structures, add hydrogens and assign Gasteiger atomic charges to each molecule. Acids, diacids and bromoacids were assigned a formal charge of –1.0 to each carboxyl group.
4. The starting structures from (3) were optimized by conjugate gradient minimization of the potential energy using the Dreiding force field (Mayo *et al.*, 1990) with Gasteiger (Gasteiger and Marsili, 1980) charges. The structures were minimized (to an RMS force of 0.2 kcal/mol/Å) and used as starting conformations in HierDock

**Table 1** Ligands for which we have calculated binding sites and binding energies to the six OR's in the present study

| A. Known experimental response | B. Common odorants and others | | |
|---|---|---|---|
| Butanol | 1-Decanol | Furaneol | Undecanal |
| Pentanol | 2,3,5,6-Tetramethylpyrazine | Furaneol glucoside | Valeraldehyde |
| Hexanol | 2,3,5-Trimethylpyrazine | Geraniol | Valerophenone |
| Heptanol | 2,3-Dimethylpyrazine | Heptanal | 2-Aminoacetophenone |
| Octanol | 2,4-Dimethylacetophenone | Hexanal | Citric acid |
| Nonanol | 2,5-Dimethylpyrazine | Hexyl octanoate | Pyruvic acid |
| Butyric acid | 2-Ethyl-3,5(6)-dimethylpyrazine | Isoamyl nonanoate | Piperidine |
| Valeric acid | 2-Ethyl-3-methoxypyrazine | Isobutyric acid | Piperine |
| Hexanoic acid | 2-Ethyl-3-methylpyrazine | Isopropyl hexanoate | 1-Propyldisulfanyl-propane |
| Heptanoic acid | 2-Ethyl-5(6)-methylpyrazine | Isopulegol | Skatole |
| Octanoic acid | 2-Ethylpyrazine | Lilial | Hexahydro-1, 3, 5-trinitro-1, 3, 5-triazine (rdx) |
| Nonanoic acid | 2-Hexanone | (+)-Linalool | Octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine (hmx) |
| Succinic acid | 2-Isobutyl-3-methoxypyrazine | Citral | Dichloro(trans-2-chlorovinyl)arsine (Lewisite) |
| Glutaric acid | 2-Isopropyl-3-methoxypyrazine | Citronellal | 2,2′-Dichlorodiethyl sulfide (mustard gas) |
| Adipic acid | Methyl-2-ethylbutyrate | Decanal | S-(+)-Carvone |
| Pimelic acid | 2-Methoxy-3-methylpyrazine | Dodecanal | *Trans*-cinnamaldehyde |
| Suberic acid | 2-Methoxypyrazine | Ethyl acetoacetate | Glucose |
| Azelaic acid | 2-Methylpyrazine | Ethyl butyrate | 3-Methyl-1-butane-1-thiol |
| 4-Bromobutyric acid | Methyl-2-methylbutyrate | Ethyl isobutyrate | Pyridine |
| 5-Bromovaleric acid | 2-Octanone | (–)-Linalool | Tridecanal |
| 6-Bromohexanoic acid | 2-Phenylethanol | Lyral | b-Terpineol |
| 7-Bromoheptanoic acid | 2-Sec-butyl-3-methoxypyrazine | (D,L)-Menthone | Butyraldehyde |
| 8-Bromooctanoic acid | 2-Sec-butylcyclohexanone | (D,D)-Menthone | Butyrophenone |
| 9-Bromononanoic acid | Ethyl vanillin | (L,L)-Menthone | Eucalyptol |
| | 3-Heptanol | (L,D)-Menthone | Ethyl salicylate |
| | Iso-amyl alcohol | Mesifurane | Eugenol |
| | 3-Phenyl-1-propanol | Methyl isobutyrate | Beta-ionone |
| | Estragole | Methyl nonanoate | Salicylic acid |
| | 5-Hydroxy-tetradecanoic acid | Methyl octanoate | |
| | Ethyl isoamyl ketone | Nonanal | |
| | Acetoin | Octanal | |
| | Acetone | Phenylethylamine | |
| | Acetophenone | Propiophenone | |
| | Benzaldehyde | R-(–)-carvone | |

Column 1 (hereafter referred to as Table 1A) is the list of 24 odorants for which Malnic *et al*. (1999) measured experimental response. The other three columns contain common odorants and tastants plus some molecules related to bioterrorism.

for locating the putative binding site of odorants in the 3-D structure of each OR and to determine their binding configuration and energy. The solvation energies for the odorants were calculated using the analytical volume generalized born (AVGB) continuum solvation approach (Zamanakos, 2002).

### Building the 3-D structure of ORs

The 3-D structure of each ORs was predicted using the MembStruk 1.0 procedure as detailed in Floriano *et al.* (2000). The main steps in the MembStruk procedure are as follows.

1. Predict the TM regions using hydropathicity analysis (Donnelly, 1993) combined with input from multi-sequence profiles calculated using the Eisenberg hydrophobicity scale (Eisenberg *et al.*, 1984). The sequence alignment profile is obtained from the sequence analysis for the set of ORs described in Floriano *et al.* (2000). The TM regions are assigned using the hydropathicity profiles and capping rules based on known helix breaker residues (Pro, Gly and charged residues). The average hydrophobicity at each sequence position is calculated over window sizes from 12 to 20, until seven peaks can clearly be identified in the profile. The sequence alignment showing the predicted TMs as well as the residues we find to be involved in odorant binding is given in Figure 4 and will be discussed in the results section.
2. Construct canonical right-hand α-helices for the predicted TM segments and add counterions $Cl^-$ and $Na^+$ to neutralize charges in Lys, Arg, Glu and Asp residues. Optimize these individual helices using conjugate gradients energy minimization followed by 200 ps of torsional dynamics using the NEIMO (Newton–Euler inverse mass operator) torsional molecular dynamics method (Mathiowetz *et al.*, 1994; Vaidehi *et al.*, 1996) at 300K, which fixes bonds and angles. This allows the helix to distort and bend in response to helix breakers such as proline and also optimize the side-chain conformations.
3. Identify lipid-accessible residues from analysis of the periodicity of the hydrophobic residues (Donnelly, 1993) in the sequence.
4. Assemble the helix bundle as follows. The helical axes were oriented according to the 7.5 Å electron density map of frog rhodopsin (Schertler, 1998). Although there are high resolution structures available for rhodopsin (Okada *et al.*, 2002; Teller *et al.*, 2001; Palczewski *et al.*, 2000), we used this low resolution one because we only wanted the orientations of the helical axes (not atomic coordinates) to build an initial template bundle. The helices were then rotated to orient their hydrophobic moments to the outside of the helix barrel. We then added an explicit lipid bilayer of 52 molecules of dilauroylphosphatidyl choline to simulate the membrane. This lipid–TM bundle was minimized to an RMS force of 0.2 kcal/mol/Å.
5. Further optimize the lipid-TMs bundle for 100 ps using constant temperature rigid body molecular dynamics (RBMD) at 300K.
6. Construct disulfide bridges as described below and add intracellular and extracellular connecting loops using the software WHATIF (Vriend, 1990).
7. Energy minimize the entire lipid-TM bundle using the surface generalized born (SGB; Ghosh *et al.*, 1998) continuum solvent method (we used dielectric constant of 60.0 to simulate the dielectric region surrounding the membrane).

All calculations were carried out with Dreiding force field (Mayo *et al.*, 1990) using CHARMM22 (MacKerell *et al.*, 1998) charges for the protein.

ORs have four highly conserved cysteines believed to form disulfide bridges (Floriano *et al.*, 2000). One of the conserved Cys is located in the extracellular loop 1 (EC1) and the other three are located in the extracellular loop 2 (EC2). Mutation analysis (Singer *et al.*, 1996) and truncation experiments (Gimelbrant *et al.*, 1999) suggest that the EC1 Cys is paired to the second Cys in EC2 and that the first and third Cys in EC2 are bound together. These correspond to Cys127–Cys209 and Cys199–Cys219 in the amino acid sequence of S25. Disulfide bridges in the loop region increase the confidence in the modeled loop structure since they restrict the loops conformational space. However, because of their high flexibility, loops are still the part of the structure with the least degree of confidence in our structure predictions.

There is one additional highly conserved Cys located in TM6 that is unpaired in all six predicted structures (in S25 this is Cys 241). S25 and S19 have additional disulfide bridges (S25: Cys132–Cys192 and Cys157–Cys171; S19: Cys60–Cys102, Cys92–Cys274 and Cys155–Cys176), all of which were assigned by mutation analysis (Singer *et al.*, 1996) and assumed in our calculations. S18 (Cys105–Cys187 and Cys177–Cys197) and S46 (Cys99–Cys181 and Cys171–Cys191) have only the two highly conserved disulfide bonds. S50 and S6 have the two conserved linkages (Cys101–Cys183 and Cys173–Cys193) plus an extra disulfide bond between Cys116 and Cys289. S6 and S50 have three Cys (Cys76, Cys116, and Cys289) located about half-way into the barrel on TMs 2, 3 and 7 that could potentially pair in any combination, consequently pulling two of the three helices closer to each other. Our predicted structures for S50 and S6 place the pair Cys116–Cys289 close enough to form an SS bond in the final structures and hence we made this bond.

The 3-D structure for S25 used in our previous publication (Floriano *et al.*, 2000) was improved by correcting some chirality problems not detected in the original structure. The binding profile of the odorants for S25 was re-calculated using the HierDock protocol reported here, which is considerably improved from the previous procedure (Floriano *et al.*, 2000).

### Predicting function for the ORs

Since there is no experimental information on the location of the binding site of the odorants, we used ScanBindSite (a variation of HierDock) to scan the entire receptor structure to locate the binding site of each odorant to each OR. No assumption was made on the nature or the location of the binding site in these receptors. Once the binding site was determined, we applied HierDock to this putative binding site to predict the best binding configuration and energy for each odorant in this site.

#### *Scanning for a putative binding site (ScanBindSite)*

Scanning of each OR for an energetically favorable binding region was performed as follows.

We partitioned the empty volume suitable for docking in each OR structure into 4–13 (depending on the size of the empty volume) small overlapping regions. A modified version of HierDock (called ScanBindSite) was used to independently dock the 24 ligands into each region. Rigid docking of each ligand into each docking region was performed using the program Dock 4.0 (Ewing and Kuntz, 1997). The maximum number of docked orientations was set to 100. Each configuration was torsion minimized. We used the option for bump filter (four bumps), energy scoring and automated matching with no anchor search in the Dock 4.0 code.

The configurations with the 10 best Dock 4.0 scores were selected for each ligand in each region. Each of these configurations (240 per region) was then energy minimized for 80 steps while keeping the protein coordinates fixed. We then eliminated any configuration whose molecular surface was <70% buried into the receptor and selected the lowest energy configuration for each ligand in each region. For each OR there was one region where most of the selected configurations cluster together. This region was then selected as the binding site. No presumption was made on the nature or the location of the binding site in these receptors.

#### *Calculating binding energy profiles using the HierDock protocol*

We used the HierDock protocol to predict the structures of the odorant/OR bound complexes and to estimate their binding affinities. HierDock has been validated for other ORs (Floriano *et al.*, 2000), other GPCRs (Vaidehi *et al.*, 2002) and for globular proteins (Wang *et al.*, 2002; Datta *et al.*, 2003; Kekenes-Huskey *et al.*, 2003). The main steps of the HierDock protocol are as follows.

