

Ligand specificity of odorant receptors

Kamil Khafizov · Claudio Anselmi · Anna Menini ·
Paolo Carloni

Received: 31 March 2006 / Accepted: 12 September 2006 / Published online: 21 November 2006
© Springer-Verlag 2006

Abstract Odorant receptors belong to class A of the G protein-coupled receptors (GPCRs) and detect a large number of structurally diverse odorant molecules. A recent structural bioinformatic analysis suggests that structural features are conserved across class A of GPCRs in spite of their low sequence identity. Based on this work, we have aligned the sequences of 29 ORs for which ligand binding data are available. Recent site-directed mutagenesis experiments on one such receptor (MOR174-9) provide information that helped to identify nine amino-acid residues involved in ligand binding. Our modeling provides a rationale for amino acids in equivalent positions in most of the odorant receptors considered and helps to identify other amino acids that could be important for ligand binding. Our findings are consistent with most of the previous models and allow predictions for site-directed

mutagenesis experiments, which could also validate our model.

Keywords Olfactory receptors · Sequence alignment · Odorant binding · G protein-coupled receptor · Homology modeling

Abbreviations

OR	odorant receptor
3D	three-dimensional
SI	sequence identity
GPCR	G protein-coupled receptor
TM	transmembrane

Introduction

The olfactory system detects and distinguishes among a large number of structurally diverse odorant molecules. In the nasal cavity of vertebrates, volatile molecules interact with odorant receptors (ORs), which are located in the cilia of olfactory sensory neurons. The odorant binding to ORs activates a transduction cascade that leads to the production of action potentials, which are transmitted to the brain (for review, see Menini et al. [1]).

Mammalian ORs were first identified in 1991 by Buck and Axel [2]. Because of the difficulty in expressing ORs on the membrane surface of heterologous cells, to date only a limited number of relations between ORs and odorant molecules could be provided (for reviews see Mombaerts [3] and Godfrey et al. [4]). One odorant can activate numerous types of ORs, while a single OR can be activated by several different odorants [5]. Each type of odorant molecule is recognized by the activation of a unique combination of ORs in a combinatorial code. ORs have

Electronic supplementary material Supplementary material is available in the online version of this article at <http://dx.doi.org/10.1007/s00894-006-0160-9> and is accessible for authorized users.

K. Khafizov · C. Anselmi · A. Menini · P. Carloni
International School for Advanced Studies,
via Beirut 4,
I-34014 Trieste, Italy

K. Khafizov · C. Anselmi · P. Carloni (✉)
INFN-DEMOCRITOS Modelling Centre for Research
in Atomistic Simulation,
via Beirut 4,
I-34014 Trieste, Italy
e-mail: carloni@sissa.it

K. Khafizov · C. Anselmi · A. Menini · P. Carloni
Italian Institute of Technology,
via Beirut 4,
I-34014 Trieste, Italy

been subdivided in two classes: [6, 7] Class #1 ORs, the so-called fish-like ORs, which share a rather high sequence identity (SI) and Class #2 ORs, mammalian-like ORs, which comprise all the others. Class #1 ORs bind structurally similar ligands [5, 8] such as aliphatic acids, alcohols, aldehydes and so on, whereas Class #2 ORs bind structurally divergent odorants [5, 8–19].

ORs belong to Class A of the GPCR superfamily, the so-called Rhodopsin-like Class [20]. They feature seven transmembrane (TM) helices in the membrane domain, an extracellular N-terminus and an intracellular C-terminus [20]. Resting on the significant divergence of SI and on the results obtained by site-directed mutagenesis experiments, helices TM3, TM4, TM5 and TM6 are believed to form the binding pocket [1, 14, 21]. Since there is still a lack of experimental structures at atomic level, structural information on ORs is essentially based on computational methods.

As for the prediction of the entire transmembrane domain, structural models of some ORs were based on the bacteriorhodopsin X-ray structure [22–24] or on a low-resolution (7.5 Å) electron-microscopy map of rhodopsin [21, 25–29]. Molecular-dynamics simulations of one of these models, the rat OR I7, [28] helped to identify the dissociative pathway along which the ligand accesses the pocket [30]. Models of ORs (MOR31-2, MOR32-4, MOR33-1, MOR42-1, MOR42-3, MOR103-15, MOR175-1 and MOR204-32) were also predicted using molecular-simulation based protocols [31–34]. These models were structurally rather different from each other and from rhodopsin, although several positions in TM3 and TM6, which are involved in ligand binding, are conserved across the family. The binding pockets of these models turned out to be constituted by helices TM3 to TM7 along with extracellular loops. Free-energy calculations based on these models turned out to be in agreement with experimentally-determined binding affinities [9–11, 15].

Man et al. [35] suggested that residues that might be involved in ligand binding in ORs are located on helices TM2 to TM7 and on the second extracellular loop. Their work was based on a comparison between mouse and human OR sequences, under the assumption that functional contact residues would be conserved among pairs of orthologous receptors, but considerably less conserved among paralogous pairs [35]. Recently, Katada et al. [14] published a study that combines computational biology with molecular biology tools. They performed several site-directed mutagenesis experiments for MOR174-9 OR to identify residues involved in ligand binding and selectivity and recognized nine crucial amino acids—Ser113, Phe206, Asn207, Thr211, Leu212, Phe252, Thr255, Ile256 and Leu259—located on helices TM3, TM5 and TM6. Models

based on the bovine rhodopsin X-ray structure, [36] which is the only GPCR for which an X-ray structure has been determined, were fully consistent with the experimental results [14].

The reliability of homology models for ORs has been questioned, based on first-principles simulations and the low SI between rhodopsin and ORs (in the case of those considered in this work it ranges between 14 and 20%) [31–34]. However, homology models with low SI (ranging from 10 to 30%), based on rhodopsin and other templates, have been reported for a variety of membrane proteins [37–39]. These models provided insights into structure/function relationships for these systems. In addition, the recent observation that structural features are well conserved across all class A GPCRs [40], which include ORs and rhodopsin, has led to the conclusion that OR rhodopsin-based structural predictions may be reliable, strengthening a posteriori the conclusions of Katada et al. [14].

Prompted by these studies, we have here extended the rhodopsin-based modeling to 29 ORs for which ligand affinity data have been measured (Table 1). Because of the aforementioned structural conservation, we postulate that: (1) the presence of conserved residues in ORs binding ligands possessing the same functional groups may provide hints on their functional role, in spite of the low SI across the family. This assumption is further substantiated a posteriori by an analysis of the sequences, which shows that, based on this assumption, ORs binding to specific ligands feature conserved residues that bind to those ligands. (2) The relative orientation of the seven TM helices is similar across the OR family, in spite of the diversity of their sequence [2, 5, 35].

