Article

Engineered odorant receptors illuminate the basis of odour discrimination

https://doi.org/10.1038/s41586-024-08126-0

Received: 15 November 2023

Accepted: 26 September 2024

Published online: 30 October 2024

Check for updates

Claire A. de March[1,2,10]✉, Ning Ma[3,10], Christian B. Billesbølle[4,10], Jeevan Tewari[1], Claudia Llinas del Torrent[4,5], Wijnand J. C. van der Velden[3], Ichie Ojiro[1,6], Ikumi Takayama[1,7], Bryan Faust[4], Linus Li[4], Nagarajan Vaidehi[3]✉, Aashish Manglik[4,8]✉ & Hiroaki Matsunami[1,9]✉

How the olfactory system detects and distinguishes odorants with diverse physicochemical properties and molecular configurations remains poorly understood. Vertebrate animals perceive odours through G protein-coupled odorant receptors (ORs)[1]. In humans, around 400 ORs enable the sense of smell. The OR family comprises two main classes: class I ORs are tuned to carboxylic acids whereas class II ORs, which represent most of the human repertoire, respond to a wide variety of odorants[2]. A fundamental challenge in understanding olfaction is the inability to visualize odorant binding to ORs. Here we uncover molecular properties of odorant–OR interactions by using engineered ORs crafted using a consensus protein design strategy[3]. Because such consensus ORs (consORs) are derived from the 17 major subfamilies of human ORs, they provide a template for modelling individual native ORs with high sequence and structural homology. The biochemical tractability of consORs enabled the determination of four cryogenic electron microscopy structures of distinct consORs with specific ligand recognition properties. The structure of a class I consOR, consOR51, showed high structural similarity to the native human receptor OR51E2 and generated a homology model of a related member of the human OR51 family with high predictive power. Structures of three class II consORs revealed distinct modes of odorant-binding and activation mechanisms between class I and class II ORs. Thus, the structures of consORs lay the groundwork for understanding molecular recognition of odorants by the OR superfamily.

Vertebrate animals perceive odours primarily through olfactory G protein-coupled receptors (GPCRs) found within sensory neurons of the olfactory epithelium. In humans, olfactory GPCRs account for over half of the class A GPCR family[2,4,5] (Extended Data Fig. 1a). Two types of GPCRs are involved in sensing odorants: a large family of ORs commonly subdivided further into class I and class II types and a separate family of trace amine-associated receptors (TAARs)[1,6]. Class II ORs are the most prevalent, accounting for 84% of all olfactory GPCRs and with 335 identified members in humans. They are followed by class I ORs (56 members) and TAARs (6 members). ORs are further divided into 17 subfamilies (class II: 1–14; class I: 51, 52 and 56) based on their amino acid sequence similarities[7].

The olfactory system needs to detect and discriminate odorants with diverse physicochemical properties and molecular structures. This challenging task is accomplished through the combinatorial activation of olfactory GPCRs, wherein multiple odorants can activate a single receptor or a single odorant can activate multiple receptors[8,9]. Each type of olfactory GPCR is responsible for detecting a particular segment of odour chemical space. Although TAARs are specialized to sense amines and class I ORs are tuned to carboxylic acids, class II ORs respond to a larger array of volatile odorants[10,11]. TAARs and class I ORs are more abundant in fish, probably because these receptors recognize water-soluble odorants. Class II ORs have undergone substantial expansion in terrestrial vertebrates, probably because they recognize a more diverse set of volatile, poorly water-soluble odorants[12,13]. The anatomical distribution of ORs in amphibian species further supports this mapping of chemical diversity to OR classes. In the model amphibian *Xenopus laevis*, class I ORs are expressed in olfactory epithelium regions dedicated to the detection of water-soluble molecules, whereas class II ORs are found in areas dedicated to the detection of volatile odorants[14].

Several advancements have started to provide an atomic perspective on how odorants are recognized by the olfactory system. We recently reported the structure of a human odorant receptor, OR51E2, bound to the odorant propionate[15]. Like most other class I ORs, OR51E2 responds to carboxylic acids. Additional studies have reported structures of human and murine TAARs bound to diverse amines[16–20]. Despite these

[1]Department of Molecular Genetics and Microbiology, Duke University, Durham, NC, USA. [2]Institut de Chimie des Substances Naturelles, UPR2301 CNRS, Université Paris-Saclay, Gif-sur-Yvette, France. [3]Department of Computational and Quantitative Medicine, Beckman Research Institute of the City of Hope, Duarte, CA, USA. [4]Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, USA. [5]Laboratory of Computational Medicine, Biostatistics Unit, Faculty of Medicine, Universitat Autònoma Barcelona, Barcelona, Spain. [6]Department of Food and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan. [7]Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, Koganei, Tokyo, Japan. [8]Department of Anesthesia and Perioperative Care, University of California, San Francisco, San Francisco, CA, USA. [9]Department of Neurobiology, Duke Institute for Brain Sciences, Duke University, Durham, NC, USA. [10]These authors contributed equally: Claire A. de March, Ning Ma, Christian B. Billesbølle. ✉e-mail: claire.de-march@cnrs.fr; NVaidehi@coh.org; aashish.manglik@ucsf.edu; hiroaki.matsunami@duke.edu

foundational insights into odorant recognition, how class II ORs interact with diverse odorants remains elusive for two reasons: (1) class II ORs share only 18–34% amino acid identity with OR51E2 and (2) class II ORs recognize a distinct set of odour chemical space compared with class I ORs and TAARs[10,11].