1. *Define docking region*. We run HierDock twice for each OR. In the first pass, the docking region for each target protein is defined as the location found in the ScanBindSite step. For the second HierDock pass, the docking region is redefined using the best known positive from the first pass. The new binding region is defined as the 1 Å around the selected known positive and all the ligands are re-docked into this narrower binding region.
2. *Protein grid calculation*. The docking step uses an energy grid for the protein contribution to the interaction energy. This grid is calculated only once per target, using the program Grid provided as part of the Dock 4.0 (Ewing and Kuntz, 1997) package.
3. *Level 0*. Generation of docked configurations of each ligand into the binding site of the target protein using the program Dock 4.0 (Ewing and Kuntz, 1997). We used the options for flexible docking, torsion drive on, energy scoring, 75% reduced van der Waals parameters for ligands, minimization before scoring and 1000 maximum scored configurations. The 500 best configurations by Dock 4.0 scores for each ligand in the library were saved and used in subsequent steps.
4. *Filter*. A combined criterion of differential solvation and level 0 scores (Dock 4.0 energies) was applied to the 500 configurations per ligand from level 0. We eliminated configurations with <75% of buried surface and those having energy >100 kcal/mol. The best 50 configurations per ligand were carried to next step.
5. *Level 1*. For each of the 50 configurations per ligand surviving filter step 4, we used the Dreiding FF with CHARMM22 charges for protein and Gasteiger charges for ligand to minimize the structure of each ligand (gas phase) for (100 steps) using fixed protein coordinates. The best five configurations by energy were selected for each ligand.
6. *Level 2*. For each of the five configurations per ligand from level 1, we used the Dreiding FF to minimize the structure of the ligand–protein complex for 100 steps, allowing all atoms of the protein to move. The best configuration by energy was selected for each ligand.
7. *Ranking*. The complexes from level 2 were scored by single point energy calculation including solvation using the AVGB (Zamanakos, 2002) continuum solvation method with fixed protein (this gives the energy of the ligand in the field of the protein) using a dielectric constant of 2.5. In addition, we calculated the energy including solvation of the free ligand in its docked conformation. The dielectric constant used for non-bonded interactions was 1.0, while the dielectric constant for the continuum solvation calculation was set to 78.2 for exterior regions and 1.3 for the interior region. The binding energies were calculated as the difference in the energies of the ligand in protein and ligand in water [BindE = E(bound_ligand_in_protein) – E(free_ligand_in_water)]. The ligand list was then grouped according to the main chemical functional group and sorted by binding energies.

### *Refining the binding site location and the binding affinities*

Each of the six ORs studied here has one or more experimentally known agonist ligand (Malnic *et al.*, 1999), that is, an odorant that causes response from cells found to express that OR. Hence we used the putative binding mode of the known odorants having the most favorable binding energy in a first HierDock run to refine the location of the binding site in each OR. Note that this procedure does not change the general location of the binding region. This only reduces its size, making the conformational sampling step more efficient. For the first HierDock run on each OR, we used the binding site from ScanBindSite (described above). We then used the bound configuration of the best (by binding energy) known ligand to define a smaller docking region surrounding the ligand with a margin of 1 Å for a second HierDock run. All ligands, including the ones used for refining the binding regions, were re-docked and re-scored in the new binding region using steps 2–7 described above. This procedure ensures that we calculate binding affinities for ligand configurations that are in similar binding modes. It also allows for a better sampling of the binding site. This is important because ligands that do not elicit experimental response may have better affinity for regions other than the one where binding of the ligand leads to activation of the receptor. However, the affinity at the active site is what should differentiate experimental positives from negatives.

Structural analysis of the complexes generated in the first HierDock pass for S18, S19 and S46 ORs, pointed to a strong interaction of the carboxyl groups in the ligands with histidine in the receptors. These histidines (His113/S18, His133/S19 and His107/S46) were treated as charge neutral during ScanBindSite and for the first HierDock pass. However, since they were found to interact with carboxyl groups in the ligands, we assumed that the ring nitrogens will be protonated and assigned a formal charge of +1.0 to these histidines, which was used for the second HierDock pass. These histidines are conserved in S18, S19 and S46 (all three recognize acids and bromo acids, but not diacids), mutated to valine in S25 (which does not recognize acids, bromo acids, or diacids) and mutated to tyrosine in S6 and S50 (both recognizes diacids, but not acids or bromo acids). The binding energies of the ligands calculated using the refined receptor structure for the second HierDock run have been used for binding energy comparisons across ligands and ORs.

### *Using perturbation to further enhance the conformational search*

The differences in binding energies for members of the same chemical class are often in the order of a few kilocalories. Since small differences in the binding mode of a ligand can cause differences in binding energy of similar magnitude, we need to ensure that the ligands being compared are scored in equivalent binding modes. To do this we start with the best binding conformation of the most energetically favorable ligand of a chemical class and generate all other ligands in the class (differing only in the carbon chain length) by adding or removing methyl groups. We then minimize (with movable protein) and score the perturbed complexes as described in level 2 of HierDock (steps 6 and 7, described above). This guarantees that different ligands in the same chemical class are scored consistently so that their relative binding energies can be used as a measurement of relative biological affinities.

This perturbation procedure was used when members of a chemical class were found to have different binding modes after HierDock. However, it was not used when the members of a chemical class were found by HierDock to have equivalent binding modes. The need for perturbation comes from the non-exhaustive conformational sampling in level 0 of HierDock.

Among the six receptors reported here, this perturbation approach was found to be necessary only for the alcohol series bound to S25. In this case, all other members of the alcohol series were perturbed from the bound configuration of hexanol in S25. For the other five receptors, we found that HierDock led to equivalent binding modes for all members of each chemical class and therefore perturbations were not required.

## Results and discussion

In this section we discuss the predicted 3-D structures of the six mouse ORs S6, S18, S19, S25 and S50 bound to their known and potential agonists. The TrEMBL identification numbers for these OR sequences are given in Table 6.

### Comparison of the predicted 3-D structure of the six ORs

The calculated percentage of sequence similarity and the corresponding root mean square deviation in (carbon alpha) coordinates (CRMS) difference between predicted structures for the ORs studied here are given in Table 2. The sequence comparison considered all amino acids in each aligned OR, while the structure comparison included only the TM regions. Members of the same protein family with high sequence similarity are expected to have a high degree of structural homology. For the family of six ORs studied here, sequence similarity ranges from 20 to 94%. The CRMS differences between the TM regions of the predicted OR structures range from 1.5 Å (between S6 and S50 with a sequence similarity of 94%) to 6.3 Å (between S19 and S25 with a sequence similarity of 25%; as discussed below, the bottom part of S19 had a distorted chain that led to larger CRMS errors).

The sequence similarity between these six ORs and bovine rhodopsin (pdb code 1F88) ranges from 11% (S6) to 18% (S46) leading to CRMS deviations of 4.0 Å (S46) to 7.3 Å (S19), with five of the six receptors in the 4.0–4.4 Å range.

Our predicted structure for OR S19 differs dramatically from the other structures, with a CRMS of 6.3 Å. This arises

**Table 2** Sequence and structural relationship among the six mouse olfactory receptors studied here

| OR | Targeted chemical class | Percentage homology | S6 | S50 | S19 | S18 | S46 | S25 |
|---|---|---|---|---|---|---|---|---|
| S6 | Diacids | Seq. | 100 | 94 | 30 | 35 | 34 | 23 |
| | | Struct. | 0.0 | 1.5 | 5.7 | 2.4 | 1.8 | 3.0 |
| S50 | Diacids | Seq. | 94 | 100 | 32 | 35 | 34 | 24 |
| | | Struct. | 1.5 | 0.0 | 5.8 | 2.6 | 2.3 | 3.1 |
| S19 | Acid, bromo-acid, alcohol | Seq. | 30 | 32 | 100 | 48 | 45 | 25 |
| | | Struct. | 5.7 | 5.8 | 0.0 | 6.1 | 5.6 | 6.3 |
| S18 | Acid, bromo-acid, alcohol | Seq. | 35 | 35 | 48 | 100 | 52 | 27 |
| | | Struct. | 2.4 | 2.6 | 6.1 | 0.0 | 2.0 | 3.2 |
| S46 | Acid, bromo-acid | Seq. | 34 | 34 | 45 | 52 | 100 | 27 |
| | | Struct. | 1.8 | 2.3 | 5.6 | 2.0 | 0.0 | 2.8 |
| S25 | Alcohol | Seq. | 23 | 24 | 25 | 27 | 27 | 100 |
| | | Struct. | 3.0 | 3.1 | 6.3 | 3.2 | 2.8 | 0.0 |
| Rhodopsin | Aldehyde | Seq. | 11 | 12 | 15 | 13 | 18 | 12 |
| | | Struct. | 4.2 | 4.4 | 7.3 | 4.4 | 4.0 | 4.3 |

The percentage sequence similarity is given in each row followed by CRMS difference (Å) of the Cα atoms in the TM regions. Both provide a measure of the structural similarities between the various predicted OR structures. The CRMS was calculated for the 119 structurally equivalent residues in the superimposed OR structures.

from TM 4 in S19, which is unraveled in the bottom one-third (C terminus). This probably resulted from an early bad contact in the helix optimization steps, but we did not go back to re-examine it because the intracellular C terminus is not expected to be involved in direct binding of the ligand.

The CRMS differences found for the six ORs can be compared with that found in sets of homologous proteins including globins, serine proteases and immunoglobulins (Chothia and Lesk, 1986). Sequence identities in the 20–50% range translate to structural differences between 1 and 2 Å. This agrees with the accuracy found for structures of globular proteins predicted using homology methods for cases with 50% sequence similarity (Marti-Renom *et al.*, 2002).

The MembStruk structure prediction algorithm for GPCRs uses only sequence information combined to physical principles to predict the 3-D structure. This leads to independent unbiased predictions for each of the six ORs. The advantage of this approach is that mistakes made on a model will not pass to models generated from it and the problems that come with aligning low homology sequences are avoided. However, this brings the disadvantage that structures of the same family may differ from each other more that they should because of the statistical nature of our approach. We plan to apply homology methods to predict the other five ORs from each of our predicted structures. Such methods might be used to obtain several predicted structures that could be averaged in some way to provide more accurate structures than can be derived solely from independent MembStruk runs.

### Predicted odorant recognition profiles for the six ORs

Table 3 and Figure 1a–f show the binding energies predicted for the series of 24 odorants for which odorant activation profiles were determined experimentally (Malnic *et al.*, 1999). It is important to note that the calculated binding energy is a necessary but not sufficient condition for activation of the receptor. Some odorants may bind to the receptor but may not activate it. In Figure 1a–f odorants that activate experimentally cells from where the ORs were isolated are fully shaded. Odorants not tested experimentally against cells expressing a particular OR are half shaded. Odorants tested and found not to cause activation are not shaded. Each odorant family is ordered with C4 at the left and C9 at the right.

The calculated recognition profiles for S50 (Figure 1a) and S6 (Figure 1b) clearly identify aliphatic diacids as the best odorant binding class for both receptors, which is amply confirmed by the experiments. Within the diacids class, the ligand with the most favorable calculated binding affinity for S50, azelaic acid (C9 diacid), is the only odorant observed to elicit response experimentally out of the 24 odorants. The predicted odorant binding energy profile for S6 (Figure 1b) is also in good agreement with the experimental profile in terms of the chemical class identified by this OR. Also, the predicted best ligand (by binding energy), azelaic acid (C9 diacid), is one of the two odorant agonists observed experimentally. However, the theory predicts that the second experimental positive (the C8 diacid suberic acid) is the third highest (more negative) binding energy while the

**Table 3** Calculated (HierDock) binding energies (kcal/mol) of the 24 odorants in Table 1for the six mouse olfactory receptors

| Class | Ligand | S6 | S50 | S18 His113+ | S19 His133+ | S46 His107+ | s25 |
|---|---|---|---|---|---|---|---|
| Alcohols | Butanol (C4) | –29.78[a] | –16.79[a] | –15.94[a] | –20.22[a] | –14.04[a] | –29.78[a] |
| | Pentanol (C5) | –20.00 | –17.84 | –23.96 | –20.96 | –16.09 | –28.80 |
| | Hexanol (C6) | –30.29 | –18.53 | –27.75 | –25.41 | –22.10 | –30.42 |
| | Heptanol (C7) | –35.04 | –26.61 | –29.97 | –26.40 | –25.78 | –31.97 |
| | Octanol (C8) | –35.86 | –27.18 | –32.32 | –29.52 | –24.60 | –28.01 |
| | Nonanol (C9) | –38.68 | –29.48 | –33.42 | –31.61 | –27.08 | –15.35 |
| Bromo acids | 4-Bromobutyric acid (C4) | –81.88 | –50.86 | –37.40 | –9.99 | –11.61 | 12.45 |
| | 5-Bromovaleric acid (C5) | –83.70 | –58.24 | –40.46 | –16.84 | –14.21 | 8.05 |
| | 6-Bromohexanoic acid (C6) | –84.54 | –41.62 | –39.48 | –19.04 | –16.09 | 5.21 |
| | 7-Bromoheptanoic acid (C7) | –87.81[a] | –64.55[a] | –39.47[a] | –20.11[a] | –28.00[a] | 3.14[a] |
| | 8-Bromooctanoic acid (C8) | –91.73 | –59.57 | –45.52 | –21.09 | –28.80 | 1.53 |
| | 9-Bromononanoic acid (C9) | –91.78[a] | –64.82[a] | –49.21[a] | –22.76[a] | –22.45[a] | 13.65[a] |
| Acids | Butyric acid (C4) | –79.08 | –60.14 | –36.38 | –10.75 | –10.30 | 13.36 |
| | Valeric acid (C5) | –81.09 | –62.12 | –37.63 | –12.02 | –12.92 | 10.43 |
| | Hexanoic acid (C6) | –82.94 | –65.69 | –39.93 | –15.13 | –13.36 | 7.40 |
| | Heptanoic acid (C7) | –87.10 | –62.81 | –42.13 | –17.88 | –15.74 | 6.75 |
| | Octanoic acid (C8) | –86.16 | –64.40 | –44.03 | –19.74 | –22.63 | 2.70 |
| | Nonanoic acid (C9) | –85.74 | –65.88 | –45.23 | –19.94 | –28.36 | 5.54 |
| Diacids | Succinic acid (C4) | –119.94[a] | –77.89[a] | –20.39[a] | 18.96[a] | 13.20[a] | 63.06[a] |
| | Glutaric acid (C5) | –132.37[a] | –79.91[a] | –24.82[a] | –1.73[a] | 17.91[a] | 56.58[a] |
| | Adipic acid (C6) | –138.79 | –80.71 | –39.56 | –11.31 | –7.43 | 62.73 |
| | Pimelic acid (C7) | –145.84 | –71.39 | 0.32 | –12.50 | –5.14 | 60.56 |
| | Suberic acid (C8) | –141.07 | –84.28 | –16.93 | –18.26 | –8.81 | 62.70 |
| | Azelaic acid (C9) | –151.40 | –86.91 | –7.15 | –15.84 | –10.40 | 70.71 |

Note that positive binding is indicated with a negative number, thus most negative is best. Ligands found experimentally (Malnic *et al.*, 1999) to activate the receptor are highlighted for each receptor.
[a]Binding energies for ligands not tested experimentally are highlighted in gray.

second best calculated binding energy (the C7 diacid pimelic acid) was not observed experimentally. However, suberic acid elicited only a weak response in the experiments (Malnic *et al.*, 1999) and it is possible that pimelic acid could elicit response at higher concentrations or if tested with longer residence time.