In addition, the structural information for MOR174-9 [14] may be included in the modeling for most ORs (Table 1), given the structural conservation among members of GPCR family [40]. Our calculations permit us to identify a number of residues located on helices TM3, TM4, TM5 and TM6, which may play a role in odorant binding for 23 out of the 29 ORs considered (Fig. 1 and Table 1). For these receptors, mutations in the putative binding site were also suggested. These mutations could be used to validate the model against molecular biology experiments.

Materials and methods

Sequence alignments

The sequences of mouse ORs for which binding affinity data are available [5, 8, 10–18] (Fig. 2), were aligned using the ClustalW multiple alignment program [41].

Table 1 Residues in the putative binding pocket of mouse ORs identified by this study

Class of odorants	Total number of carbons in aliphatic chain (<i>N</i>)	ORs	Main residues in binding pocket
c	9	MOR8-2 (S86)	3 (H), 4 (S), 5 (A), VI (F), VII (P), VIII (V), IX (L)
	7–9	MOR22-2 (S41)	3 (H), 4 (Q), II (T), III (G), VII (P), 13 (F)
	7–9	MOR23-1	3 (H), 4 (Q), II (A), III (I), VII (L), 13 (F)
	5–6	MOR31-2 (S18)	3 (H), 4 (F), 5 (V), 9 (S), II (A), VI (F), VII (P), 13 (F)
	6–10	MOR31-4	3 (H), 4 (F), 5 (I), 9 (S), II (L), VI (F), VII (P), 13 (F)
	5	MOR31-6	3 (H), 4 (F), 5 (V), 9 (S), II (V), III (M), VI (F), VII (P), 13 (F)
	9–10	MOR32-4 (S46)	3 (H), <i>II</i> (S), III (M), V (V), VII (P), 13 (F)
	10	MOR32-5	3 (H), <i>II</i> (S), III (V), V (L), VII (P), 13 (F)
	8–10	MOR32-11	3 (H), <i>II</i> (S), V (M), VII (P), 13 (F)
	6–9	MOR33-1 (S19)	3 (H), 4 (F), 5 (A), 9 (T), III (V), VII (P), 13 (F)
	7–9	MOR40-1 (S51)	3 (N), 4 (L), 5 (T), III (L), VII (V), 13 (V)
	8–9	MOR40-4 (S83)	3 (H), 4 (V), 5 (A), III (L), VII (V), 13 (V)
	7–9	MOR106-1 (S1)	5 (T), II (F), III (V), VIII (V), 13 (V), IX (M) or 3 (F), II (F), III (V), <i>VII</i> (T), 13 (V), IX (M)
	9	MOR203-1	1 (L), 2 (V), 9 (I), 10 (F), VI (F), <i>VII</i> (T), 13 (Y) or 2 (V), 4 (V), 5 (T), I (C), 10 (F), II (F), <i>III</i> (N), 12 (I), VIII (L)
	a	8–9	MOR22-2 (S41)
7–9		MOR33-1 (S19)	3 (H) or 9 (T), III (V), VII (P), 13 (F)
8–9		MOR40-1 (S51)	3 (N) or 5 (T), 4 (L), III (L), VII (V), 13 (V)
9		MOR40-4 (S83)	3 (H), 4 (V), 5 (A), III (L), VII (V), 13 (V)
5–7		MOR106-13P (S3)	3 (F), 5 (S), I (C), II (L), III (P), 13 (V), IX (M) or 3 (F), I (C), II (L), III (P), 12 (L), <i>VII</i> (S), 13 (V), IX (M)
6–7		MOR204-32 (S25)	3 (V), 5 (T), III (I), VIII (I), 13 (F), IX (I) or 3 (V), III (I), 11 (L), <i>VII</i> (T), 13 (F) or <i>I</i> (S), III (I), V (V), VI (F), VIII (I)
d	6–9	MOR13-6 (S85)	3 (H), 4 (S), III (A), VII (P), 13 (S), IX (V)
	9	MOR42-1 (S50)	3 (Y), 4 (I), 5 (V), II (F), 12 (T) or <i>VII</i> (S) or 13 (S) or IX (S)
	9	MOR42-3 (S6)	3 (Y), 4 (I), 5 (V), 9 (T), II (F), 12 (T) or <i>VII</i> (S) or 13 (S) or IX (S)
e	5	MOR31-6	2 (V), II (V), III (M), VI (F), <i>VIII</i> (S), 14 (L)
	6–10	MOR103-15 [20]	2 (F), 3 (L), 5 (C), I (C), 9 (I), II (F), III (I), VI (F), <i>VIII</i> (S), IX (I)
g	7	MOR175-1 (OR912-93)	2 (F), 3 (S), 5 (V), II (I), III (L), VI (L), VIII (L), IX (I)
		MOR171-2 (M71)	2 (F), 3 (L), 4 (V), 5 (I), I (C), 9 (F), II (F), VI (F), <i>VIII</i> (T), IX (M)
		MOR258-5	2 (F), 3 (L), 4 (G), 5 (I), 9 (I), II (I), III (L), VI (M), <i>VIII</i> (C), IX (M) or 2 (F), 3 (L), 4 (G), 5 (I), <i>I</i> (S), 9 (I), 10 (L), II (I), III (L), VI (M), IX (M)

The class of odorants, the total number of carbon atoms in the ligand, the names of the receptors are here reported using the same labelling schemes of Fig. 2. Commonly used OR names are reported in parenthesis. Positions of residues likely to be involved in ligand–protein hydrogen bonds are in italics.