To understand how the sequence diversity of ORs enables the recognition of diverse odorants, we used a combination of odorant–receptor engineering and cryogenic electron microscopy (cryo-EM) to unravel fundamental features of odorant recognition in class I and class II ORs. Together with molecular dynamics (MD) simulations and mutagenesis studies, we also reveal key differences in how each of these families recognizes odorants and important similarities in how odorants activate their receptors. Furthermore, because our engineering strategy facilitates the structure determination of otherwise technically recalcitrant ORs, it provides a way to model the thousands of ORs encoded across vertebrate genomes.

consORs are highly expressed

A fundamental challenge to the study of vertebrate ORs is the low expression levels of native receptors in heterologous cell systems[21]. Our recent structure determination of human OR51E2 relied on the identification of an OR that is atypically highly expressed in model cell lines, probably because it is ectopically expressed and strongly conserved during evolution[15]. Most other vertebrate ORs have remained recalcitrant to overexpression in heterologous cell lines, even with co-expression of dedicated OR chaperones[22–24]. Owing to these fundamental challenges in the biochemical study of OR function, we applied a previously established 'consensus' strategy for engineering thermostable proteins[25–27]. Although initially described for immunoglobulins[28] and enzymes[29], we previously demonstrated that consORs can be designed using individual members of a subfamily of human ORs[3]. Such consORs are expressed in heterologous cells at levels that approach other non-olfactory class A GPCRs and can therefore produce structures of ORs[30]. Notably, consORs are a strong starting point for modelling individual native ORs as they have high sequence identity to each individual member of an OR subfamily. Values range from 58 to 66% average sequence identity depending on the subfamily (Supplementary Table 1). Moreover, consORs are often activated by similar odorants as their corresponding native ORs[3].

We initially applied the consensus approach to study the human OR51 subfamily, which belongs to class I ORs that recognize carboxylic acid odorants. After aligning 23 members of the OR51 subfamily, we designed a consensus construct (consOR51) that retained the most common amino acid at each aligned position (Fig. 1a). A phylogenetic analysis of consOR51 that compared it with the native sequences of OR51 subfamily members showed that the consensus construct lies at the root of the extant sequences, which range from 45 to 74% amino acid identity when compared with consOR51 (Fig. 1b and Supplementary Table 1). The majority of individual OR51 subfamily members failed to express at measurable levels in HEK293T cells, with the exception of OR51E2 and, to a lesser extent, OR51E1. By contrast, consOR51 was expressed at levels higher than OR51E2 (Fig. 1c and Extended Data Fig. 2). In a GloSensor cAMP production assay, consOR51 showed significant increases in the GloSensor signal at baseline, which suggests that the consensus construct has high basal activity in the absence of an odorant (Fig. 2d, all raw signalling data are available in Supplementary Table 2).

Encouraged by the surface expression levels of consOR51, we determined a cryo-EM structure of consOR51. Following the successful strategy used for structure determination of OR51E2, we designed a construct that fused consOR51 with a carboxy-terminal mini-$G\alpha_s$ protein[15,31]. Because consOR51 is constitutively active, we purified the consOR51–mini-$G\alpha_s$ fusion protein in the absence of an odorant agonist. Consistent with the increased cell surface expression of consOR51 compared with OR51E2, we observed greater protein purification yields for consOR51–mini-$G\alpha_s$ compared with OR51E2–mini-$G\alpha_s$ (data not shown). We further added $G\beta_1\gamma_2$ and the stabilizing nanobody Nb35 to produce a complex amenable for single-particle cryo-EM studies, which produced a map of consOR51 bound to the G_s heterotrimer with 3.2 Å resolution (Fig. 1d, Extended Data Fig. 3 and Extended Data Table 1). We did not observe an odorant bound to consOR51, probably because consOR51 is stable in the active state even in the absence of an odorant (Fig. 1e). Application of the consensus strategy therefore enables strong expression of model ORs, which makes them amenable to structure determination.

The consOR51 structure templates the OR51 family

We first compared the structures of consOR51 and human OR51E2 to understand how well consOR constructs recapitulate the structure of native ORs. consOR51 shares 58% sequence identity (73% similarity) with OR51E2. The overall structure of consOR51 and OR51E2 are highly similar, with an overall root mean square deviation (RMSD) of 1.3 Å (Fig. 2a). Although the overall architecture of the extracellular loops is highly similar between consOR51 and OR51E2, the intracellular ends of transmembrane helix 5 (TM5) and TM6 deviate slightly between consOR51 and OR51E2. These differences could be due to the high basal activity of consOR51.