The ranking of ligands by predicted binding energy for S18 (Figure 1c), S19 (Figure 1d) and S46 (Figure 1e) correlates well to the experimental recognition profiles within the same chemical class. Thus ligands found to elicit experimental response have more favorable binding energies than ligands that do not elicit response. However, the energy differences cannot clearly separate them apart. For example, the calculated binding energies of the alcohols that activate the ORs S18 and S19 are not comparable to the binding energies of acids. This discrepancy in binding energies across chemical classes could arise from many factors such as dominance of the electrostatic contribution to the binding energy of the charged ligands with insufficient desolvation contribution, or underestimation of the hydrogen bond contribution to the energy of the receptor–ligand complex. Moreover, the binding energies are calculated at 0K temperature and do not include entropic effects.

The predicted binding profile for S25 is in good agreement with the experimental one. The alcohols are predicted to bind much more strongly than any of the other three classes. Within the alcohol series, heptanol and hexanol, which were found to elicit response experimentally (Malnic *et al.*, 1999), have the best (more negative) binding energies, although butanol (not tested experimentally), pentanol and octanol are very close in energy to the experimental positives.

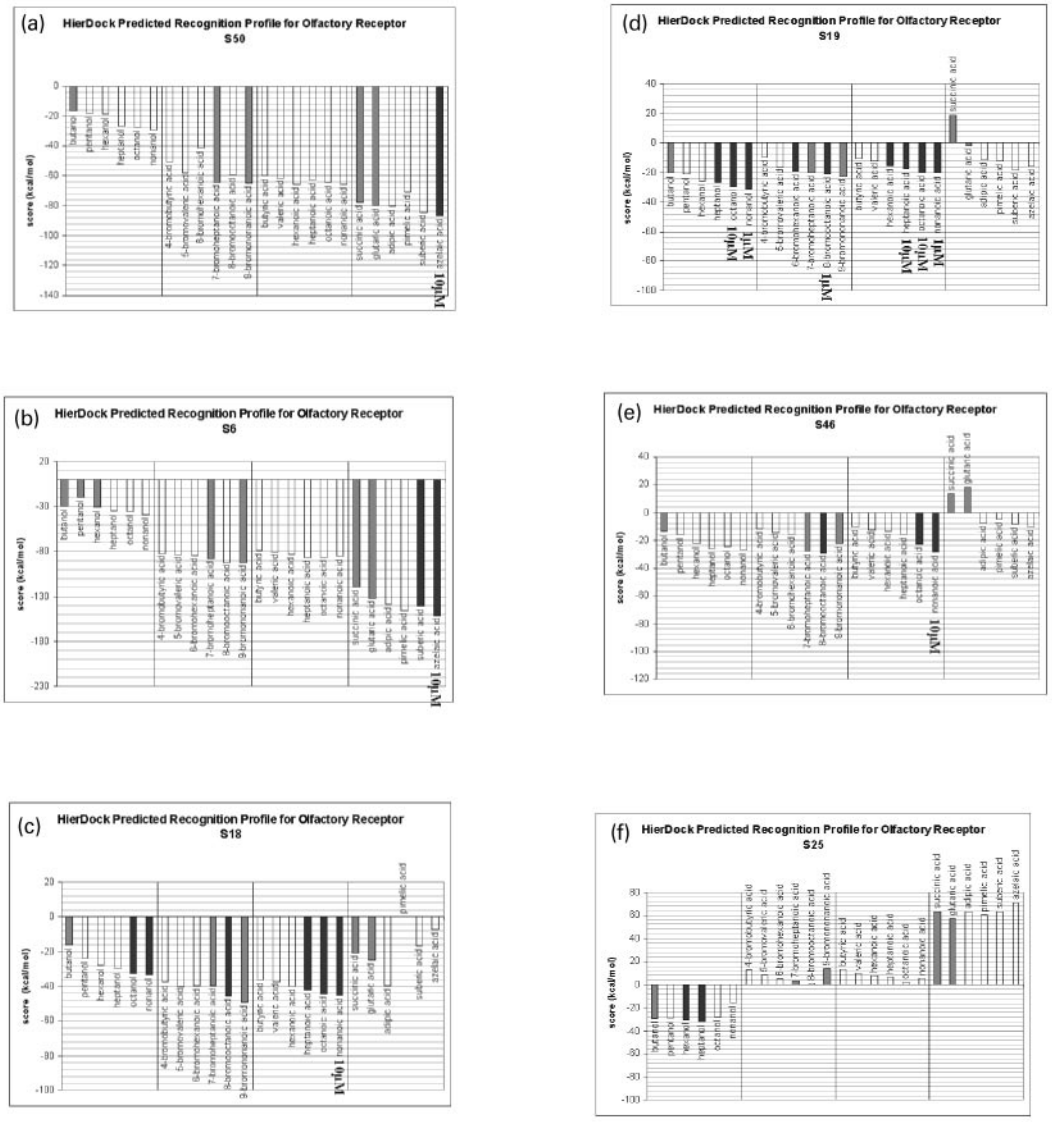



**Figure 1** Calculated (HierDock) binding profiles for mouse olfactory receptors **(a)** S50, **(b)** S6, **(c)** S18, **(d)** S19, **(e)** S46 and **(f)** S46. Odorants found to elicit experimental response (Malnic *et al*., 1999) are filled in black. Odorants that were not tested experimentally are filled in grey. Odorants that do not elicit response experimentally are represented as non-shaded bars. The experimental concentration thresholds (Malnic *et al*., 1999) are provided to show their correlation with the calculated binding energies.

Overall, all three receptors give good correlation for each of the four chemical classes between the predicted binding energies and the observed activation profile. However, there seem to be differences in the absolute values for the activation threshold for different odorant classes. Also, the absolute binding energies for different receptors vary significantly between the various cases. This discrepancy in binding energies for different chemical classes could arise from

1. slight errors in the 3-D structure (rotation or translation of one or more helices by an Ångstrom or so that might have differential effects on different classes);
2. approximations in the solvation of the active site where we use a non-variable dielectric constant and ignore the possible role of water (and metals such as Zn) buried in the active site; or
3. differential charging of the bound ligand with respect to the ligand isolated in solvent (we assume that all acids have the same net charge in the protein and in solution.

The calculated odorant binding profiles also correlate with the experimental concentration threshold. Thus, nonanoic acid (which elicits experimental responses at 10 μM for S18 and S46 and at 1 μM for S19) has higher binding affinity than the acids that elicit experimental response at 10 μM (octanoic and heptanoic acids in S19) and 100 μM (octanoic acid in S46; octanoic and heptanoic acids in S18; hexanoic acid in S19). Correlation between concentration threshold and binding energy is also found for S19 binding to alcohols

where nonanol (1 μM) has better (more negative) binding energy than octanol (10 μM), which is better than heptanol (100 μM). The same holds true for S19 and bromo acids, with 8-bromo octanoic acid (1 μM) ranking better than 6-bromo hexanoic acid (100 μM). In all these cases, the calculated binding decreases with the decrease in carbon chain length. Three odorants not tested experimentally show better binding energy than known ligands and are consistent with the trend described above: 9-bromo nonanoic acid for S18; 9-bromononanoic acid and 7-bromoheptanoic acid for S19. We predict that these ligands should bind and likely serve as agonists.

### Location and general features of the binding sites

Figure 2 shows that the binding sites (side view and the extracellular top view) of odorants in all six ORs are located between TMs 3, 4, 5 and 6 and lie ~10 Å below the extracellular loops of the helical barrel. Figure 3 superimposes the predicted binding site for azelaic acid in S6 (receptor and ligand in blue) and compares it to the binding site of *cis*-retinal in bovine rhodopsin crystal structure (pdb code: 1F88) colored in red. This shows that the binding sites for ORs are in the same general location as retinal in rhodopsin, even though all potential binding sites were determined by independent scanning of the receptors. Figure 3 also shows the differences between the crystal structure of bovine rhodopsin and our predicted structure for OR S6.

Residues within 3.5Å of the experimental agonists for all six olfactory receptors are shown in Figure 5 and Table 4. The same set of residues is also marked in the sequence alignment of these ORs shown in Figure 4. Residues that form hydrogen bonds with the ligand are shown in bold font in Table 4. Amino acid positions in the sequence alignment (Figure 4) found to be involved in odorant binding for four ORs (S18, S19, S25 and S46) are highlighted in Table 4. These are positions 6, 9, 10 and 13 in TM3, and position 19 in TM6. Consensus positions 6, 9 and 13 in TM3 are in the same face of that helix. These four ORs recognize odorants from the same chemical classes, while the other two (S6 and S50) recognize a class of odorants (diacids) not recognized by any of the other four ORs.

We assume that these amino acid positions found to be consistently involved in binding of odorants in S18, S19, S46 and S25 are indeed responsible for odorant recognition. We will now examine whether this information on the binding site can be used to explain the differences in the experimental odorant activation profiles for these ORs.

The ORs that recognize acids and bromo acids (S18, S19 and S46) have a His in position TM3-6 (this notation means sixth residue on TM3 starting from the amino end of the helix 3), a residue with a short hydrophobic side chain (Val, Ala or Gly) at position TM3-10 and a Pro in position TM6-19.

The only OR among the six considered here that does *not* recognize acids (S25) has:

- a Val in TM3-6;
- a polar Thr at position TM3-10;
- another Thr in TM6-19.

The Thr284 at position TM6-19 in S25 is hydrogen-bonded to Lys302 (EC3), which is the anchor residue for the hydroxyl group.

The two ORs that recognize both alcohols and acids (S18 and S19) have Phe at position TM3-9, which is occupied by a Thr in the OR S46, that recognizes only acids and by a Gly in the OR S25 that is activated only by alcohols. Figure 6 shows a comparison between these positions in S18 (recognizes alcohols and mono-acids) compared with S25 (alcohols but not mono-acids) and S46 (mono-acids but not alcohols). Together, these four positions seem to explain the differences in chemical class preference for S18, S19, S25 and S46. Three of them (TM3-6, TM3-9 and TM3-10) are considered hypervariable positions in OR sequence alignments and have been suggested to be involved in differentiating among odorants (Malnic *et al.*, 1999; Pilpel and Lancet, 1999). The fifth consensus position TM3-13 (also a hypervariable position according to the literature; Malnic *et al.*, 1999; Pilpel and Lancet, 1999) does not seem to account for differences in preference among the Malnic 24 odorant set and may be involved in the binding of some other odorant not included in that set.