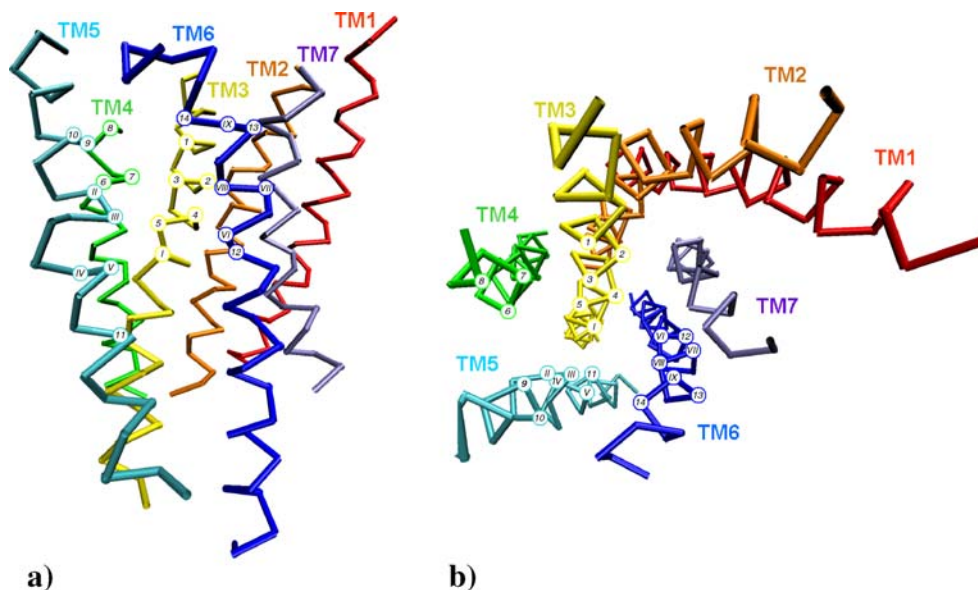
Three-dimensional structural models

Three-dimensional (3D) structural models of ORs (Fig. 1) were constructed using the MODELLER 6.2 program [42]. The calculations were based on the X-ray structure of bovine rhodopsin [43]. Sequences of ORs were aligned against the rhodopsin sequence using ClustalW [41]. The experimental information about residues that bind eugenol in the MOR174-9 receptor [14] was included into the alignment. To achieve this, (1) we considered the positions of the residues known to bind eugenol in the MOR174-9 receptor, which are believed to

face the active-site cavity [14]. (2) We looked at the alignments of MOR174-9 with the other receptors and identified the corresponding positions of the latter proteins. These positions were assumed to belong to residues facing the active-site cavity. (3) The alignments of OR sequences against bovine rhodopsin sequence were modified manually according to this information. The alignments are available as supplementary material.

Energy minimization based on the parm99 AMBER force field [44] was carried out for all the 3D-models investigated here. The analysis presented in the **Results and discussion** section refers to such models.

Fig. 1 Side (a) and top (b) views of a representative structural model of ORs (C_{α} carbons in TM helices are only shown). The positions of the residues that may be involved in ligand binding are here indicated. Nine of them, indicated by *I* to *IX*, coincide with those previously obtained for MOR174-9 receptor [14]. Figures were created with VMD 1.8.2 program. [47]



Results and discussion

Comparisons among ORs

Here we attempt to provide a rationale for ligand binding to mouse ORs (Fig. 2) by first investigating the role of the nine residues, identified as constituents of the binding site in MOR174-9 (Ser113, Phe206, Asn207, Thr211, Leu212, Phe252, Thr255, Ile256 and Leu259) [14]. For purposes of clarity, we will refer to the positions occupied by these residues, as well as by the residues in equivalent positions in the other ORs, as positions *I–IX* throughout the text. We constructed approximate 3D-models for the 15 Class #1 ORs (Fig. 2) and the 13 Class #2 ORs (MOR103-15, MOR106-1, MOR106-13P, MOR171-2, MOR175-1, MOR203-1, MOR204-32, MOR258-5). However, for ORs MOR118-1, MOR136-6, MOR174-4, MOR267-13 and MOR276-1, the experimentally available information and/or the similarity with other ORs is not sufficient to make reliable predictions for the binding pocket regions. Consequently, these ORs were not considered further.

We will first focus on Class #1 ORs, which share some common ligands, and later on Class #2 ORs, whose ligands are more variable.

Class #1 ORs

The odorant specificities of 15 Class #1 ORs have been tested experimentally [5, 8]. Aliphatic carboxylic acids were found to bind to 12 of the 15 Class #1 ORs, while dicarboxylic acids bind to the remaining three: MOR42-1, MOR42-3 and MOR13-6 (Table 1). We investigate here the role of residues in positions *I–IX* in the binding of aliphatic

carboxylic and dicarboxylic acids, in which the charged groups are at the ends of the chains, for the experimentally tested Class #1 ORs.

Position *IV* always has an Asp group. We suggest, at a speculative level, that this Asp could interact with a basic position in TM4 (indicated by a red arrow in Fig. 2) conserved across this class of ORs. Thereby, this position may serve structural and/or functional purposes, as previously noticed by Malnic et al. [5] Position *I* has polar residues in most ORs, with the exception of MOR42-1 and

Fig. 2 Sequences of all mouse ORs, for which ligand affinity data have been measured [5, 8, 10–18]. In the absence of a commonly agreed recent nomenclature for ORs, we have used the nomenclature proposed by Zhang and Firestein [7], in which Class #1 OR subfamilies are given numbers lower than 100 and Class #2 OR subfamilies are given numbers higher than 100. (For a comparison among different OR names see Table 5 by Godfrey et al. [4]). Only the sequences corresponding to helices TM3, TM4, TM5 and TM6 are here reported. Basic, acid, polar and Cys residues are colored in blue, red, green and yellow, respectively. Residues that have been shown to bind eugenol in MOR174-9 [14] are reported inside black boxes. Positions *I* to *IX* and *I* to *14*, discussed in the text, are labelled. The position in TM4, mentioned in the text, is indicated with a red arrow above the alignment. Classes and specific ligands, which bind to corresponding ORs, are labelled on the right side. These are: *a* Alcohols, *c* carboxylic acids, *b* bromocarboxylic acids, *d* dicarboxylic acids, *e* alkylic aldehydes, *f* ketones, *g* benzaldehyde, *h* citronellal, *i* eugenol, *j* carvone, *k* acetophenone, *l* piperonal, *m* ethyl-vanillin, *n* vanillin, *o* citral, *p* limonene, *q* coumarin, *r* isovaleric acid, *s* cinnamaldehyde, *t* 2-octenal, *u* lylal, *v* chromanone, *w* 2-coumaranone [5, 8, 10–18]. Small differences in the response profiles of ORs to odorants have been found using different experimental systems by Malnic et al. [5] and Saito et al. [8] In case of discrepancy, we used the most recent experimental results obtained by Saito et al. [8] because their newly developed system of OR expression permits to perform site-directed mutagenesis experiments and to functionally test the odorant ligand specificity, whereas the pioneer studies by Malnic et al. [5] were obtained on isolated olfactory sensory neurons