A potential utility of consORs is that they may enable accurate modelling of the odorant-binding pocket of native ORs. We therefore compared how well consOR51 recapitulates the binding pocket of OR51E2 (Fig. 2b,c). Although our structure of consOR51 was obtained without an odorant, comparison of the binding pockets of consOR51 and OR51E2 revealed high similarity in the identity of many amino acids in this region and the conformation of side chains that engage odorants. Perhaps most notable is a conserved arginine residue in class I ORs ($R264^{6\times59}$ in consOR51 and $R262^{6\times59}$ in OR51E2; superscripts represent the modified Ballesteros–Weinstein numbering system for GPCRs[32–34]). We have previously demonstrated that engaging the carboxylic acid of propionate by $R262^{6\times59}$ in OR51E2 is crucial for receptor activation[15]. In consOR51, we observed that $R264^{6\times59}$ is poised to make a similar contact with a carboxylic acid in the odorant-binding pocket (Fig. 2b,c). Other residues that engage the propionate carboxylic acid moiety in OR51E2 are similarly poised to interact with a carboxylic acid in consOR51. For OR51E2, we previously demonstrated that hydrophobic interactions between the aliphatic tail of fatty acids and the odorant-binding pocket confer fatty-acid-mediated activity and selectivity. As expected, residues in this region diverge between OR51E2 and consOR51. A notable difference occurs at position 3×37, which is a bulky aromatic in consOR51 ($F110^{3\times37}$) compared with a small aliphatic side chain in OR51E2 ($A108^{3\times37}$). Previous studies have shown that bulky amino acids at this position increase the basal activity of ORs[35]. Mutation of consOR51 at this position to glycine (consOR51(F110G)) produced significantly reduced basal activity and a gain of odorant-dependent responses (Fig. 2d). The increased space at position 4×57 (F155 in OR51E2, I157 in consOR51) accommodates longer chain fatty acids[15] and, as expected, consOR51(F110G) responded to longer chain length fatty acids (EC_{50} ranking: C10 < C9 < C6 < C5 < C8 < C7 < C4 < C3 < C2; C*n*, where *n* is the number of fatty acid carbons; Supplementary Table 2).

We next sought to understand whether the consOR51 structure may enable accurate homology modelling of a different OR51 family member. Here we focused on OR51E1, which shares 62% sequence identity with consOR51 (76% sequence similarity) (Fig. 2e and Supplementary Table 1). Although OR51E2 is selective for the short-chain fatty acids acetate and propionate, OR51E1 responds to longer chain fatty acids[9,36]. Indeed, in a GloSensor cAMP accumulation assay, OR51E1 responded to a range of fatty acids, with a preference for pentanoate ($pEC_{50} = -4.46 \pm 0.05$; Fig. 2f). We generated a homology model of