### Comparing the binding sites for diacids in the homologues S6 and S50

The olfactory receptors S6 and S50 share 94% sequence identity. The structural homology of the MembStruk predicted structures for S6 and S50 is high showing a CRMS difference of only 1.5Å in the TM domain. The main deviations are in the loop conformations, which are not significant since the loops are conformationally flexible. S50 recognizes only azelaic acid (C9 diacid), while S6 recognizes both azelaic acid and suberic acid (C8 diacid). The mutations from S50 to S6 responsible for changing the activity are in TM2 (Leu82–Val), TM3 (Val112–Ile, Ser113–Val, Met114–Phe) and TM5 (Ala202–Val, Ile205–Thr, Ile206–Val, Thr208–Ile, Thr220–Ala).

Figures 4 and 5 show that of the 13 residues found within 3.5 Å of agonist ligands, only four residues (Tyr109, Leu204, Arg207 and Ser261) are common to both S6 and S50. We find that a single mutation at Ser113 (TM3)–Val, in S50 to S6 is directly involved in odorant binding. However, 8 out of 16 mutations between S50 and S6 are in TMs 3 and 5 and are involved in forming the shape of the active odorant binding site in S50 and S6. Another structural difference is that S6 has significant contributions from residues in EC2, which in S6 is somewhat folded towards the opening of the helix barrel, whereas it is in an 'open' conformation for S50. Since loops are the least accurate parts of our predicted 3-D structures, loop residues identified with binding should be taken with some reservation. It should be noted, however, that

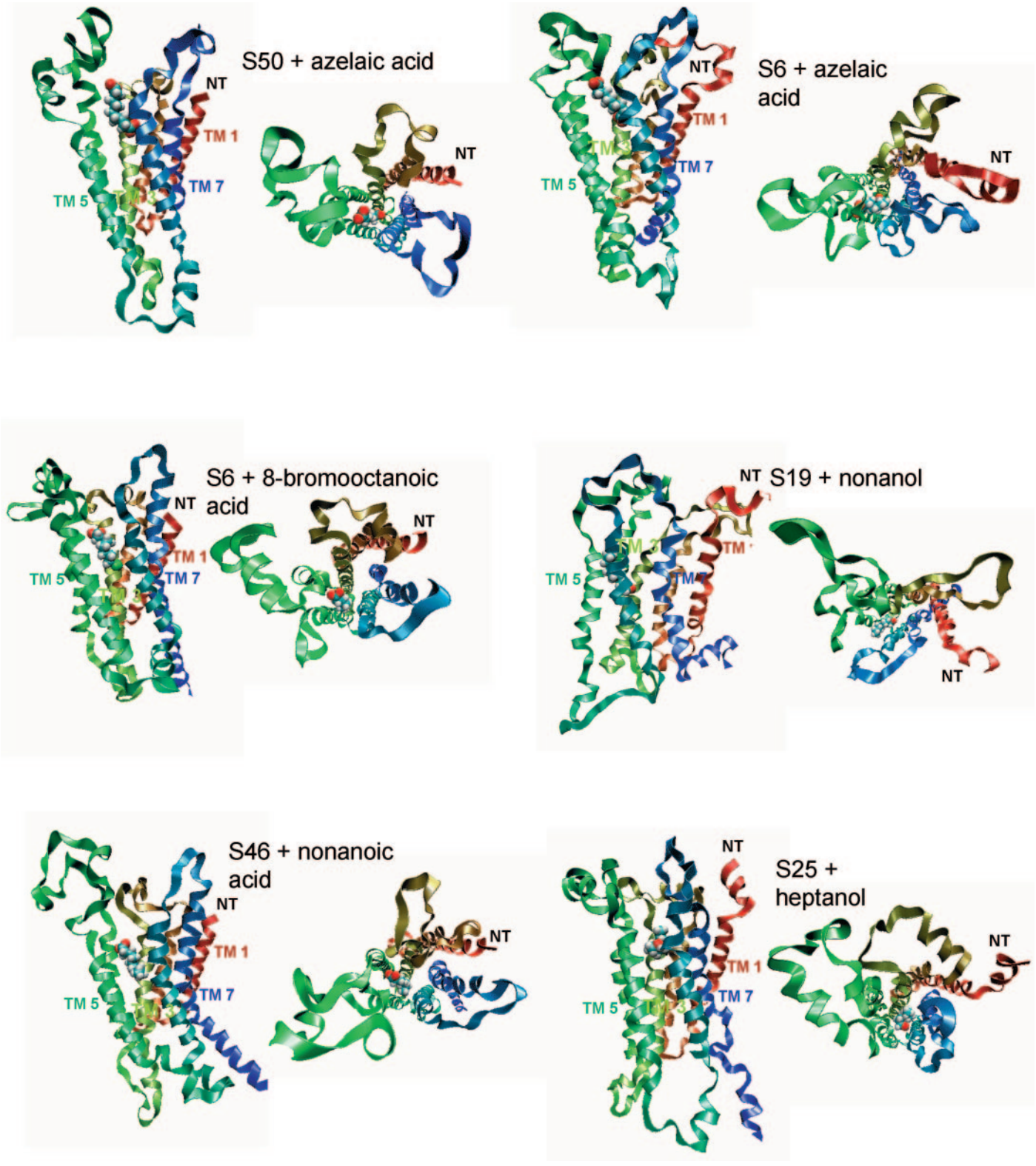



**Figure 2** Front and top (extracellular end) view of the 3-D structures of mouse olfactory receptors S6, S50, S18, S19, S25 and S46 with their respective agonists bound to the HierDock predicted binding sites. The TMs are, counter-clockwise from the back: TM 1 (red-brown), TM2 (ochre), TM3 (moss green), TM4 (lime green), TM5 (green), TM6 (light blue) and TM7 (blue). Figures were generated using VMD (Humphrey *et al.*, 1998).

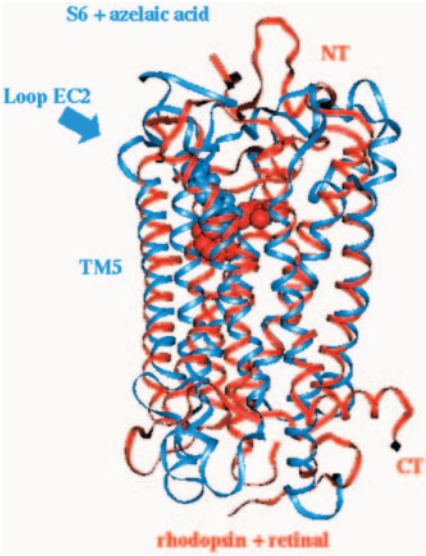



**Figure 3** Structural superposition of mouse olfactory receptor S6 (with azelaic acid bound to predicted binding site) to bovine rhodopsin bound to retinal (pdb code 1F88) showing the differences in the protein structures and the similarity in the binding sites location. S6 and azelaic acid are blue; rhodopsin and retinal are red. The OR structures and the locations of their binding sites were each derived independently. No structural information about the binding site location, length of the trans-membrane domains, tilt of the helices or loop configuration from the rhodopsin structure was used in predicting the structure of the ORs.

EC2 is folded into the binding site in the crystal structure of rhodopsin with bound *cis*-retinal.

### The binding site for alcohols in S25 compared with previous work

The calculated binding energy profile of the alcohol binding receptor S25 shows good agreement with the experimental profile as described in above. The binding energies reported here for odorants in S25 use the refined HierDock protocol described in this paper. The absolute binding energies are different from the previous published work (Floriano *et al.*, 2000), but the relative binding energy profile remains identical. Both studies lead to very good agreement between calculated and experimental profiles in terms of the chemical class recognized by the OR and, within that chemical class, the ranking of ligands by their likelihood of activating the receptor. We also found the same residues to be involved in binding and the same critical binding feature in the hydrogen bond (HB) to Lys302.

### Comparing the binding modes of acids and alcohols with ORs S18, S19 and S46

The binding site of mono-acids, bromoacids and alcohols in this family involves residues in TMs 3, 4, 5 and 6, and EC2. Odorants bind with their long axis generally parallel to the TMs, with their carboxyl group pointing toward the extracellular side of the helical barrel. Moreover their orientation is consistent with the idea that the N-terminal region of an alpha helix favors the binding of negatively charged ligands because of stabilization due to the positive charge of the helix dipole (Hol *et al.*, 1978). The binding mode (Figure 4) of alcohols to S19 involves a HB to Ser140 (TM3), while the binding mode for acids has His133 (TM3) as the anchor point. The hydrophobic tails of alcohols and acids are

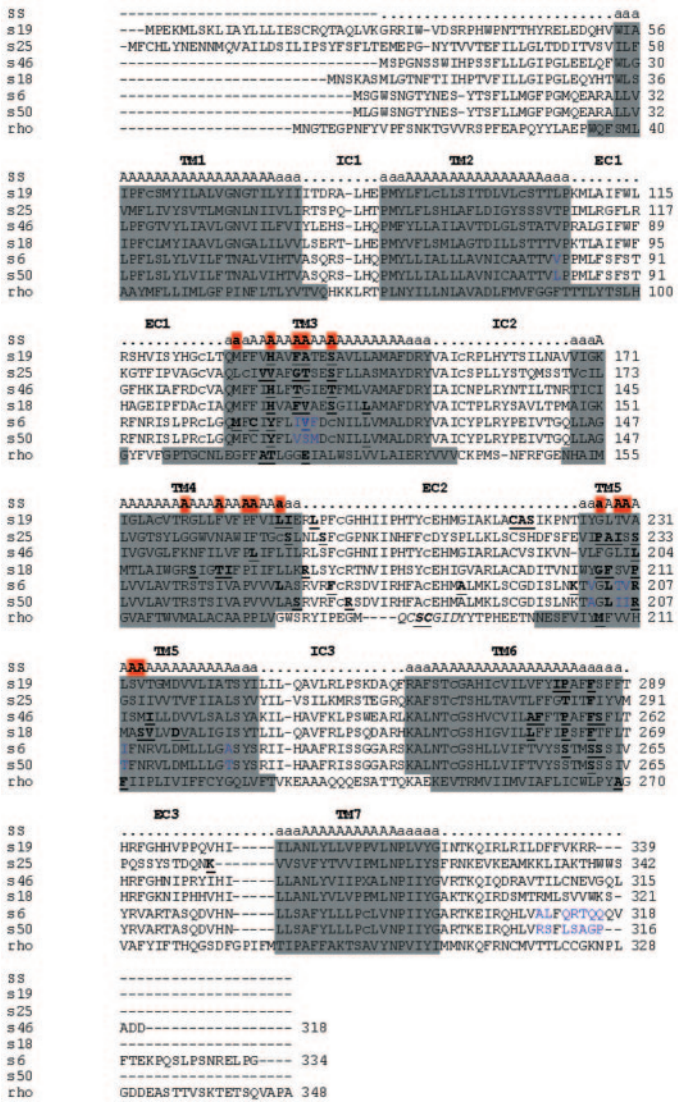
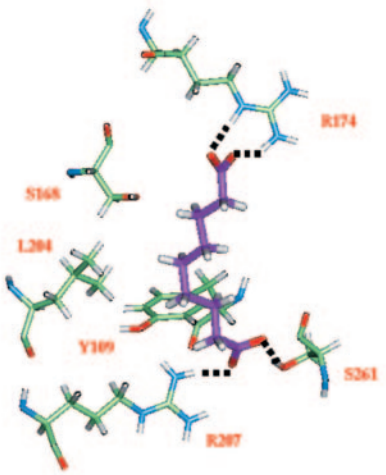



**Figure 4** Sequence alignment for mouse olfactory receptors S6, S50, S18, S19, S25, S46 and bovine rhodopsin. Predicted trans-membrane regions are highlighted. Residues found within 3.5 Å of an experimental agonist are in bold underlined font. Cys involved in disulfide bonds are in lower case letter. Blue letters represent the amino acid differences in the 94% homologous pair S6 and S50. Hypervariable positions are highlighted in the secondary structure assignment rows. There have been suggestions in the literature that hypervarient positions are involved in odorant recognition (Malnic *et al.*, 1999; Pilpel and Lancet, 1999). Trans-membrane helices (TM) 1–7, extracellular loops (EC) 1–3 and intracellular loops (IC) 1–3 are identified in the alignment. For bovine rhodopsin, the secondary structure assignment was from the pdb file 1F88. Some of the residues found to be involved in binding in our models are also hypervariable positions. Trans-membrane domains 3, 5 and 6 have positions consistently involved in binding for most ORs studied here.