TM3

		<i>1</i>	<i>23</i>	<i>45</i>	<i>1</i>																													
MOR8-2	S	P	N	E	C	F	A	L	F	F	I	H	F	S	A	M	E	S	S	V	L	L	I	M	S	F	R	V	I	A	I	128		
MOR13-6	A	F	D	A	C	L	A	M	F	F	F	H	F	F	F	S	W	T	E	S	G	I	L	L	A	M	S	F	R	V	A	I	128	
MOR22-2	S	F	D	A	C	L	L	O	M	L	I	H	S	F	O	A	F	E	S	G	I	L	L	A	M	A	F	R	V	A	I	128		
MOR23-1	A	F	D	A	C	L	T	O	L	F	F	I	H	A	L	O	G	M	E	S	G	V	L	L	A	M	A	F	R	V	A	I	128	
MOR31-2	P	F	D	A	C	I	A	O	M	F	F	I	H	V	A	F	V	A	E	S	G	I	L	L	A	M	A	F	R	V	A	I	134	
MOR31-4	S	L	D	R	C	I	T	O	L	F	F	I	H	S	T	F	I	S	E	S	G	I	L	L	V	M	A	F	R	V	I	A	I	128
MOR31-6	A	F	D	A	C	V	T	O	V	F	F	V	H	M	F	V	G	E	S	A	I	L	L	A	M	A	F	R	F	I	A	I	128	
MOR32-4	A	F	R	D	C	V	A	O	M	F	F	I	H	L	F	T	G	I	E	T	F	M	L	V	A	M	A	F	R	V	I	A	I	128
MOR32-5	A	F	G	A	C	I	T	O	M	Y	I	I	H	T	C	T	G	L	E	S	V	V	L	L	I	M	A	F	R	V	I	A	I	128
MOR32-11	S	F	G	G	C	L	A	O	M	F	F	I	H	V	F	T	G	M	E	T	V	L	L	V	M	A	F	R	F	V	A	I	128	
MOR33-1	S	Y	H	G	C	L	T	O	M	F	F	V	H	A	V	F	A	T	E	S	A	V	L	L	A	M	A	F	R	V	A	I	154	
MOR40-1	G	F	S	S	C	F	L	O	M	F	V	M	N	S	F	L	T	M	S	C	I	F	V	C	M	A	I	R	V	A	I	129		
MOR40-4	S	L	P	R	C	L	O	M	Y	A	I	H	C	F	V	A	M	E	S	G	I	F	V	C	M	A	I	R	V	A	I	132		
MOR42-1	S	L	P	R	C	L	O	M	F	C	I	Y	F	L	I	V	F	D	C	N	I	L	L	V	M	A	I	R	V	A	I	130		
MOR42-3	S	L	P	R	C	L	O	M	F	C	I	Y	F	L	I	V	F	D	C	N	I	F	V	C	M	A	I	R	V	A	I	130		
MOR103-15	S	F	A	C	M	T	O	L	Y	F	F	L	G	L	C	T	E	C	V	L	L	A	V	M	A	F	R	V	A	I	131			
MOR106-1	S	F	S	G	C	F	L	O	F	Y	F	F	F	S	L	G	T	T	E	C	L	F	L	A	V	M	A	F	R	V	L	A	I	136
MOR106-13P	S	F	S	G	C	F	L	O	F	Y	F	F	F	S	L	G	S	T	E	C	F	L	A	V	M	A	F	R	V	L	A	I	130	
MOR118-1	S	L	L	G	C	F	L	O	A	F	L	Y	F	F	L	G	T	T	E	F	L	L	A	V	M	S	F	R	V	A	I	124		
MOR136-6	S	H	A	G	C	V	S	C	I	Y	F	F	L	F	G	C	I	D	N	F	L	L	S	M	A	F	R	V	A	I	126			
MOR171-2	S	Y	V	G	C	M	S	O	L	Y	F	F	L	V	F	V	I	A	E	C	M	L	T	V	M	A	F	R	V	A	I	126		
MOR174-4	S	F	V	G	C	I	V	O	Y	F	L	C	V	F	V	V	T	E	A	F	L	L	V	M	A	F	R	F	V	A	I	127		
MOR174-9	S	F	V	G	C	L	V	O	Y	F	F	F	F	C	T	F	V	V	T	E	S	F	L	L	V	M	A	F	R	F	V	A	I	127
MOR175-1	S	F	L	G	C	A	I	O	L	C	L	F	S	O	F	V	V	T	E	C	F	L	L	A	S	M	A	F	R	V	A	I	126	
MOR203-1	S	F	T	G	C	L	A	O	L	L	V	Y	G	I	F	V	P	S	D	C	F	I	L	A	A	M	A	F	R	V	A	I	126	
MOR204-32	P	V	A	G	C	V	A	L	C	I	V	V	A	F	T	S	E	S	F	L	L	A	S	M	A	F	R	V	A	I	156			
MOR258-5	S	Y	V	G	C	V	A	O	Y	I	F	L	M	L	G	I	T	E	S	W	L	F	A	I	M	A	F	R	V	A	I	129		
MOR267-13	S	L	A	G	C	A	O	M	F	F	F	V	L	A	T	N	N	C	F	L	L	T	A	M	A	F	R	V	A	I	126			
MOR276-1	S	F	A	G	C	I	O	I	G	F	F	V	L	V	G	S	F	L	L	L	G	L	M	A	F	R	V	A	I	95				