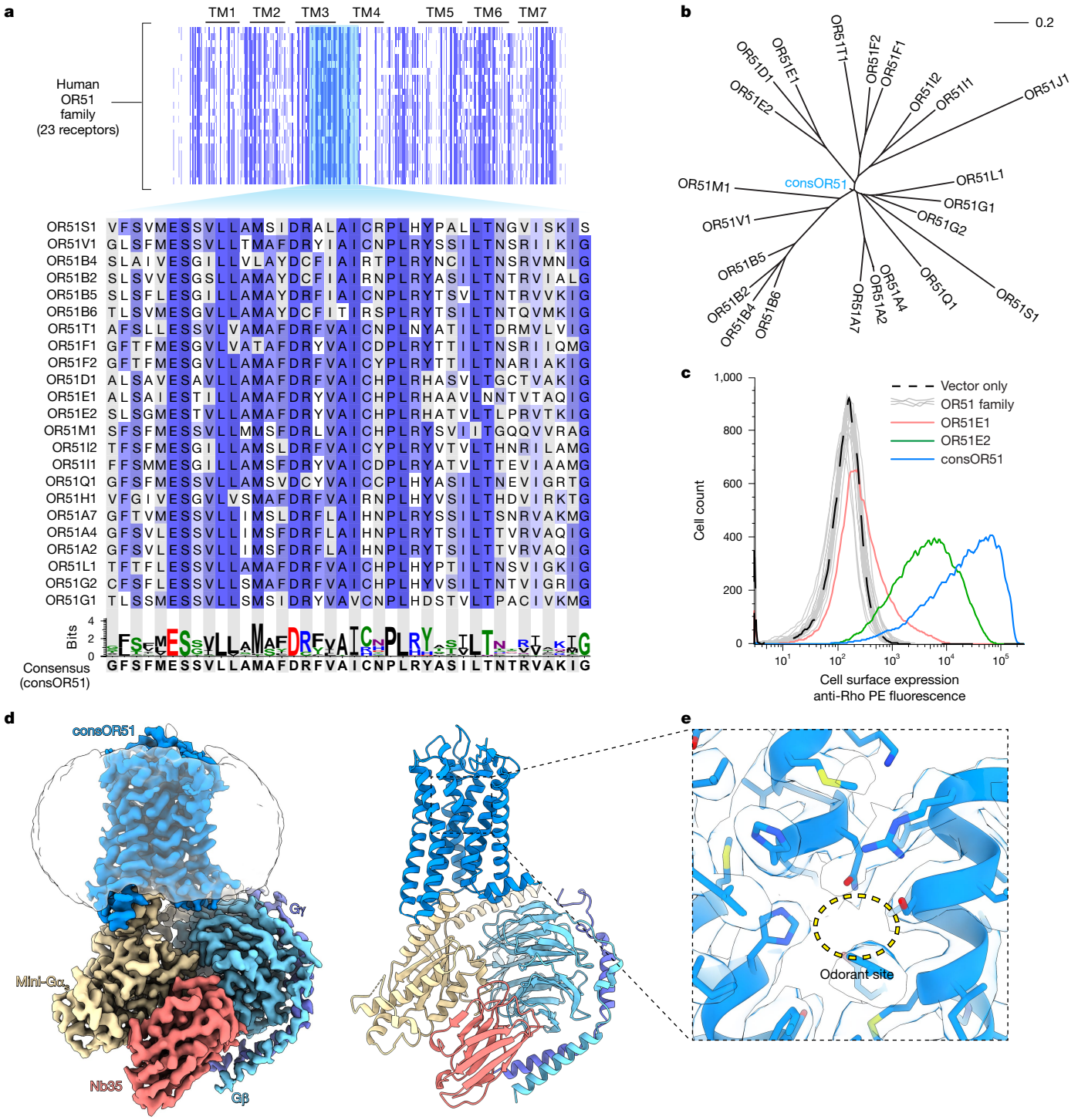


Fig. 1 | consOR strategy. a, consOR design strategy. All 23 human OR51 subfamily sequences were aligned, and the most conserved amino acid was selected at each position to create a consensus sequence. The conserved region in TM3 of the OR51 subfamily is highlighted. **b**, Phylogenetic tree of the OR51 subfamily including consOR51, which occupies the root of the subfamily tree. **c**, Cell surface expression of HEK293 cells transiently transfected with vector control, individual OR51 family members or consOR51. Most OR51 family members were poorly expressed at the cell surface, except for OR51E2. consOR51 showed a substantial increase in cell surface expression. **d**, Cryo-EM density map of consOR51 in complex with a G_s heterotrimer and the stabilizing nanobody Nb35. **e**, Zoom-in view of the putative odorant-binding site in consOR51 shows the lack of an identifiable density for an odorant.

OR51E1 using the structure of consOR51 as template and docked pentanoate into this model (Fig. 2f). Similar to the binding pose of propionate in OR51E2, the carboxylic acid of pentanoate engages an ionic and hydrogen bonding network anchored by R264$^{6\times59}$. A distinct set of residues in the divergent part of the cavity enables the longer aliphatic chain of pentanoate to bind in the OR51E1 pocket. For OR51E2, we previously demonstrated that substituting small amino acids for residues that contact the aliphatic tail of propionate changes the fatty acid preference. In particular, mutation of F155$^{4\times57}$ to alanine in OR51E2 made it responsive to longer chain fatty acids[15]. To test this model in OR51E1, we mutated residues M158$^{4\times57}$ and I205$^{5\times43}$ to alanine (M158$^{4\times57}$A and I205$^{5\times43}$A, respectively), and we predicted that introducing more