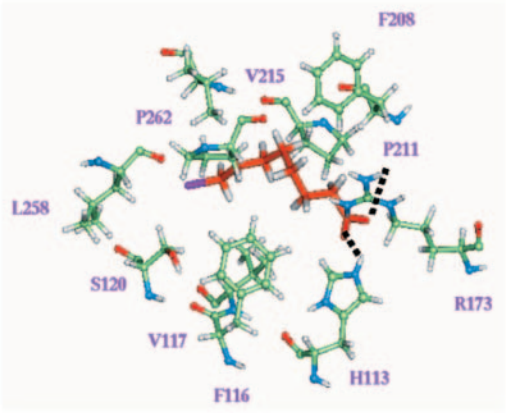
**Table 4** Residues involved in binding of experimental agonists for the six mouse olfactory receptors

| Structural element | Position | S6 | S50 | S25 | S18 | S19 | S46 | Fingerprint for aliphatic acids | Fingerprint for aliphatic alcohols |
|---|---|---|---|---|---|---|---|---|---|
| TM3 | 2 | M105 | | | | | | | |
| | 5 | | | V134 | | | | | |
| | 6 | Y109 | Y109 | V135 | **H113** | **H133** | **H107** | His | |
| | 9 | | | G138 | F116 | F136 | T110 | | Non-polar or Gly |
| | 10 | V113 | | T139 | V117 | A137 | G111 | Non-polar or Gly | |
| | 13 | | | S142 | S120 | **S140** | T114 | Ser, Tyr or Thr | Ser |
| | 17 | | | | L124 | | | | |
| TM4 | 13 | | | | S160 | | | | |
| | 16 | | | | T163 | | | | |
| | 17 | | | | I164 | | | | |
| | 20 | | | | | | L161 | | |
| | 23 | L166 | | | | L190 | | | |
| | 24 | | | S194 | | I191 | | | |
| | 25 | | S168 | | | | | | |
| EC2 | 1 | | | | **R173** | | | | |
| | 2 | | | | | L194 | | | |
| | 4 | F172 | | | | | | | |
| | 6 | | **R174** | | | | | | |
| | 19 | A187 | | | | | | | |
| | 25 | | | | | C217 | | | |
| | 26 | | | | | A218 | | | |
| | 27 | | | | | S219 | | | |
| | 32 | **K200** | | | | | | | |
| TM5 | 3 | | | | G207 | | | | |
| | 4 | L204 | L204 | A230 | F208 | | | | |
| | 5 | | | I231 | | | | | |
| | 7 | **R207** | **R207** | S233 | P211 | | L204 | | |
| | 10 | | | | V215 | | | | |
| | 11 | | | | **S214** | | I208 | | |
| | 14 | | | | **D218** | | | | |
| TM6 | 15 | | | | L258 | | A251 | | |
| | 16 | | | | | | F252 | | |
| | 18 | | | | | I281 | | | |
| | 19 | S258 | | **T284** | P262 | P282 | P255 | Pro | Polar or Pro |
| | 22 | S261 | S261 | F287 | F265 | F285 | F258 | Phe | Phe |
| | 23 | | | | | | S259 | | |
| EC3 | 12 | | | **K302** | | | | | |

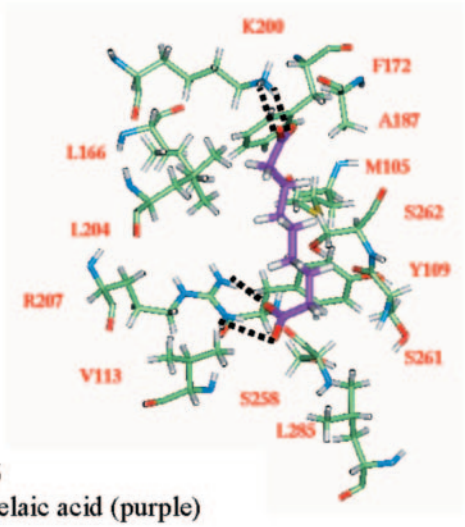
Residues involved in ligand–protein hydrogen bonds are in bold font. Only residues within 3.5 Å of a recognizable ligand are marked. Positions found to be consistently involved in binding in four of the receptors (consensus positions) are highlighted in gray. The differences in the experimental recognition profiles of mono-acids and alcohols binders can be explained on the basis of these positions alone. Sequence fingerprints for aliphatic acids and aliphatic alcohols recognition were derived from these consensus positions and used to identify similar ORs from the mouse genome.



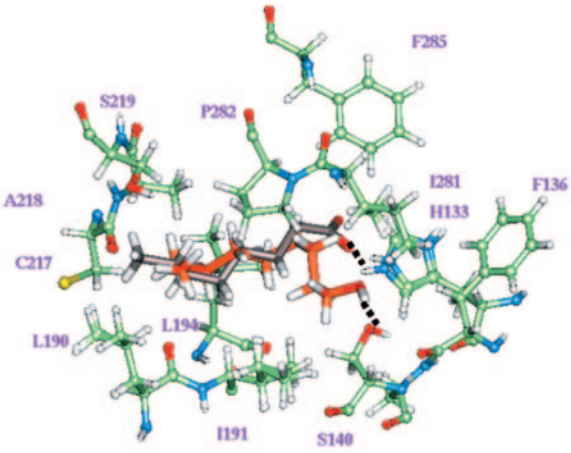

S50
azelaic acid (purple)



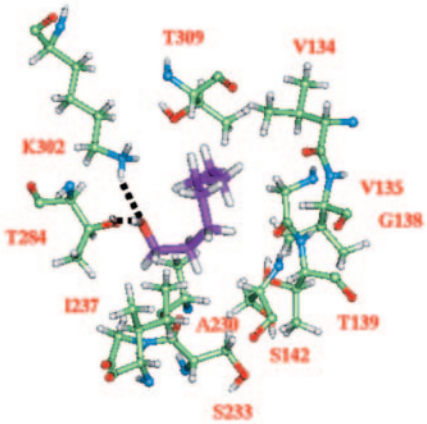

S18
8-bromo-octanoic acid (orange)



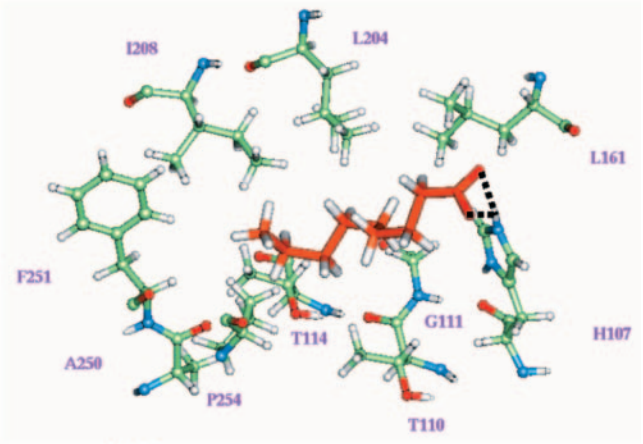

S6
azelaic acid (purple)




S19
nonanol (orange) and octanoic acid (gray)




S25
Heptanol (purple)




S46
Nonanoic acid (orange)

**Figure 5** Details of the binding modes for the six mouse olfactory receptors: azelaic acid/S6, azelaic acid/S50, 8-bromo-octanoic acid/S18, (nonanol and octanoic acid)/S19, nonanoic acid/ S46 and heptanal/ S25. Hydrogen bonds to anchor residues in the binding site are traced in the figures. Different ORs use different binding modes to recognize the same ligands. Different anchor points within the same binding site can also be used for different functional groups.

bound similarly, with Leu190 (TM4), Ile191 (TM4), Leu194 (EC2), Ile 281 (TM6), Pro282 (TM6) and Phe285 (TM6) providing a hydrophobic pocket around it. The hydroxyl and carboxyl groups point in opposite directions in order to reach their anchors. This shows that the same OR can bind to two different odorants in two different binding modes.

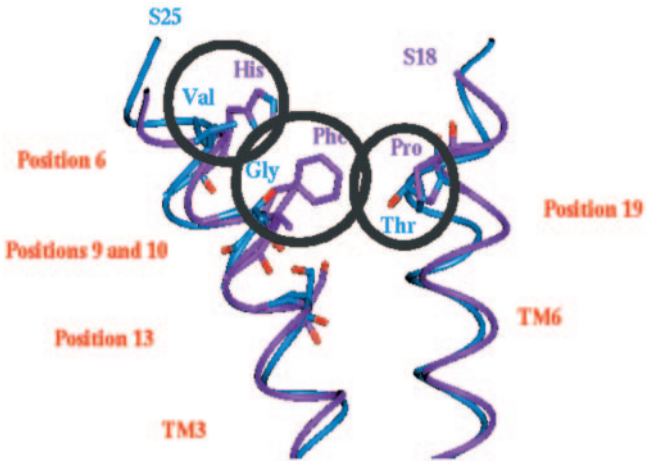



**Figure 6** Schematic diagram showing that the residues found to be involved in odorant binding are in conserved alignment positions across ORs S18, S19, S25 and S46. These four sequence alignment positions in TM3 and TM6 can explain the differences in the experimental response profiles. We compare S18 (recognizes long alcohols and mono-acids) to S25 (medium alcohols but not mono-acids) and S46 (long mono-acids but not alcohols). The ORs that recognize mono-acids (S18, S19 and S46) have a His in position TM3-6, a short hydrophobic side chain residue (V, A or G) at position TM3-10 and a Pro in position TM6-19, while the only OR in the set that does not recognize acids (S25) has a Val in TM3-6, a polar Thr at position TM3-10 and another Thr in TM6-19. The Thr284 at position TM6-19 in S25 is hydrogen-bonded to Lys302 (EC3), which is the anchor residue for the hydroxyl group. The two ORs that recognize both alcohols and mono-acids (S18 and S19) have Phe at position TM3-9, which is occupied by a Thr in the acids-only S46 and by a Gly in the alcohols-only S25.

### Comparing critical residues for binding to available data for other GPCRs

Above, we listed common positions in the ORs sequence alignment that we identify as critical for odor recognition and discussed their agreement with sequence hypervariability intrinsic to OR function. Now the question is how well these positions compare to experimental data for ligand binding to other GPCRs. For this comparison we chose a set of family A type (rhodopsin-like) GPCRs that have abundant experimental data because of their pharmaceutical relevance: muscarinic acetylcholine receptor, β2-adrenergic receptor, dopamine receptor, type-1B angiotensin II receptor and purinergic receptor. We mapped reported experimental data (Strader *et al.*, 1994; Wieland *et al.*, 1995; Bourdon *et al.*, 1997; Alberts *et al.*, 1998; Allman *et al.*, 2000; Klabunde and Hessler, 2002; Bissantz *et al.*, 2003) on residue positions involved in binding for those receptors and then compared their important positions for binding to the ones identified here for ORs. These are shown in Figure 7.

The sequences in Figure 7 were aligned as follows (no gaps were allowed in the TMs):

1. TM3 sequences were aligned based on the conserved pattern (D, E, or H)RY;
2. TM4 was aligned based on the conserved Trp;
3. TMs 5 and 6 were aligned using clustalW (Thompson *et al.*, 1994).

| | TM3 | TM4 | TM5 | TM6 |
|---|---|---|---|---|
| OR_S19 | QMFFV**H**AVFATESAVLLAMAFDRY | VIGKIGLACVTRGLLFVFPFVILIE | IYGLTVALSVTGMDVVLIATSYI | RAFSTCGAHICVILVFYIPAF**F**SFF |
| OR_S25 | QLCIV**V**AFGTSESFLLASMAYDRY | VCILLVGTSYLGGWVNAWIFTGCSL | VIPAISSGSIIVVTVFIIALSYV | KAFSTCTSHLTAVTLFFGTII**T**FIYV |
| OR_S46 | QMFFI**H**LFTGIETFMLVAMAFDRY | TICIIVGVGLFKNFILVFPLIFLIL | VLFGLLILISMILLDVVLSALSYA | KALNTCGSHVCVILAFFTPAF**F**SFL |
| OR_S6 | QMFCI**Y**FLIVFDCNILLVMALDRY | LLAGLVVLAVTRSTSIVAPVVVLAS | TVGLTVRIFNRVLDMLLLGASYS | KALNTCGSHLLVIFTVYSSTMS**S**SI |
| OR_S50 | QMFCI**Y**FLVSMDCNILLVMALDRY | LLAGLVVLAVTRSTSIVAPVVVLAS | TAGLIIRTFNRVLDMLLLGTSYS | KALNTCGSHLLVIFTVYSSTMS**S**SI |
| OR_S18 | QMFFI**H**VA**F**VAES**S**ILLAMAFDRY | AIGKMTLAIWGRSIGTIFPIIFLLK | WYGFSVPMASVLVDVALIGISYT | KALNTCGSHIGVILLFFIPSF**F**TFL |
| rhodopsin | NLEGFF**A**T**L**GG**E**IALW**S**LVVLAIERY | AIMGVAFTWVMALACAAPPLV | FVIY**M**FVVH**F**IIPLIVIFFCYG | VTRMVIIMVIAFLICWLPY**A**GVAFY |
| acm1_human | LWLALD**Y**VASNASVMNLLLISFDRY | AALMIGLA**W**LVSFVLWAPAILF**W** | I**I**TFGT**A**MAA**A**FYLPVTVMC**T**LYW | LSAILLAFILTWTPY**N**IMVLV |
| b2ar_human | EFWTSI**D**VLC**V**TASIETLCVIAVDRY | RVIILMV**W**IVSGLTSFLPIQMHWY | QAYAIA**SS**IV**S**FYVPLVIMVFVYS | LGIIMGTFTLCWLP**FF**IV**N**IVHVI |
| d3dr_human | VFVTL**D**VMMCTASILNLCAISIDRY | VALMITAV**W**VLAFAVSCPLLFGF | PDFVIY**SS**VV**S**FYLPFGVTVLVYA | MVAIVLGAFIVCWLP**FF**LTHVL |
| ag2s_rat | IASAS**V**SFNLYASVFLLTDRY | LVAKVTCIIIWLMAGLASLP | IGLGLT**K**NILGFVFPFLIILTS | IIMAIVLFFFFSWVPHQIFTFL |
| p2yr_human | Q**R**FI**FH**VNLYGSILFLTCISAHRY | NAICISVLVWLIVVVAISPILF | MCT**T**VAMFCVPLVLILGCYG | LTVFAVSYIPF**H**VM**K**TMNLR |