TM4

				<i>67</i>	<i>8</i>																				
MOR8-2	S	A	R	V	L	O	I	G	L	A	F	S	L	K	N	V	L	L	I	L	P	F	V	163	c
MOR13-6	L	A	R	V	A	H	M	G	M	S	I	I	R	S	F	C	M	V	F	P	L	P	F	163	b, d
MOR22-2	P	O	L	T	T	C	L	G	A	L	L	R	S	L	I	T	F	P	L	I	L	163	c, a, b		
MOR23-1	P	L	F	L	R	V	V	L	V	A	I	R	A	T	V	L	V	G	V	L	P	I	163	c	
MOR31-2	P	M	A	I	G	K	M	T	L	A	I	W	G	R	S	I	G	T	I	F	P	I	I	169	c
MOR31-4	N	S	L	I	G	K	I	G	V	G	I	F	L	R	S	Y	G	T	I	F	P	I	I	163	c
MOR31-6	W	S	T	V	G	R	I	A	L	A	I	V	R	S	C	I	I	F	P	V	I	F	163	c, r, e	
MOR32-4	N	R	I	C	I	I	V	G	V	L	F	K	N	F	I	L	V	F	P	L	I	F	163	c, b	
MOR32-5	N	K	V	I	A	I	L	G	I	I	I	V	R	T	L	I	F	V	T	P	F	I	F	163	c
MOR32-11	N	K	I	S	L	L	S	V	V	I	G	R	N	L	I	L	V	I	P	F	V	F	163	c	
MOR33-1	A	V	V	I	C	K	I	G	L	A	C	V	T	R	G	L	L	F	V	F	P	F	V	189	c, a, b
MOR40-1	D	O	F	V	V	R	A	A	I	F	V	A	A	R	N	G	L	I	M	P	I	P	I	162	c, a, b
MOR40-4	E	S	F	V	V	K	A	T	V	I	M	A	I	R	N	F	V	A	P	M	S	V	P	167	c, a, b
MOR42-1	G	O	L	L	A	G	L	V	L	A	V	T	R	S	T	S	I	V	A	P	V	V	165	d	
MOR42-3	G	O	L	L	A	G	L	V	L	A	V	T	R	S	T	S	I	V	A	P	V	V	165	d	
MOR103-15	S	R	L	C	V	Q	M	A	A	G	S	W	A	G	G	F	G	I	S	M	V	K	V	166	e, o, h, s, t
MOR106-1	R	R	L	C	C	I	L	V	S	S	C	W	L	I	G	F	L	S	P	I	P	I	F	171	c, b
MOR106-13P	G	R	L	C	N	I	L	V	S	C	W	I	L	G	F	L	F	P	V	P	I	I	165	a	
MOR118-1	K	R	V	C	V	L	V	F	C	S	W	M	S	G	L	L	L	I	I	V	P	S	159	h	
MOR136-6	Q	S	L	C	V	L	L	V	M	V	S	W	A	F	S	S	N	G	L	V	H	L	161	j	
MOR171-2	P	A	L	C	S	L	L	V	A	F	V	A	V	G	L	I	S	A	I	T	G	160	k, g		
MOR174-4	Q	K	L	C	I	L	V	V	G	S	Y	A	W	G	F	T	C	S	L	L	T	C	162	m, n	
MOR174-9	Q	R	L	C	A	M	L	V	L	G	S	Y	A	W	G	V	C	S	L	L	T	C	162	i, m, n	
MOR175-1	Q	R	V	C	V	L	V	I	G	P	S	I	G	F	I	S	T	M	V	H	I	161	f, e		
MOR203-1	Q	R	L	C	I	L	L	L	G	S	Y	T	M	G	F	L	N	A	T	V	N	T	161	c	
MOR204-32	S	T	V	C	I	L	L	V	G	T	S	Y	L	G	G	W	N	A	W	I	F	T	191	a	
MOR258-5	P	L	R	G	S	L	V	A	F	C	G	F	W	G	I	T	C	A	L	I	Y	T	164	q, l, g, v, w	
MOR267-13	K	G	M	C	A	L	L	V	C	S	L	G	T	G	L	M	A	V	L	H	V	P	160	u	
MOR276-1	Q	K	V	C	L	I	A	G	S	W	A	F	G	L	I	G	I	O	M	V	130	p			

TM5

		<i>10</i>	<i>III</i>	<i>V</i>																							
		<i>9</i>	<i>II</i>	<i>IV</i>	<i>II</i>																						
MOR8-2	E	N	V	I	Y	C	L	F	V	A	L	T	G	I	L	I	T	F	I	F	M	S	Y	A	L	221	
MOR13-6	I	N	S	M	Y	G	L	F	I	V	I	S	A	F	V	S	V	L	I	L	L	S	Y	V	L	222	
MOR22-2	I	N	K	I	C	G	L	L	V	A	F	A	I	L	G	F	I	V	F	I	F	S	Y	V	R	222	
MOR23-1	I	N	K	S	Y	G	L	F	V	A	F	A	I	L	G	F	I	V	F	I	S	Y	I	L	222		
MOR31-2	V	N	I	W	Y	C	F	S	V	P	M	A	S	V	L	V	I	V	A	L	I	S	Y	T	L	228	
MOR31-4	V	N	I	W	Y	C	F	S	V	L	L	T	V	V	L	V	V	L	I	F	V	S	Y	V	L	222	
MOR31-6	V	N	I	W	Y	C	F	S	V	P	I	V	M	V	I	V	I	V	L	I	A	V	S	Y	L	222	
MOR32-4	V	N	V	I	F	G	L	-	I	L	S	M	L	L	V	L	I	A	L	S	Y	A	K	221			
MOR32-5	V	N	V	I	Y	G	L	-	V	A	F	S	V	G	I	L	S	V	I	F	S	Y	I	R	221		
MOR32-11	I	N	I	Y	G	L	-	V	V	I	S	H	I	L	V	M	I	L	I	A	S	Y	V	L	221		
MOR33-1	P	N	T	I	Y	G	L	I	V	A	L	S	V	T	G	M	V	L	I	A	T	S	Y	I	L	248	
MOR40-1	E	N	K	L	Y	O	F	V	I	G	W	T	L	L	G	S	L	I	L	I	V	L	S	Y	F	222	
MOR40-4	I	N	S	I	N	O	L	F	L	A	W	T	I	M	G	S	L	A	L	I	M	I	S	Y	L	226	
MOR42-1	L	N	K	T	A	G	L	I	I	R	T	F	N	R	V	L	M	L	L	L	G	T	S	Y	S	R	224
MOR42-3	L	N	K	T	V	E	L	I	V	R	I	F	N	R	V	L	M	L	L	L	G	A	S	Y	S	R	224
MOR103-15	T	A	L	L	T	F	I	L	A	I	F	I	L	L	G	P	L	S	V	T	G	A	S	Y			

MOR42-3, which have a Cys in position *I*. Most ORs show polar residues at position *IX* and at least one polar residue in positions *II* and *III*. Therefore, some of the residues in positions *I–IV* and *IX* might serve as polar/charged anchors for the carboxylate group present in the odorants. By contrast, positions *V–VIII* are mostly occupied by apolar residues, except for MOR42-1 and MOR42-3, in which polar residues are mostly present. As a result, positions *V–VIII* might form a hydrophobic pocket to accommodate the aliphatic tail of the ligands, except for MOR42-1 and MOR42-3, in which the polar pocket might bind the additional carboxylate group present in dicarboxylic acids. In MOR13-6, which also binds to dicarboxylic acids, the only polar residue that may bind the second carboxylate group is the Ser residue in position *V*, in addition to positions *I* and *II*. Nevertheless, at present, it is difficult to state whether this Ser residue is involved in the binding.

Class #2 ORs

Ligands binding to Class #2 ORs exhibit a very large structural diversity. In the experimentally tested receptors, positions *I*, *II* and *IX* are occupied mostly by apolar groups, positions *IV* and *VII* by polar groups, whilst the chemical nature of positions *III*, *V*, *VI* and *VIII* is similar to that of Class #1 ORs. However, as expected from the great variability of Class #2 ligands, no significant common features could be identified for these positions.