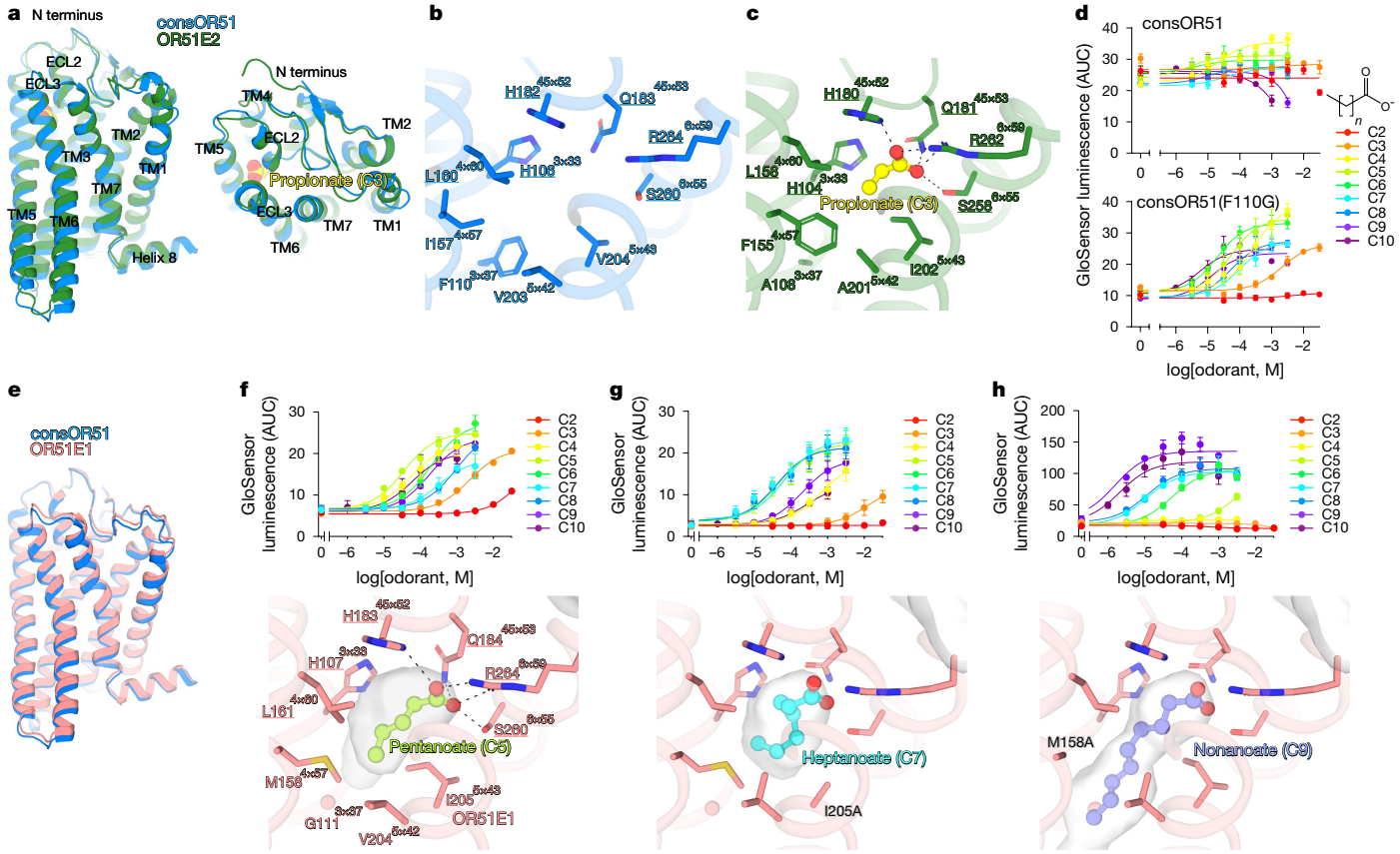


Fig. 2 | The structure of consOR51 provides insight into native OR51 family members. a, Comparison of the cryo-EM structures of consOR51 and human OR51E2 indicates the high degree of similarity in the seven transmembrane domains and the extracellular loops. **b**,**c**, Close-up view of the odorant-binding pocket in consOR51 (**b**) compared with the propionate-binding pocket of OR51E2 (**c**). Conserved side chains show similar rotamers. **d**, consOR51 is constitutively active in a GloSensor cAMP production assay. Introduction of the F110G mutation in consOR51 led to suppressed basal activity and responses to fatty acids of varying aliphatic chain length. Data points are the mean ± s.e.m. from $n = 3$ replicates. AUC, area under curve. **e**, Homology model of human OR51E1 was constructed using consOR51. **f**, Top, OR51E1 recognizes long-chain fatty acids, with a preference for pentanoate (C5). Bottom, docked structure of pentanoate in the OR51E1 homology model, with the grey surface highlighting the binding-pocket cavity. **g**,**h**, The I205$^{5\times43}$A (**g**) and M158$^{4\times57}$A (**h**) mutants of OR51E1 preferentially respond to longer chain fatty acids. Data points in **d**, **f**–**h** are the mean ± s.e.m. from $n = 3$ replicates.

space in this region would change the fatty acid preference of OR51E1. Indeed the I205$^{5\times43}$A mutant showed a preference for heptanoate and octanoate, whereas the M158$^{4\times57}$A mutant had a preference for the longer chain of nonanoate and decanoate (Fig. 2g,h). Docking of heptanoate in the I205$^{5\times43}$A mutant and nonanoate in the M158$^{4\times57}$A mutant showed that the cavity accommodates longer chain fatty acids.

With these studies, we speculate that consORs probably show high structural similarity to individual native ORs and that homology modelling of native ORs from a consOR can enable predictive models of odorant binding.

consOR1 is a representative class II OR

Attempts to express and purify class II ORs have been even more challenging than class I ORs, probably because class II ORs are generally more poorly folded and induce stronger endoplasmic reticulum stress responses[37]. Class II ORs recognize a broad range of odorants with significant structural diversity[9,38–40]. Among class II ORs, the human OR1A1 receptor has previously been characterized as a broadly tuned receptor that recognizes highly diverse odorants, including allyl phenyl acetate, dihydrojasmone, menthols and carvones[41,42]. We therefore sought to understand how the binding pocket of OR1A1 leads to its specific odorant recognition profile.

We started by using the consensus approach to generate consOR1, a construct that shares 63% sequence identity (77% sequence similarity) with native OR1A1 (Fig. 3a and Supplementary Table 1). In contrast to consOR51, consOR1 was not constitutively active and responded strongly to the odorant L-menthol (Fig. 3d and Extended Data Fig. 5). Like OR1A1, consOR1 responded to a diverse set of odorants, a result that highlights the distinct ability of consORs to recapitulate features of native ORs (Extended Data Fig. 5). Using a similar strategy as for consOR51, we determined a cryo-EM structure of consOR1 bound to L-menthol with a nominal resolution of 3.3 Å (Fig. 3b, Extended Data Fig. 4 and Extended Data Table 1).