**Figure 7** Comparison between residues predicted as critical for odorant recognition in ORs and residues experimentally found to be involved in agonist binding for other GPCRs. Transmembrane domains for rhodopsin (1F88), muscarinic acetylcholine receptor (acm1_human), β2-adrenergic receptor (b2ar_human), dopamine receptor (d3dr_human), type-1B angiotensin II receptor (ag2s_rat) and purinergic receptor (p2yr_human) correspond to the regions assigned in the swiss-prot files. For each OR, positions found to be involved in binding are in underlined font with consensus positions (i.e. common to all six cases presented in this paper) bold underlined. TM residues found within 3.5 Å of retinal in the rhodopsin structure 1F88 are underlined in the rhodopsin sequence. Residues experimentally found to be involved in binding (Strader *et al.*, 1994; Wieland *et al.*, 1995; Bourdon *et al.*, 1997; Alberts *et al.*, 1998; Allman *et al.*, 2000; Klabunde and Hessler, 2002; Bissantz *et al.*, 2003) for muscarinic acetylcholine (Allman *et al.*, 2000), β2-adrenergic receptor, dopamine, type-1B angiotensin II (Klabunde and Hessler, 2002) and purinergic receptors (Klabunde and Hessler, 2002) are also bold underlined in their respective sequences. The Trp residue half way through TM4 that is conserved in most GPCR (Rhee *et al.*, 2000) is not conserved in ORs. In fact, for the six ORs studied here, only S18 has this Trp. The binding sites in the ORs were predicted based on interaction energies between odorants and ORs with no use of experimental information. All four OR consensus positions in TM3 have been implicated in ligand binding for other GPCRs. The OR consensus position in TM6 is in the same region of the helix as in the other receptors; however, not exactly in the same position. One of the two residues in TM5 that we found to be involved in binding for some but not all six ORs studied here is also in good agreement with the data for the other GPCRs.

The Trp residue half way through TM4 that is conserved in most GPCR (Rhee *et al.*, 2000) is not conserved in ORs. In fact, out of the six ORs studied here, only S18 has this Trp, so we used S18 for aligning TM4 of the ORs to the other GPCRs. The alignment for ORs and rhodopsin shown in Figure 4 was based on the complete amino acid sequences and hence it differs from the TM-based alignment for rhodopsin TMs in Figure 7.

All four OR consensus positions in TM3 have been implicated in ligand binding for other GPCRs. The OR consensus position in TM6 is in the same region of the helix as for the other receptors; however, it is not in exactly the same position. One of the two residues in TM5 that we found to be involved in binding for some but not all six ORs studied here is also in good agreement with the data for the other GPCRs. The good agreement between our predictions and the available data for these other GPCRs is one indication that our methods provide accurate structures and binding site, since the binding sites in the ORs were predicted based on interaction energies between odorants and ORs with no use of experimental information.

## Screening odorant libraries to find potential agonists

We docked all ligands shown in Table 1B to all six OR structures studied here and sorted them in order of decreasing binding affinity. The top 10 potential ligands by binding affinity for each OR are shown in Table 5. These ligands are candidates to serve as agonists or antagonists to these ORs. We included sugars in our screening set, even though they are generally odorless, in order to provide data for comparing the binding affinities to ORs of odorants versus tastants (such as glucose, a natural agonist for taste receptors, which are functionally related members of the GPCR family). This may provide insights into how function diversity evolved in the GPCR family, and more specifically, among the chemosensory receptors.

Binding of a ligand to an OR is a necessary but not sufficient condition for activation and hence calculated binding energies alone cannot predict which molecules will elicit response *in vivo*. A ligand could bind strongly and not activate the G protein. Also, the calculated binding energies do not take into account many physical–chemical properties that may differentiate structurally similar ligands *in vivo*. However, we have shown in the previous sections that there is correlation between calculated binding energies and the chemical classes recognized by each OR. The binding energies also indicate very well odorants most likely to activate the receptor within each of the four odorant classes of Malnic, even though the energy distribution may not clearly separate active and inactive ligands. In addition, there is some correlation when comparing ligands of different classes. Thus we assume that ligands with better (more negative) calculated binding energies than the best experimental positives are likely to activate the ORs, while ligands with less favorable binding energy than the best experimental positives are not likely to activate the ORs. These ligands can be tested experimentally.

The experimental procedure used in identifying the six ORs studied here may lead to functional relationships. Experimentally, calcium imaging was used (Malnic *et al.*, 1999) to select from 647 mouse olfactory neurons from the dorsal nasal septum those that responded to one or more of the odorants in Table 1A. Then, RT–PCR was used to identify the OR genes expressed by the responsive neurons. Odorants were first tested at 100 μM, then at 10 and 1 μM, depending on response. Consequently, only ORs sharing affinity to those odorants were selected and identified, making them likely to share affinity for other odorants.

We find that some ligands have favorable binding energy to more than one of the six OR studied here. This may arise because this set of ORs was selected by their common affinity to aliphatic alcohols and acids with 6–9 carbons. This may explain their common preference for the aldehydes and esters in Table 5.

The most frequent hits are (number of ORs for which they rank among the top 10 binders/total number of ORs considered between parentheses):

- furaneol glucoside (6/6);
- ethyl vanillin (6/6);
- glucose (5/6);
- furaneol (2,5-dimethyl-4-hydroxy-3(2H)-furanone) (4/6);
- mesifurane (4/6);
- eugenol (4/6);
- decanol (4/6);
- tridecanal (3/6);
- sec-butyl-3-methoxypyrazine (3/6).

Interestingly, the ligands showing the best calculated binding affinities to the ORs studied in our work are important aroma and/or flavor components of nutrient sources mice are known to like, such as fruits, vegetables and cheese (Bramwell *et al.*, 1969; Buttery *et al.*, 1969, 1990; Murray *et al.*, 1970; Maga and Sizer, 1973; Cronin and Stanton, 1976; Ohloff, 1978; Wu *et al.*, 1991; Duke, 1992; Belay and Poole, 1993). This is also true for the Malnic compounds (Table 1A) that elicit experimental response from these ORs (Murray *et al.*, 1970; Pabst *et al.*, 1991; Wu *et al.*, 1991; Teranishi *et al.*, 1992; Krammer *et al.*, 1994; Sanz *et al.*, 1995; Forney, 2001; Jordan *et al.*, 2001; Suriyaphan *et al.*, 2001; Jackson and Linskens, 2002; Lavid *et al.*, 2002). In particular, furaneol and its methoxy derivative (mesifurane) are among the most important volatile compounds in the aroma of strawberries (Sanz *et al.*, 1995; Forney, 2001; Lavid *et al.*, 2002). The concentration of these compounds along with the furaneol glucoside derivative, sharply increase during fruit ripening, with maximum values at the ripe stage (Forney, 2001; Lavid *et al.*, 2002). Furaneol is also an important component in the aroma and flavor of many fruits including pineapples (Wu *et al.*, 1991), tomatoes

**Table 5** Suggested ligands for the six olfactory receptors

| Class | S50 ligands | bindE | Class | S6 ligands | bindE |
|---|---|---|---|---|---|
| Diacid | Azelaic acid (1) | –86.91 | Diacid | Azelaic acid (1) | –151.40 |
| Triacid | Citric acid | –78.03 | Triacid | Citric acid | –140.82 |
| Monoacid | Pyruvic acid | –53.60 | Monoacid | Pyruvic acid | –104.65 |
| Monosaccharide | Glucose | –48.44 | Glycoside | Furaneol glucoside | –81.42 |
| Glycoside | Furaneol glucoside | –41.88 | Monosaccharide | Glucose | –56.38 |
| Aldehyde, phenol | Ethyl vanillin | –33.67 | Aldehyde, phenol | Ethyl vanillin | –50.77 |
| Alcohol | Eugenol | –32.88 | Alcohol | Furaneol | –44.41 |
| Ester | Hexyl octanoate | –28.30 | Ether | Mesifurane | –39.16 |
| Aldehyde | Benzaldehyde | –27.93 | Ether,amide | Piperine | –38.33 |
| Alcohol | 1-Decanol | –27.66 | Aldehyde | Lyral | –36.73 |
| Alcohol | 3-Phenyl-1-propanol | –27.39 | Ketone | Butyrophenone | –36.41 |
| | S46 ligands | | | S19 ligands | |
| Glycoside | Furaneol glucoside | –58.84 | Glycoside | Furaneol glucoside | –62.54 |
| Alcohol | 1-Decanol | –31.44 | Monosaccharide | Glucose | –51.36 |
| Ester | Isoamyl nonanoate | –30.09 | Alcohol | Eugenol | –42.57 |
| Bromoacid | 8-Bromooctanoic acid (4) | –28.80 | Alcohol | Furaneol | –42.28 |
| Alcohol | Eugenol | –28.53 | Ether | Mesifurane | –40.67 |
| Acid, alcohol | 5-Hydroxy-tetradecanoic acid | –27.44 | Aldehyde, phenol | Ethyl vanillin | –33.37 |
| Aldehyde | Tridecanal | –26.67 | Aldehyde | Undecanal | –32.88 |
| Aldehyde, phenol | Ethyl vanillin | –26.31 | Alcohol | 1-Decanol | –32.84 |
| Ether | Mesifurane | –26.14 | Ether | 2-Sec-butyl-3-methoxypyrazine | –32.05 |
| Ether | 2-Sec-butyl-3-methoxypyrazine | –26.05 | Ketone | Propiophenone | –31.79 |
| Alcohol | Furaneol | –25.93 | Alcohol | Nonanol (11) | –31.61 |
| | S18 ligands | | | S25 ligands | |
| Glycoside | Furaneol glucoside | –70.06 | Glycoside | Furaneol glucoside | –71.26 |
| Bromoacid | 8-bromooctanoic acid (3) | –45.52 | Monosaccharide | Glucose | –48.75 |
| Monosaccharide | Glucose | –40.47 | Alcohol | Furaneol | –37.25 |
| Aldehyde | Tridecanal | –39.80 | Ether | 2-Sec-butyl-3-methoxypyrazine | –32.99 |
| Alcohol | Eugenol | –39.01 | Alcohol | 1-Decanol | –32.70 |
| Aldehyde | Undecanal | –38.78 | Ether | 2-Isopropyl-3-methoxypyrazine | –32.20 |
| Aldehyde | Dodecanal | –35.82 | Aldehyde | Tridecanal | –32.13 |
| Aldehyde | Nonanal | –35.19 | Alcohol | Heptanol (8) | –31.97 |
| Alcohol | Geraniol | –34.75 | Aldehyde, phenol | Ethyl vanillin | –31.51 |
| Aldehyde | Ethyl vanillin | –34.63 | Ether | Mesifurane | –30.74 |
| Ester | Isoamyl nonanoate | –32.76 | Ester | Hexyl octanoate | –30.42 |

These were selected as the top 10 by calculated binding affinity among compounds in Table 1A. The known experimental agonist with best predicted binding affinity is shaded with its global ranking between parenthesis shown as reference. Calculated binding affinities (bindE) are in kcal/mol.

(Krammer *et al.*, 1994), mangos (Lavid *et al.*, 2002), grapes (Gunata *et al.*, 1985) and raspberries (Pabst *et al.*, 1991).