Models built on rhodopsin's structure

By constructing 3D-models of ORs, we could identify 14 other positions (hereafter *I–14* for the sake of clarity) located on helices TM3, TM4, TM5 and TM6, which face the active-site cavity and, therefore, may play an important role in binding the odorants (Fig. 1). In this paper we analyze the possible role of all the residues we identified as putative constituents of the binding pocket and the ligand specificity for some ORs (Table 1 and Fig. 2). The main residues forming the binding pocket are summarized in Table 1. Residues that may form H-bond with ligands are in italics. As our analysis is based on static models, at present it is not possible to identify if some additional residues exist that are not located in the cavity but still might participate to the binding in a dynamic manner.

Most of the Class #1 ORs that bind to *n*-aliphatic carboxylic acids have a His residue in position 3, except for MOR40-1 (Asn), MOR42-1 and MOR42-3 (Tyr), (Fig. 2). Some of these ORs also bind to *n*-aliphatic alcohols. His in position 3 could form an H-bond with the polar group of the ligands. As can be seen from Fig. 1, usually there is a second polar residue present (positions 4, 5 and 9) that can form a second H-bond with the carboxylate group of

carboxylic acids. The aliphatic tail of these ligands might be surrounded by apolar residues in positions *10*, *II* and *III* at TM5 and *12*, *VI*, *VII*, *VIII*, *13* and *IX* at TM6.

The main difficulties in predicting which residues are responsible for binding carboxylic acids and/or alcohols are expected for Class #2 ORs: MOR106-1, MOR106-13P, MOR203-1 and MOR204-32. They show sequence diversity and therefore, as can be seen from Table 1, the distribution of polar and/or charged residues in the binding sites is very different within the class and with respect to Class #1 ORs. As a result, it is not possible to formulate a unique model of the binding pocket. For MOR106-1 and MOR203-1, which bind carboxylic acids, we can propose two different binding models. As for MOR106-1, the ligand carboxylate group can bind either Thr in positions 5 or *VII*. In the MOR203-1 case, one possibility is that Thr in position *VII* and Tyr in position 13 bind the odorant carboxylate group; the other is that the ligand is coordinated by Thr in position 5 and Asn in position *III*.

MOR106-13P and MOR204-32 bind alcohols. The first could bind the ligand hydroxyl group with Ser either in position 5 or *VII*. The second could bind the ligand either with Thr in position 5 or *VII*, or with Ser in position *I*.

MOR42-1 and MOR42-3, which bind aliphatic dicarboxylic acids, exhibit polar residues in positions 3 at TM3, and in positions *11*, *VII*, *13*, *IX* at TM6. MOR42-3 also shows a polar residue in position 9 at TM5. Thereby, we may assume that there are two available sites for the binding of the two carboxylate groups. MOR13-6, which also binds aliphatic dicarboxylic acids, exhibits polar residues in positions 5 at TM3 and 13 at TM6. These also could be two sites for binding the two carboxylate groups.

Some ORs that bind aliphatic or aromatic aldehydes, such as MOR103-15, MOR31-6 and MOR171-2 (Table 1 and Fig. 2), exhibit a polar residue at position *VIII*, which contains mostly apolar residues for other ORs. These residues could form H-bonds with the carbonyl group of the ligands. Two other ORs that bind aldehydes, MOR175-1 and MOR258-5, do not have polar residues in position *VIII*. However MOR175-1, which also binds ketones, exhibits polar groups in positions 3 and *VII*, which can form H-bonds. MOR258-5 shows a Ser in position *I*. As MOR258-5 binds ligands that may form several H-bonds, Ser in position *I* is, at present, the only candidate in the putative binding site likely to be involved in H-bond formation with ligands.

Comparison with other models

Our proposal that position 3 forms H-bonds to ligands in most of the ORs that bind carboxylic acids and in MOR175-1 is consistent with previous models [31]. Also our hypothesis that Thr in position *VII* of MOR204-32

forms a H-bond with its ligands (heptanol and hexanol) is consistent with Floriano et al.'s observations [31].

Man et al. proposed a set of 22 amino-acid positions that are important for the ligand binding site [35]. Sixteen of them are located in helices TM3, TM4, TM5 and TM6 [35]. Ten of these positions, located mostly in helices TM3 and TM4 (*I*, 2, 3, 4, 5, *I*, 7, 8, *III*, *VIII*) are also proposed to be involved in ligand binding in this

work, where the main discrepancies pertain to helices TM5 and TM6.

TM4 is predicted to be part of the stable core of the Class A of GPCRs [45, 46]. This helix has a position conserved in most proteins of Class A and contains Trp (Trp or Tyr in many ORs), which H-bonds to a conserved Asn on TM2. In addition, we have shown that in all Class #1 ORs there is a position in TM4 that is always occupied by a basic residue when that position *IV* in TM5 presents an