The binding pocket of consOR1 is largely hydrophobic with a few amino acids that provide either hydrogen bond donors or acceptors. The cryo-EM density for L-menthol supported a binding pose with the hydroxyl group of the odorant engaging N109$^{3\times37}$ in the binding pocket (Fig. 3c). L-Menthol makes van der Waals contacts with many residues in the consOR1 binding pocket. Alanine mutagenesis of binding pocket residues showed that all three polar residues in the consOR1 binding pocket are important for L-menthol activity (Fig. 3e). The importance of hydrophobic contacts is more varied. Mutation of binding pocket residues in TM5, TM6 and extracellular loop 2 (ECL2) led to more major disruption in L-menthol activity compared with positions in TM3. Alanine mutation of a residue outside the binding cavity, F260$^{6\times56}$, did not lead to a change in consOR1 responses (Extended Data Fig. 5). We conclude that many residues in the consOR1 binding pocket contribute to L-menthol binding and efficacy. It is likely that many other class II ORs show similar modes of odorant recognition. That is, a combination of

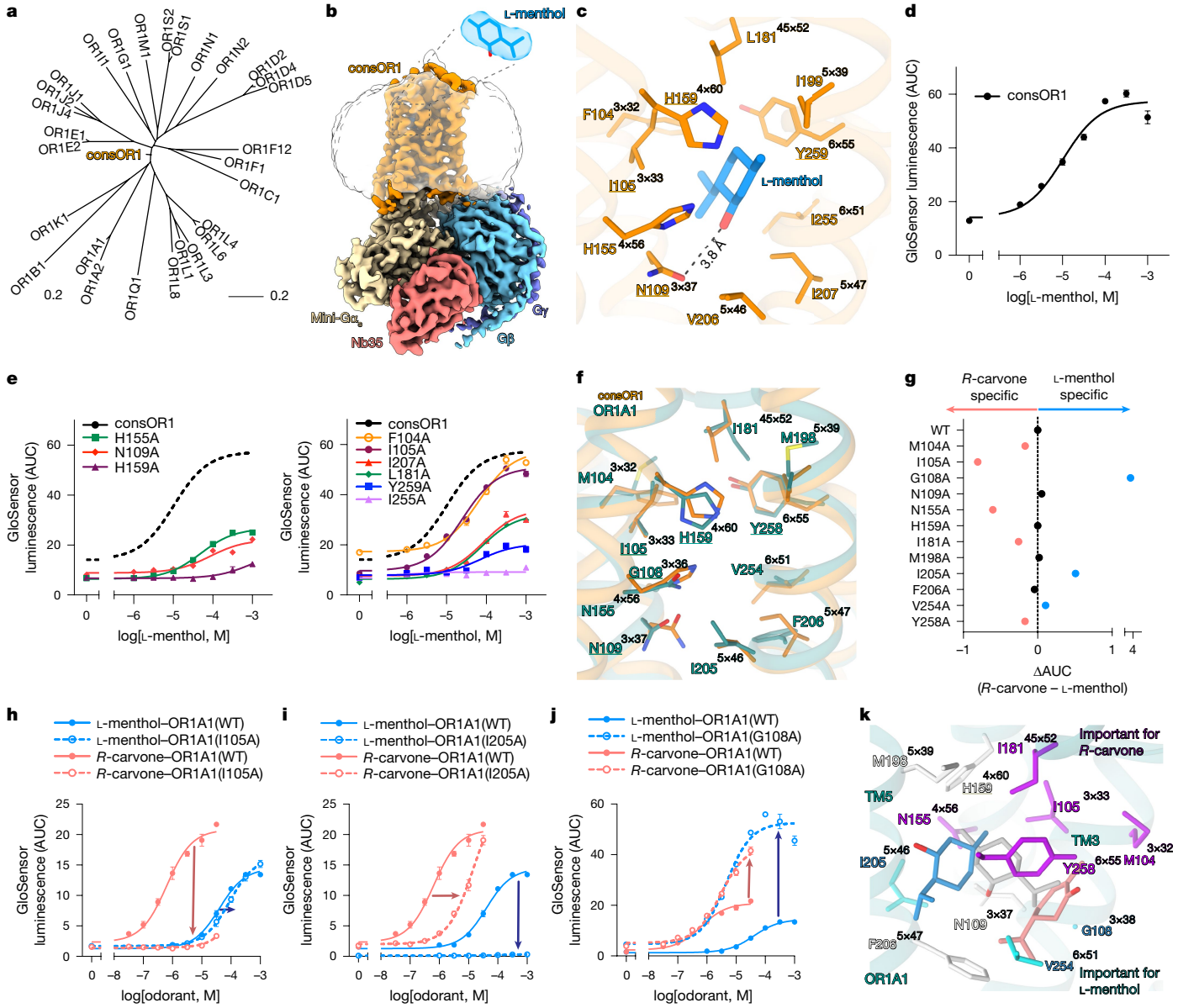


Fig. 3 | The structure of consOR1 provides insight into human OR1A1. **a**, Phylogenetic tree of the human OR1 subfamily including consOR1. **b**, Cryo-EM map of the consOR1–G_s complex. Insert shows the cryo-EM density for L-menthol. **c**, View of the consOR1 odorant-binding pocket within 5 Å with a single hydrogen bond shown as dashed lines. **d**, Dose–response for L-menthol activation of consOR1. **e**, Mutagenesis studies of consOR1 in a cAMP accumulation assay. **f**, Homology model of OR1A1 based on consOR1. **g**, OR1A1 mutants differentially affect *R*-carvone and L-menthol activity. The area under the dose–response curve was calculated for each OR1A1 mutant activated by either odorant ($n = 3$). For each odorant, AUC values were normalized to wild type (WT) OR1A1. Subtraction of normalized AUCs revealed a differential effect of mutations. **h**, The $I105^{5\times46}A$ mutation in OR1A1 had a larger effect on *R*-carvone activity than L-menthol. **i**, By contrast, OR1A1($I205^{5\times46}A$) had a greater impact on the response to L-menthol. **j**, OR1A1($G108^{3\times36}A$) increased the responses to *R*-carvone and L-menthol with a greater impact on L-menthol. **k**, Mapping the effect of mutations in **j** onto the homology modelled structure of OR1A1 shows that mutations that affect L-menthol and *R*-carvone cluster in distinct regions of the OR1A1 binding pocket. The best scoring docking results are shown for both odorants as solid sticks (L-menthol in blue and *R*-carvone in coral). Residues with a ΔAUC > 0.1 are coloured as cyan or magenta. For all cell assays, data points are the mean ± s.e.m. from $n = 3$ replicates.

many distributed hydrophobic contacts combined with a limited set of hydrogen-bonding interactions.

consOR1 templates a native class II OR

We next sought to understand odorant recognition by the native receptor OR1A1. Like consOR1, OR1A1 responds to L-menthol with micromolar potency ($pEC_{50} = -4.41 \pm 0.06$; Extended Data Fig. 5). We additionally identified several other odorants with activity at OR1A1 and focused on the molecular recognition of another terpenoid odorant, *R*-carvone (Extended Data Fig. 5). Compared to L-menthol, *R*-carvone is more potent at OR1A1 ($pEC_{50} = -6.29 \pm 0.07$; Extended Data Fig. 5). Both L-menthol and *R*-carvone are primarily hydrophobic ligands but harbour a single hydrogen bond donor or acceptor. To understand how OR1A1 recognizes these distinct terpenoids, we generated a homology model of OR1A1 based on the structure of consOR1 bound to L-menthol (Fig. 3f). This model was used for docking studies of L-menthol and *R*-carvone. In both cases, docking did not identify a single pose of the odorant within the OR1A1 binding pocket (Extended Data Fig. 6). Instead, both L-menthol and *R*-carvone dock to OR1A1 in multiple