Our results support the idea that natural agonists for ORs are likely to be found among environmentally related substances.

The high affinity of the glycosidically bound furaneol to all six ORs studied here may indicate the complex role ORs play in flavor perception. Glycosidically bound aroma compounds are considered important flavor precursors (Gunata *et al.*, 1985; Pabst *et al.*, 1991; Wu *et al.*, 1991; Krammer *et al.*, 1994; Jackson and Linskens, 2002), and binding of those compounds to ORs may not lead to specific odor descriptor but instead to an overall 'pleasant' sensation.

## Other ORs sharing similar odorant preferences

One question that arises from the theory of combinatorial coding for odors is how many receptors are used to uniquely identify an odorant. We cannot answer this question until we finish predicting the structure and odorant binding of all ORs in the mouse or in the human genomes. However, we can provide a rough estimate of how many ORs could potentially recognize the same class of chemical compounds as the six ORs studied here. To do this we derived simplified recognition patterns for identification of aliphatic alcohol and aliphatic acid from the amino acids we found to be directly involved in the binding pocket of these odorants in the six ORs we studied. We found that these binding sites lead to specific positions in the ORs sequence alignment (see Table 4). We will now use these patterns to identify related ORs by pattern matching. In order to do this, we searched for all sequences in the TrEMBL database (http://ca.expasy.org/sprot/sprot-search.html) matching 'mouse' in the organism and 'MOR[number]' in the descriptor fields. The sequences resulting from this query were extracted from the database. Incomplete and redundant sequences (homology >97%) were eliminated from the set. Sequence alignment was performed using ClustalW (Thompson *et al.*, 1994) and the TM assignment for S25 was used to predict the TM domains in the alignment. Sequence alignment was also used to identify the 'S' olfactory receptors in the TrEMBL set. The final set had 869 mouse OR sequences.

The pattern (shown in Table 4) of important amino acids derived for binding of aliphatic acids with 6–9 carbons to S18, S19 and S46, is:

- His at position 6 in TM3;
- any non-polar or Gly residues at TM3/10;
- Ser, Tyr or Thr at TM3/13, Pro at TM6/19;
- Phe at TM6/22.

We found that 32 out of 869 OR sequences match this pattern making them potential receptors for similar aliphatic acids. The TrEMBL codes for these sequences are given in Table 6.

The pattern for recognition of aliphatic alcohols with 6–9 carbons (shown in Table 4) was derived as:

- non-polar residue or Gly at TM3/9;
- Ser at TM3/13;
- polar or Pro residues at TM6/19;
- Phe at TM6/22.

This pattern is satisfied by 36 sequences from the 869 set. The TrEMBL codes for these sequences are given in Table 6. Of these 36, 12 were in common with the 32 recognizing acids.

## Summary and conclusions

Our previous studies showed that MembStruk predicts the transmembrane part of the 3-D structures of GPCRs to ~3Å CRMS error compared with the crystal structures of bovine rhodopsin and bacteriorhodopsin (Vaidehi *et al.*, 2002). In addition, the studies of binding of retinal to bovine rhodopsin, epinephrine to β1 and β2 adrenergic receptors and dopamine to dopamine receptor are sufficiently accurate to account for experimentally measured ligand binding function for these receptors (Vaidehi *et al.*, 2002).

This paper shows that the predicted structures for six ORs lead to predicted odorant binding sites for the 24 experimentally studied odorants that lead to calculated binding energy profiles that correlate well with the experimentally determined activation profiles. In addition, the features common to all six ORs found in the predicted binding modes of known agonists are in good agreement with experimental data for related proteins such as muscarinic acetylcholine, β2-adrenergic, dopamine, type-1B angiotensin II and purinergic receptors (Strader *et al.*, 1994; Wieland *et al.*, 1995; Bourdon *et al.*, 1997; Alberts *et al.*, 1998; Allman *et al.*, 2000; Klabunde and Hessler, 2002; Bissantz *et al.*, 2003), as seen above.

The good correlation between experimental and calculated binding profiles for the six ORs studied here and the good agreement between predicted binding mode and mutation studies for other GPCRs validates that the 3-D structures built with MembStruk capture the binding site character sufficiently well to use them in predicting the recognition profiles for other potential agonists. In addition, the accuracy of these predictions validates the efficacy of HierDock for predicting binding site and binding energy. We should stress that the binding sites in the ORs were predicted based on interaction energies between odorants and ORs only, with no use of experimental information except to validate the predictions.

It may be argued that our 3-D models at the accuracy they are expected to be (~3Å in the TM region) are not suitable for detailing binding modes. However, the intrinsically less specific character of the binding of odorants to ORs, compared with the degree of conformational specificity required for agonist binding in other systems, imply binding

**Table 6** TrEMBL codes for the sequences extracted from the OR genome database set (comprised of 869 mouse OR amino acid sequences) by matching the recognition pattern deduced for 6–9 carbon aliphatic alcohols and aliphatic acids (highlighted in Table 4 based on predicted binding sites for mouse olfactory receptors: S18, S19, S46 and S25

| (a) Aliphatic acids | (b) Aliphatic alcohols |
|---|---|
| **S19 (Q9WU90)** | **S19 (Q9WU90)** |
| **S18 (Q9WU89)** | **S18 (Q9WU89)** |
| S46 (Q9WU93) | S25 (Q8VEZ0) |
| Q8VF03 | Q8VEV6 |
| Q8VF06 | Q8VF62 |
| Q8VF27 | Q8VFQ6 |
| **Q8VG19** | Q8VFR5 |
| Q8VG23 | Q8VFR6 |
| Q8VG24 | Q8VFR7 |
| Q8VG26 | Q8VFR9 |
| Q8VG28 | Q8VFV5 |
| **Q8VG78** | Q8VFV6 |
| Q8VG79 | Q8VFV7 |
| Q8VG84 | Q8VFV8 |
| **Q8VGA1** | Q8VFW0 |
| Q8VGU9 | Q8VFW3 |
| **Q8VGV3** | Q8VFW9 |
| **Q8VGV4** | Q8VFX1 |
| Q8VGV5 | Q8VG13 |
| Q8VGV6 | Q8VG17 |
| Q8VGV8 | **Q8VG19** |
| Q8VGW2 | **Q8VG78** |
| Q8VGW3 | **Q8VGA1** |
| Q8VGW4 | Q8VGE2 |
| **Q8VGW5** | **Q8VGV3** |
| Q8VGX4 | **Q8VGV4** |
| **Q8VGX8** | Q8VGW1 |
| **Q8VGY0** | **Q8VGW5** |
| **Q8VGY3** | **Q8VGX8** |
| Q8VGZ9 | **Q8VGY0** |
| Q8VH00 | **Q8VGY3** |
| **Q8VH03** | **Q8VH03** |
| **Q8VH18** | Q8VH09 |
| Q9EQQ6 | **Q8VH18** |
| **Q9WVD9** | Q9EQ89 |
| | Q9EQA7 |
| | Q9EQB0 |
| | Q9EQB6 |
| | **Q9WVD9** |

Bold face indicates receptors predicted to recognize both alcohols and acids. The TrEMBL codes for the 'S' ORs are shown in parenthesis.

interactions that are much less dependent on a specific binding conformation. That means that the features in common to all six independently generated 3-D models are very likely to represent the most important interactions for odorant recognition in ORs.

For the six mouse ORs studied here, the binding sites are located in the same region, between TM helices 3, 4, 5 and 6. This binding region contains a number of hypervariable residues among the ORs, consistent with their involvement in binding, as proposed in the literature (Malnic *et al.*, 1999; Pilpel and Lancet, 1999). However, we have been able to identify specific conserved amino acid motifs that we suggest are used as signatures for recognizing specific chemical functional groups (alcohol and acid).

Two classes of ORs have been identified in vertebrates (Freitag *et al.*, 1998): class I (fish-like receptors) and class II (mammalian-like receptors). Fish ORs respond to water-soluble odorants such as amino acids (Ivanova and Caprio, 1993), while mammalian ORs respond to volatile compounds (Mezler *et al.*, 2001). It has been suggested that class I ORs may be specialized in detection of water-soluble odorants, while class II detect volatiles (Freitag *et al.*, 1998; Mezler *et al.*, 2001). The two classes differ by the length of EC3 (longer in the fish-like class I ORs) and the sequence variability in TMs 3–5, although no obvious class-specific amino acid motif has been detected in the TM domains (Freitag *et al.*, 1998; Mezler *et al.*, 2001). Of the six ORs studied here, five (S18, S19, S46, S6 and S50) are selective for water-soluble odorants as expected for the fish-like class I ORs (Zhang and Firestein, 2002), while S25 is selective for alcohols, consistent with class II. Since we find that the preference of our class I receptors for acids involves amino acid differences at positions TM3-6, TM3-9 and TM6-19 when compared with S25, we suggest that these positions might lead to class-specific fingerprints.

We should emphasize that our predictions of binding sites and binding energies of the odorants do not indicate whether binding will lead to activation of the G protein, which is essential to odor recognition. Other factors not accounted for in our receptor/ligand docking studies include: transport of the ligands to the ORs, latency (reaction time for recognition), ability to promote the conformational changes in the receptor necessary for signaling and concentration thresholds. On the other hand, the studies by Malnic measure only activation of the ORs. That is, they only measure agonists. It is possible that some ligands which we find to bind strongly might not activate the ORs (that is they may serve as antagonists). Such a concept of good binder with no activation or 'antagonists' has not yet been addressed in the olfactory receptor literature.

Our studies raise points about the metrics used to organize the odorant/olfactory receptor space. Most studies of the olfactory system have tested the activation of the ORs by chemically related structures. This assumes implicitly that there is some direct relationship between odor perception

and chemical functional groups of odorants. This may be the case but an alternate hypothesis is the olfactory system may be organized in ways that depend on the relationship between odors in odorant space. The olfactory system might be organized around a built-in 'understanding' of the biosynthetic relationships between different molecules found in nature or it could be organized to recognize odorants with similar perception (Mamlouk *et al.*, 2003). Our study suggests that odorants having biosynthetic relationships in odor space could be recognized by a specific set of ORs. However, to put such analyses on a firm foundation requires determining the structure and odorant binding properties of all ORs in the genome. This work lays the basis for such a study by validating the computational techniques required.

The two sets of mouse ORs identified as potential receptors for aliphatic alcohols and aliphatic acids may form basic arrays for unique combinatorial coding of single molecules in these classes. This prediction is testable experimentally. We speculate that these sets may also code odorants such as esters and aldehydes.

Our studies suggest that the set of six ORs presented here are likely to respond to aldehydes and esters with chains longer than 9 carbons, as found in flavors related to mice diet. We suggest that such compounds may elicit response at concentration closer to physiological thresholds than the set of aliphatic alcohols and acids used in our simulations.

## Acknowledgements

We want to thank Professor Linda Buck and Dr Betina Malnic (Fred Hutchinson Cancer Research Center) for many helpful discussions about the experiments. We also want to thank Dr Michael Singer, Professor Gordon Shepherd (Yale Medical School) and Professor James Bower (University of Texas, San Antonio) for many insightful suggestions. In addition we thank Mr Spencer Hall for helpful discussions. This research was initiated with support by an ARO-MURI grant (Dr Robert Campbell) and completed with funding from NIHBRGRO1-GM625523, NIH-R29AI40567 and NIH-HD36385. The computational facilities were provided by a SUR grant from IBM and a DURIP grant from ARO. The facilities of the Materials and Process Simulation Center are also supported by DURIP-ONR, DOE (ASCI ASAP), NSF, MURI-ONR, General Motors, ChevronTexaco, Seiko–Epson, Beckman Institute and Asahi Kasei.

## References

**Alberts, G.L., Pregenzer, J.F.** and **Bin Im, W.** (1998) *Contributions of cysteine 114 of the human D3 dopamine receptor to ligand binding and sensitivity to external oxidizing agents*. Br. J. Pharmacol., 125, 705–710.

**Allman, K., Page, K.M., Curtis, C.A.M.** and **Hulme, E.C.** (2000) *Scanning mutagenesis identifies amino acid side chains in transmembrane domain 5 of the M1 muscarinic receptor that participate in binding the acetyl methyl group of acetylcholine*. Mol. Pharmacol., 58, 175–184.

**Belay, M.T.** and **Poole, C.F.** (1993) *Determination of vanillin and related flavor compounds in natural vanilla extracts and vanilla-flavored foods by thin-layer chromatography and automated multiple development*. Chromatographia 37, 365–373.