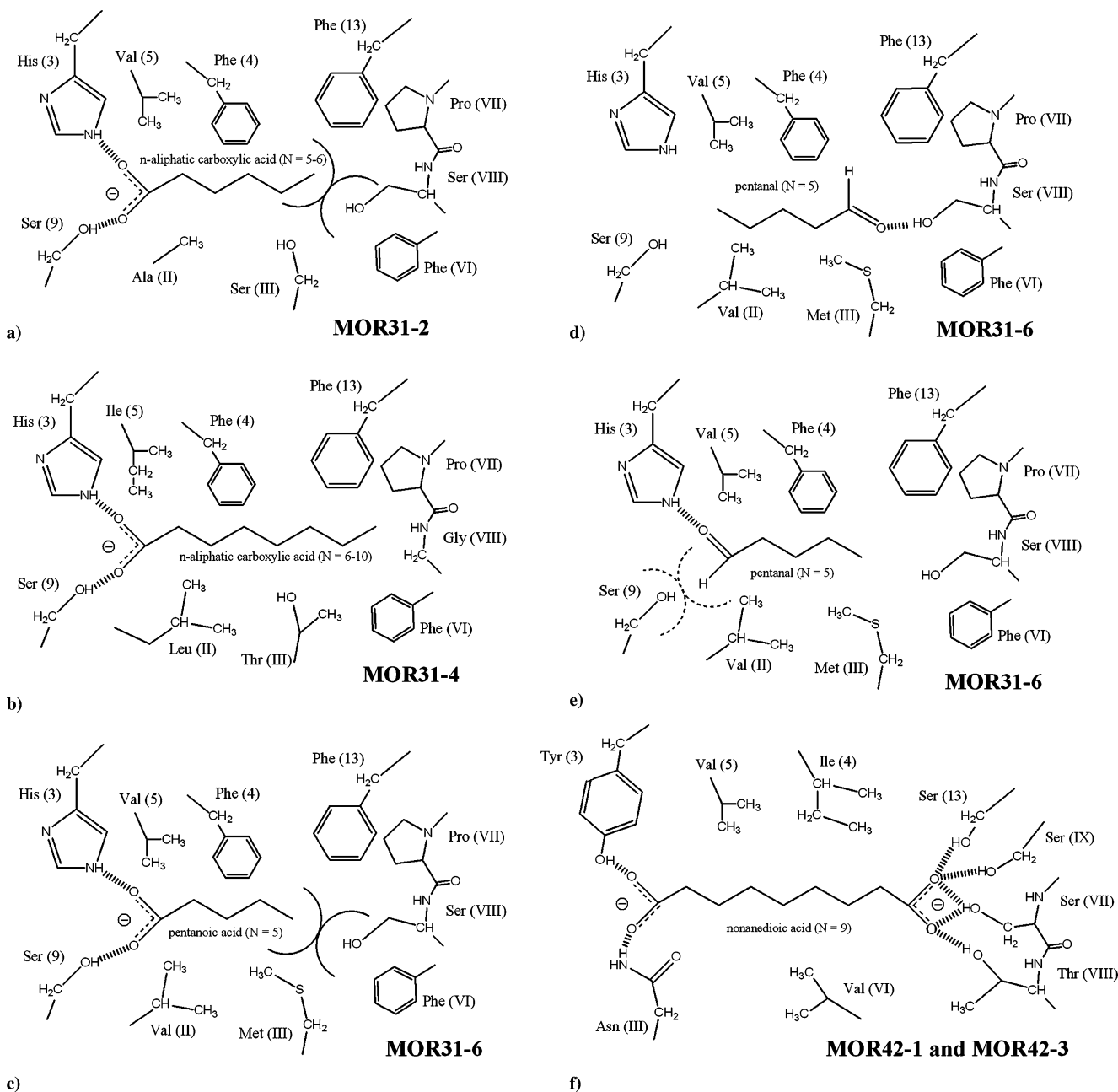


Fig. 3 Schematic representations of the interactions between several ORs, for which mutagenesis experiments have been predicted, and their ligands. **a–c** Examples of binding aliphatic carboxylic acids in MOR31-2, MOR31-4 and MOR31-6, respectively. *Half-rounds* in (**a**) and (**c**) indicate unfavorable interactions between aliphatic tails of ligands and polar residues in position *VIII*. **d** and **e** Right and

wrong orientation of the pentanal bound to the binding site of MOR31-6, respectively. **f** Binding of dicarboxylic acids in MOR42-1 and MOR42-3. A redundant number of possible hydrogen bonds is shown due to several candidate residues suitable for forming H-bond with the second carboxylate group of the ligand

Asp residue. We have therefore assumed that this information is sufficient to include TM4 helix in our model for all Class #1 ORs.

Finally, the *ab initio* model of MOR103-15 has pointed out that Lys164, located in TM4 (position 8), is the crucial residue for binding aldehydes [32]. This model is supported by free-energy calculations, which proved to be in agreement with experiment. However, our modeling suggests that this residue is disposed in the neighbor of the binding pocket, probably forming a salt bridge with Asp285, whereas Ser in TM6 (position VIII) is the best candidate for binding the carbonyl group of aldehydes: firstly, most ORs that bind aldehydes have a polar or a Cys residue in position VIII in TM6; secondly, it has been shown that in MOR174-9, the most crucial residue for binding the carbonyl group is Ser in TM3 (position 3), [33] despite this OR binding an aldehyde—heptanal—and also having the basic residue His in TM4. MOR171-2 has a Glu residue in position 8 and binds ligands acetophenone and benzaldehyde, which cannot be donors of H-bonds. Finally, MOR276-1 has the polar residue Gln in this position, but it binds an absolutely apolar ligand—limonene. Consequently, this issue could be resolved by including molecular biology data.

Concluding remarks

Our calculations suggest that: (1) residues in positions I–IV, IX and V–VIII might form electrostatic and van der Waals interactions for most ligands of Class #1 ORs; (2) in Class #2 OR, positions I, II and IX are occupied mostly by apolar groups, positions IV and VII by polar groups, and the nature of residues at positions III, V, VI and VIII is similar to that of Class #1 ORs; (3) other 14 positions (indicated as 1–14 for clarity purposes), in addition to I–IX, may play an important role for ligand binding.

It is clear that the possibility of performing further mutagenesis experiments could dramatically improve the general comprehension of the odorant recognition process and eventually provide a straightforward validation to our models. We have here emphasized the achievement of Saito et al., [8] who have been able to express a set of 11 ORs by using identified accessory proteins. Therefore, it should now be experimentally possible to test the effect of site-directed mutagenesis experiments on ligand specificity for this set of ORs.

As for those ORs that bind carboxylic acids (MOR23-1, MOR31-2, MOR31-4, MOR31-6, MOR32-5, MOR32-11 and MOR203-1), [8] the critical experiment is the substitution of the polar groups, which constitute the binding site, with apolar residues (see Table 1). Apart from MOR203-1,

which shows many polar residues facing the binding pocket and possibly participating to the carboxylic acid binding, in order to observe a change in ligand binding affinity, His in position 3 seems the best candidate for mutation (Fig. 3). Since all ORs that bind carboxylic acids show two very close polar residues in positions 3 and 9 (MOR31-2, MOR31-4 and MOR31-6), 3 and 4 (MOR23-1), 3 and II (MOR32-5, MOR32-11), 5 and III or VII and 13 (MOR203-1) (Fig. 2), a double mutation could be necessary to observe a complete loss of affinity for carboxylic acids.

Some mutations could be suggested to modify the affinity for odorants belonging to the same chemical class. For example, MOR31-4 binds *n*-aliphatic carboxylic acids with a chain length ranging from six to ten carbon atoms, contrary to MOR31-6, which can bind only pentanoic acid (and pentanal) (Fig. 3 and Table 1) [8]. We believe that the main reason for their different behavior is given by the nature of the amino-acid residues in position VIII (a Ser for MOR31-6 and a Gly for MOR31-4). Therefore, in principle MOR31-4 mutation of Gly257 to Ser or bulky residues should reduce OR affinity for long-chain carboxylic acids. However, the very low accuracy of our models does not allow us to establish this point firmly. On the contrary, the mutation of Ser256 in MOR31-6 into Gly should allow the OR to bind longer carboxylic acids. In addition, the mutation of Ser256 into Gly or apolar residues should reduce the affinity of MOR31-6 for pentanal, validating the hypothesis that position VIII is involved in the binding of aldehydes.