orientations with a distributed set of van der Waals contacts. Despite the shared terpenoid scaffold of both odorants, docking revealed that L-menthol and *R*-carvone engage distinct subpockets in OR1A1 that are different from the position of L-menthol bound to consOR1 in the cryo-EM structure. In OR1A1, L-menthol engages residues in TM5 more extensively, whereas *R*-carvone engages the other side of the pocket composed primarily of residues in TM3 (Extended Data Fig. 6).

To test these docking predictions, we assessed the activity of L-menthol and *R*-carvone against alanine mutants of each binding pocket residue (Fig. 3g and Extended Data Fig. 6). We proposed that these mutations may differentially affect the activity of L-menthol and *R*-carvone owing to their distinct engagement of the OR1A1 pocket. Three mutations, N109$^{3\times37}$A, F206$^{5\times47}$A and H159$^{4\times60}$A, were deleterious for both L-menthol and *R*-carvone activity, whereas M198$^{5\times39}$A had no effect on OR1A1 responses (Extended Data Fig. 6). Other mutations more selectively affected either L-menthol or *R*-carvone activity. For example, I105$^{3\times33}$A led to loss of responses for *R*-carvone but had a negligible effect on potency and E_{max} for L-menthol (Fig. 3h). By contrast, I205$^{5\times46}$A prevented OR1A1 responses to L-menthol but had a moderate effect on *R*-carvone (Fig. 3i). The mutation G108$^{3\times36}$A increased the response to both odorants, with a greater increase in the potency and E_{max} of L-menthol (Fig. 3j). To more easily capture the combined effects of efficacy and potency, we calculated the integrated area under the dose–response curve for each mutant (Methods). Comparison of this metric revealed that OR1A1 mutations have differential effects on the activity of L-menthol and *R*-carvone (Fig. 3g). Concordant with our docking analysis, OR1A1 binding pocket mutations in TM3 more strongly affected *R*-carvone activity, whereas mutations in TM5 more strongly affected L-menthol activity (Fig. 3k).

These docking and mutagenesis studies highlight the complex mode of odorant recognition for a broadly tuned class II OR, which probably involves many different odorant binding poses. Different odorants probably engage a single odorant receptor binding pocket in distinct ways, which further adds complexity to molecular recognition in the OR system.

Dynamics in consOR1 ligand recognition

The flexibility of *R*-carvone docking to OR1A1 and the site-directed mutagenesis data suggested that odorants can bind class II ORs without a single, well-defined binding pose. Our previous studies of OR51E2 showed that propionate is not flexible in its binding site and persistently adopts a single pose that is constrained by an ionic interaction. Compared with highly water-soluble class I odorants like propionate, class II OR ligands are more hydrophobic, often with only a single hydrogen bond donor or acceptor (Extended Data Fig. 1a). We therefore sought to understand the structural dynamics of odorant binding to class II ORs.