**Bissantz, C., Bernard, P., Hibert, M.** and **Rognan, D.** (2003) *Protein-based virtual screening of chemical databases II. Are homology models of G-protein coupled receptors suitable targets?* Proteins Struct. Funct. Genet., 50, 5–25.

**Bourdon, H., TrumppKallmeyer, S., Schreuder, H., Hoflack, J., Hibert, M.** and **Wermuth, C.G.** (1997) *Modelling of the binding site of the human M(1) receptor: experimental validation and refinement*. J. Comput. Aid. Mol. Des., 11, 317–332.

**Bozza, T.C.** and **Kauer, J.S.** (1998) *Odorant response properties of convergent olfactory receptor neurons*. J. Neurosci., 18, 4560–4569.

**Bramwell, A.F., Burrell, J.W.K.** and **Riezebos, G.** (1969) *Characterisation of pyrazines in galbanum oil*. Tetrahedron Lett., 37, 3215.

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

**Buttery, R.G., Ling, L.C.** and **Guadagni, D.G.** (1969) *Food volatiles—volatilities of aldehydes ketones and esters in dilute water solution*. J. Agric. Food Chem., 17, 385–389.

**Buttery, R.G., Teranishi, R., Ling, L.C.** and **Turnbaugh, J.G.** (1990) *Tomato aroma components—identification of glycoside hydrolysis volatiles*. J. Agric. Food Chem., 38, 336–340.

**Chothia, C.** and **Lesk, A.M.** (1986) *The relation between the divergence of sequence and structure in proteins*. EMBO J., 5, 823–826.

**Cronin, D.A.** and **Stanton, P.** (1976) *2-Methoxy-3-sec-butylpyrazine—important contributor to carrot aroma*. J. Sci. Food. Agric., 27, 145–151.

**Datta, D., Nagarajan, V., Floriano, W.B., Kim, K.S., Prasadarao, N.V.** and **Goddard, W.A. III** (2003) *Interaction of E-coli outer membrane protein A with sugars on the receptors of the brain microvascular endothelial cells*. Proteins Struct. Funct. Genet., 50, 213–221.

**Donnelly, D.** (1993) *Modeling alpha-helical transmembrane domains*. Biochem. Soc. Trans., 21, 36–39.

**Duchamp-Viret, P., Chaput, M.A.** and **Duchamp, A.** (1999) *Odor response properties of rat olfactory receptor neurons*. Science, 284, 2171–2174.

**Duke, J.A.** (1992) Handbook of Phytochemical Constituents of GRAS Herbs and Other Economic Plants. CRC Press, Boca Raton, FL.

**Eisenberg, D., Weiss, R.M.** and **Terwilliger, T.C.** (1984) *The hydrophobic moment detects periodicity in protein hydrophobicity*. Proc. Natl Acad. Sci. USA, 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., Singer, M.S., Shepherd, G.M.** and **Goddard, W.A. III** (2000) *Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor*. Proc. Natl Acad. Sci. USA, 97, 10712–10716.

**Forney, C.F.** (2001) *Horticultural and other factors affecting aroma volatile composition of small fruit*. Horttechnology, 11, 529–538.

**Freitag, J., Ludwig, G., Andreini, I., Rössler, P.** and **Breer, H.** (1998) *Olfactory receptors in aquatic and terrestrial vertebrates*. J. Comp. Physiol. A, 183, 635–650.

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

**Ghosh, A., Rapp, C.S.** and **Friesner, R.A.** (1998) *Generalized born model based on a surface integral formulation*. J. Phys. Chem. B, 102, 10983–10990.

**Gimelbrant, A.A., Stoss, T.D., Landers, T.M.** and **McClintock, T.S.** (1999) *Truncation releases olfactory receptors from the endoplasmic reticulum of heterologous cells*. J. Neurochem., 72, 2301–2311.

**Glusman, G., Bahar, A., Sharon, D., Pilpel, Y., White, J.** and **Lancet, D.** (2000) *The olfactory receptor gene superfamily: data mining, classification, and nomenclature*. Mamm. Genome, 11, 1016–1023.

**Gunata, Y.Z., Bayonove, C.L., Baumes, R.L.** and **Cordonnier, R.E.** (1985) *The aroma of grapes. I. Extraction and determination of free and glycosidically bound fraction of some grape aroma components*. J. Chromatogr., 331, 83–90.

**Hol, W.G.J., van Duijnen, P.T.** and **Berendsen, H.J.C.** (1978) *The alpha-helix dipole and the properties of proteins*. Nature, 273, 443–446.

**Humphrey, W., Dalke, A.** and **Schulten, K.** (1998) *VMD—visual molecular dynamics*. J. Mol. Graphics, 14, 33–38.

**Ivanova, T.T.** and **Caprio, J.** (1993) *Odorant receptors activated by amino acids in sensory neurons of the channel catfish* Ictalurus punctatus. J. Gen. Physiol., 102, 1085–1105.

**Jackson, J.F.** and **Linskens, H.F.** (eds) (2002) Analysis of Taste and Aroma. Springer, New York.

**Jordan, M.J., Tandon, K., Shaw, P.E.** and **Goodner, K.L.** (2001) *Aromatic profile of aqueous banana essence and banana fruit by gas chromatography–mass spectrometry (GC–MS) and gas chromatography–olfactometry (GC–O)*. J. Agric. Food Chem., 49, 4813–4817.

**Kekenes-Huskey, P.M., Vaidehi, N., Floriano, W.B.** and **Goddard, W.A. III** (2003) *Amino acid screening for phenylalanyl t-RNA synthetase*. J. Am. Chem. Soc., 107, 11549–11557.

**Klabunde, T.** and **Hessler, G.** (2002) *Drug design strategies for targeting G-protein-coupled receptors*. ChemBioChem, 3, 928–944.

**Krammer, G.E., Takeoka, G.R.** and **Buttery, R.G.** (1994) *Isolation and identification of 2,5-dimethyl-4-hydroxy-3(2H)-furanone glucoside from tomatoes*. J. Agric. Food Chem., 42, 1595–1597.

**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.

**Lancet, D.** and **Ben-Airie, N.** (1993) *Olfactory receptors*. Curr. Biol., 3, 668–674.

**Lavid, N. Schwab, W., Kafkas, E., Koch-Dean, M., Bar, E., Larkov, O., Ravid, U.** and **Lewinsohn, E.** (2002) *Aroma biosynthesis in strawberry: S-adenosylmethionine:furaneol O-methyltransferase activity in ripening fruits*. J. Agric. Food Chem., 50, 4025–4030.

**MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D.** and **Karplus, M.** (1998) *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. J. Phys. Chem. B, 102, 3586–3616.

**Maga, J.A.** and **Sizer, C.E.** (1973) *Pyrazines in foods—review*. J. Agric. Food Chem., 21, 22–30.

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

**Mamlouk, A.M., Chee-Ruiter, C., Hofmann, U.G.** and **Bower, J.M.** (2003) *Quantifying olfactory perception: mapping olfactory perception space by using multidimensional scaling and self-organizing maps*. Neurocomputing, 52, 591–597.

**Marti-Renom, M.A., Madhusudhan, M.S., Fiser, A., Rost, B.** and **Sali, A.** (2002) *Reliability of assessment of protein structure prediction methods*. Structure, 10, 435–440.

**Mathiowetz, A.M., Jain, A., Karasawa, N.** and **Goddard, W.A. III** (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, W.A. III** (1990) *Dreiding—a generic force-field for molecular simulations*. J. Phys. Chem., 94, 8897–8909.

**Mezler, M., Fleischer, J.** and **Breer, H.** (2001) *Characteristic features and ligand specificity of the two olfactory receptor classes from* Xenopus laevis. J. Exp. Biol., 204, 2987–2997.

**Mombaerts, P., Wang, F., Dulac, C., Chao, S.K., Nemes, A., Mendelsohn, M., Edmondson, J.** and **Axel, R.** (1996) *Visualizing an olfactory sensory map*. Cell, 87, 675–686.

**Mori, K., Nagao, H.** and **Yoshihara, Y.** (1999) *The olfactory bulb: coding and processing of odor molecule information*. Science, 286, 711–715.

**Murray, K.E., Shipton, J.** and **Whitfield, F.B.** (1970) *2-Methoxypyrazines and flavour of green peas (*Pisum-satium*)*. Chem. Ind. Lond., 27, 897–000.

**Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E.M.** and **Shichida, Y.** (2002) *Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography*. Proc. Natl Acad. Sci. USA, 99, 5982–5987.

**Ohloff, G.** (1978) *Recent developments in the field of' naturally occurring aroma components*. Prog. Chem. Org. Nat. Prod., 35, 431–527.

**Pabst, A., Barron, D., Etievant, P.** and **Schreier, P.** (1991) *Studies on the enzymatic hydrolysis of bound aroma constituents from raspberry fruit pulp*. J. Agric. Food Chem., 39, 173–175.

**Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M.** and **Miyano, M.** (2000) *Crystal structure of rhodopsin: a G protein-coupled receptor*. Science, 289, 739–745.

**Pilpel, Y.** and **Lancet, D.** (1999) *The variable and conserved interfaces of modeled olfactory receptor proteins*. Protein Sci., 8, 969–977.

**Rhee, M.-H., Nevo, I., Bayewitch, M.L., Zagoory, O.** and **Vogel, Z.** (2000) *Functional role of tryptophan residues in the fourth transmembrane domain of the CB2 cannabinoid receptor*. J. Neurochem., 75, 2485–2491.

**Rubin, B.D.** and **Katz, L.C.** (1999) *Optical imaging of odorant representations in the mammalian olfactory bulb*. Neuron, 23, 499–511.

**Sanz, C., Richardson, D.G.** and **Perez, A.G.** (1995) *2,5-Dimethyl-4-hydroxy-3(2h)-furanone and derivatives in strawberries during ripening*. ACS Symp. Ser., 596, 268–275.

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

**Shi, L.** and **Javitch, J.A.** (2002) *The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop*. Annu. Rev. Pharm. Toxicol., 42, 437–467.

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

**Singer, M.S., WeisingerLewin, Y., Lancet, D.** and **Shepherd, G.M.** (1996) *Positive selection moments identify potential functional residues in human olfactory receptors*. Recept. Chan., 4, 141–147.

**Strader, C.D., Fong, T.M., Tota, M.R., Underwood, D.** and **Dixon, R.A.F.** (1994) *Structure and function of G protein-coupled receptors.* Annu. Rev. Biochem., 63, 101–132.

**Suriyaphan, O., Drake, M., Chen, X.Q.** and **Cadwallader, K.R.** (2001) *Characteristic aroma components of British farmhouse cheddar cheese*. J. Agric. Food Chem., 49, 1382–1387.

**Teller, D.C., Okada, T., Behnke, C.A., Palczewski, K.** and **Stenkamp, R.E.** (2001) *Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs)*. Biochemistry, 40, 7761–7772.

**Teranishi, R., Buttery, R.G.** and **Sugisawa, H.** (eds) (1992) Bioactive Volatile Compounds from Plants. American Chemical Society, Washington, DC.

**Thompson, J.D., Higgins, D.G.** and **Gibson, T.J.** (1994) *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice*. Nucleic Acids Res., 22, 4673–4680.

**Vaidehi, N., Jain, A.** and **Goddard, W.A. III** (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., Choi, E.J., Zamanakos, G.** and **Goddard, W.A. III** (2002) *Prediction of structure and function of G-protein coupled receptors*. Proc. Natl Acad. Sci. USA, 99, 12622–12627.

**Vriend, G.** (1990) *What if—a molecular modeling and drug design program*. J. Mol. Graph., 8, 52–56.

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

**Wieland, H.A., Willim, K.D., Entzeroth, M., Wienen, W., Rudolf, K., Eberlein, W., Engel, W.** and **Doods, H.N.** (1995) *Subtype selectivity and antagonistic profile of the nonpeptide y1 receptor antagonist BIBP-3226*. J. Pharmacol. Exp. Ther., 275, 143–149.

**Wu, P., Kuo, M.-C., Hartmann, T.G., Rosen, R.T.** and **Ho, C.-T.** (1991) *Free and glycosidically bound aroma compounds in pineapple (*Ananas comosus *L. Merr.)*. J. Agric. Food Chem., 39, 170–172.

**Zamanakos, G.** (2002) *A fast and accurate analytical method for the computation of solvent effects in molecular simulations*. PhD thesis, California Institute of Technology, Pasadena, CA.

**Zhang, X.M.** 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.

**Zozulya, S., Echeverri, F.** and **Nguyen, T.** (2001) *The human olfactory receptor repertoire*. Genome Biol., 2, 0018.1–0018.12.

*Accepted February 10, 2004*