In the case of MOR42-1 and MOR42-3, which bind dicarboxylic acids, we proposed the presence in the binding pocket of two polar regions, constituted by several residues (Fig. 3 and Table 1). We can advance that mutations of the polar residues present in one of these two regions with apolar residues should alter ORs affinity for dicarboxylic acids, whereas the possibility of binding at least one carboxylate group should be conserved.

Acknowledgements We would like to thank Kazushige Touhara and Sayako Katada (University of Tokyo, Japan) for supplying the sequence alignment of bovine rhodopsin and MOR174-9 and the 3D-model structure of the latter, Dr. Alejandro Giorgetti (Università di Roma “La Sapienza”, Italy) for useful discussions and Manuela Schipizza Lough (SISSA/ISAS) for reading the text. This work was supported by INFM-Democritos.

References

1. Menini A, Lagostena L, Boccaccio A (2004) *News Physiol Sci* 19:101–104
2. Buck L, Axel R (1991) *Cell* 65:175–187
3. Mombaerts P (2004) *Nat Rev Neurosci* 5:263–278
4. Godfrey PA, Malnic B, Buck LB (2004) *Proc Natl Acad Sci USA* 101:2156–2161
5. Malnic B, Hirono J, Sato T, Buck LB (1999) *Cell* 96:713–723

6. Freitag J, Beck A, Ludwig G, von Buchholtz L, Breer H (1999) *Gene* 226:165–174
7. Zhang X, Firestein S (2002) *Nat Neurosci* 5:124–133
8. Saito H, Kubota M, Roberts RW, Chi Q, Matsunami H (2004) *Cell* 119:679–691
9. Araneda RC, Kini AD, Firestein S (2000) *Nat Neurosci* 3:1248–1255
10. Bozza T, Feinstein P, Zheng C, Mombaerts P (2002) *J Neurosci* 22:3033–3043
11. Gaillard I, Rouquier S, Pin JP, Mollard P, Richard S, Barnabe C, Demaille J, Giorgi D (2002) *Eur J Neurosci* 15:409–418
12. Kajiyama K, Inaki K, Tanaka M, Haga T, Kataoka H, Touhara K (2001) *J Neurosci* 21:6018–6025
13. Katada S, Nakagawa T, Kataoka H, Touhara K (2003) *Biochem Biophys Res Commun* 305:964–969
14. Katada S, Hirokawa T, Oka Y, Suwa M, Touhara K (2005) *J Neurosci* 25:1806–1815
15. Krautwurst D, Yau KW, Reed RR (1998) *Cell* 95:917–926
16. Levasseur G, Persuy MA, Grebert D, Remy JJ, Salesse R, Pajot-Augy E (2003) *Eur J Biochem* 270:2905–2912
17. Oka Y, Omura M, Kataoka H, Touhara K (2004) *EMBO J* 23:120–126
18. Touhara K, Sengoku S, Inaki K, Tsuboi A, Hirono J, Sato T, Sakano H, Haga T (1999) *Proc Natl Acad Sci USA* 96:4040–4045
19. Zhao H, Ivic L, Otaki JM, Hashimoto M, Mikoshiba K, Firestein S (1998) *Science* 279:237–242
20. Kolakowski LF Jr (1994) *Receptors Channels* 2:1–7
21. Floriano WB, Vaidehi N, Goddard WA III, Singer MS, Shepherd GM (2000) *Proc Natl Acad Sci USA* 97:10712–10716
22. Pebay-Peyroula E, Rummel G, Rosenbusch JP, Landau EM (1997) *Science* 277:1676–1681
23. Singer MS, Shepherd GM (1994) *Neuroreport* 5:1297–1300
24. Singer MS, Oliveira L, Vriend G, Shepherd GM (1995) *Receptors Channels* 3:89–95
25. Afshar M, Hubbard RE, Demaille J (1998) *Biochimie* 80:129–135
26. Schertler GF, Villa C, Henderson R (1993) *Nature* 362:770–772
27. Schertler GF (1998) *Eye* 12 (Pt 3b):504–510
28. Singer MS (2000) *Chem Senses* 25:155–165
29. Vaidehi N, Floriano WB, Trabanino R, Hall SE, Freddolino P, Choi EJ, Zamanakos G, Goddard WA III (2002) *Proc Natl Acad Sci USA* 99:12622–12627
30. Lai PC, Singer MS, Crasto CJ (2005) *Chem Senses* 30:781–792
31. Floriano WB, Vaidehi N, Goddard WA III (2004) *Chem Senses* 29:269–290
32. Hall SE, Floriano WB, Vaidehi N, Goddard WA III (2004) *Chem Senses* 29:595–616
33. Hummel P, Vaidehi N, Floriano WB, Hall SE, Goddard WA III (2005) *Protein Sci* 14:703–710
34. Trabanino RJ, Hall SE, Vaidehi N, Floriano WB, Kam VW, Goddard WA III (2004) *Biophys J* 86:1904–1921
35. Man O, Gilad Y, Lancet D (2004) *Protein Sci* 13:240–254
36. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) *Science* 289:739–745
37. Khalid S, Bond PJ, Deol SS, Sansom MS (2006) *Proteins* 63:6
38. Oakhill JS, Sutton BJ, Gorringer AR, Evans RW (2005) *Protein Eng Des Sel* 18:221–228
39. Zhang Y, Sham YY, Rajamani R, Gao J, Portoghese PS (2005) *Chembiochem* 6:853–859
40. Eilers M, Hornak V, Smith SO, Konopka JB (2005) *Biochemistry* 44:8959–8975
41. Thompson JD, Higgins DG, Gibson TJ (1994) *Nucleic Acids Res* 22:4673–4680
42. Sali A, Blundell TL (1993) *J Mol Biol* 234:779–815
43. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V (2004) *J Mol Biol* 342:571–583
44. Cornell WD, Cieplak P, Bayly CI, Gould IR, Caldwell JW, Kollman PA (1995) *J Am Chem Soc* 117:5179–5197
45. Lin JC, Parrish W, Eilers M, Smith SO, Konopka JB (2003) *Biochemistry* 42:293–301
46. Liu W, Eilers M, Patel AB, Smith SO (2004) *J Mol Biol* 337:713–729
47. Humphrey W, Dalke A, Schulten K (1996) *J Mol Graph* 14:33–38