We turned to all-atom MD simulations to examine the flexibility of L-menthol in the consOR1 binding pocket. To understand how the G protein and ligand influence consOR1 flexibility, we performed simulations under the following conditions: (1) consOR1 bound to L-menthol and mini-Gα_s; (2) consOR1 bound to L-menthol without mini-Gα_s; and (3) consOR1 alone (Fig. 4a). Each simulation was performed in five replicates and each replicate was evolved over 1 μs (Extended Data Fig. 7). As expected based on simulations of other GPCRs[43–45], removal of mini-G$_s$ and L-menthol led to an increase in structural flexibility of consOR1 (Fig. 4a). Notably, L-menthol was highly dynamic within the ligand-binding pocket of consOR1 (Fig. 4b–d). In the absence of G protein, L-menthol explored a broader range of the odorant-binding site, with a ligand RMSD of 6.1 Å, compared with the cryo-EM structure of L-menthol bound to consOR1. In simulations of consOR1 bound to mini-Gα_s, the flexibility of L-menthol was reduced, with a ligand RMSD of 4.2 Å. The flexibility of L-menthol is in contrast to the relative stability of propionate bound to OR51E2 (Fig. 4e–g). Our previous simulations of propionate bound to OR51E2 revealed an overall ligand RMSD of 2.1 Å

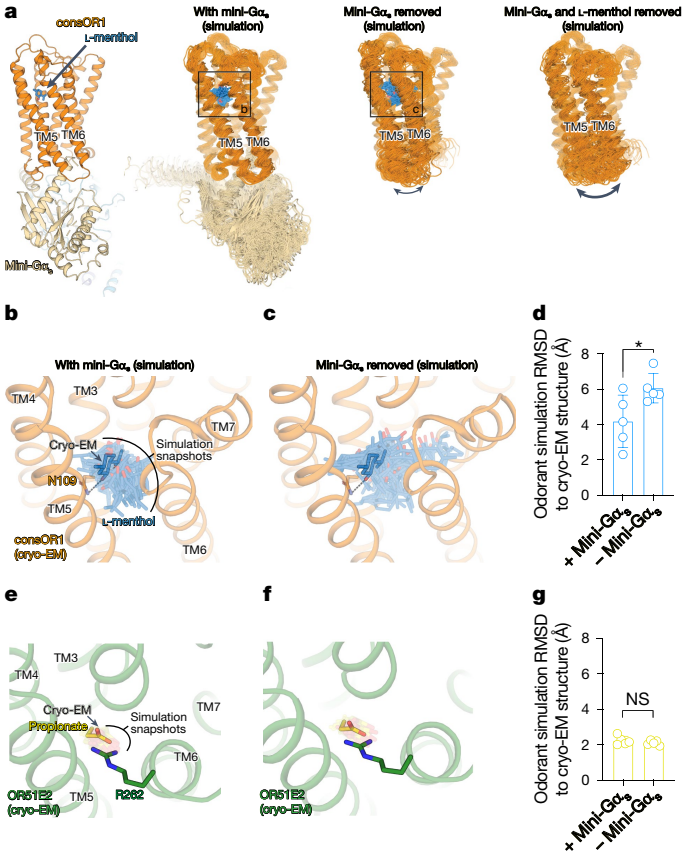


Fig. 4 | Structural flexibility in odorant binding. a, MD simulations were performed for consOR1, with mini-Gα_s and L-menthol, without mini-Gα_s but with L-menthol, and without both mini-Gα_s and L-menthol. Snapshots from 100 ns intervals are shown from the MD simulations, which were initiated by 5 random seeds. In consOR1, TM5 and TM6 are more flexible in the absence of mini-Gα_s, and even more dynamic in the absence of mini-Gα_s and L-menthol. **b**,**c**, L-Menthol is dynamic in the binding pocket of consOR1 in simulations with mini-Gα_s (**b**), and shows even greater flexibility in simulations without mini-Gα_s (**c**). **d**, The RMSD of L-menthol compared to the cryo-EM pose for each simulation replicate ($n = 5$ random velocities to initiate simulation, data are presented as the mean ± s.d., $P = 0.039$). *$P < 0.05$. **e**,**f**, In simulations of OR51E2, propionate is constrained within the ligand-binding pocket and makes a persistent interaction with R262. **g**, The RMSD of propionate compared to the cryo-EM pose for each simulation replicate ($n = 5$ random velocities to initiate simulation, data are presented as the mean ± s.d., $P = 0.341$). NS, not significant. *P* values were calculated by two-sided Welch's *t*-test.

and 2.4 Å for simulations performed with and without the mini-Gα_s, respectively. The increased flexibility of L-menthol in simulations of consOR1 without mini-Gα_s correlated with an increase in the volume of the consOR1 binding pocket. With mini-Gα_s, L-menthol explored a consOR1 pocket that encloses 250 Å^3. In the absence of mini-Gα_s, the pocket expanded to 450 Å^3 (Extended Data Fig. 7). The increased volume of the consOR1 ligand binding pocket arises from an outward movement of ECL3 and the extracellular sides of TM6 and TM7.

Taken together, these simulation and mutagenesis studies suggest that odorants bind class II ORs with significantly greater flexibility compared with class I ORs. Furthermore, our simulations show that binding of the G protein decreases odorant flexibility in a class II OR binding pocket.

A shared class II OR activation motif

We next sought to expand the consOR strategy to other class II ORs, with the aim of understanding both shared and distinct features between

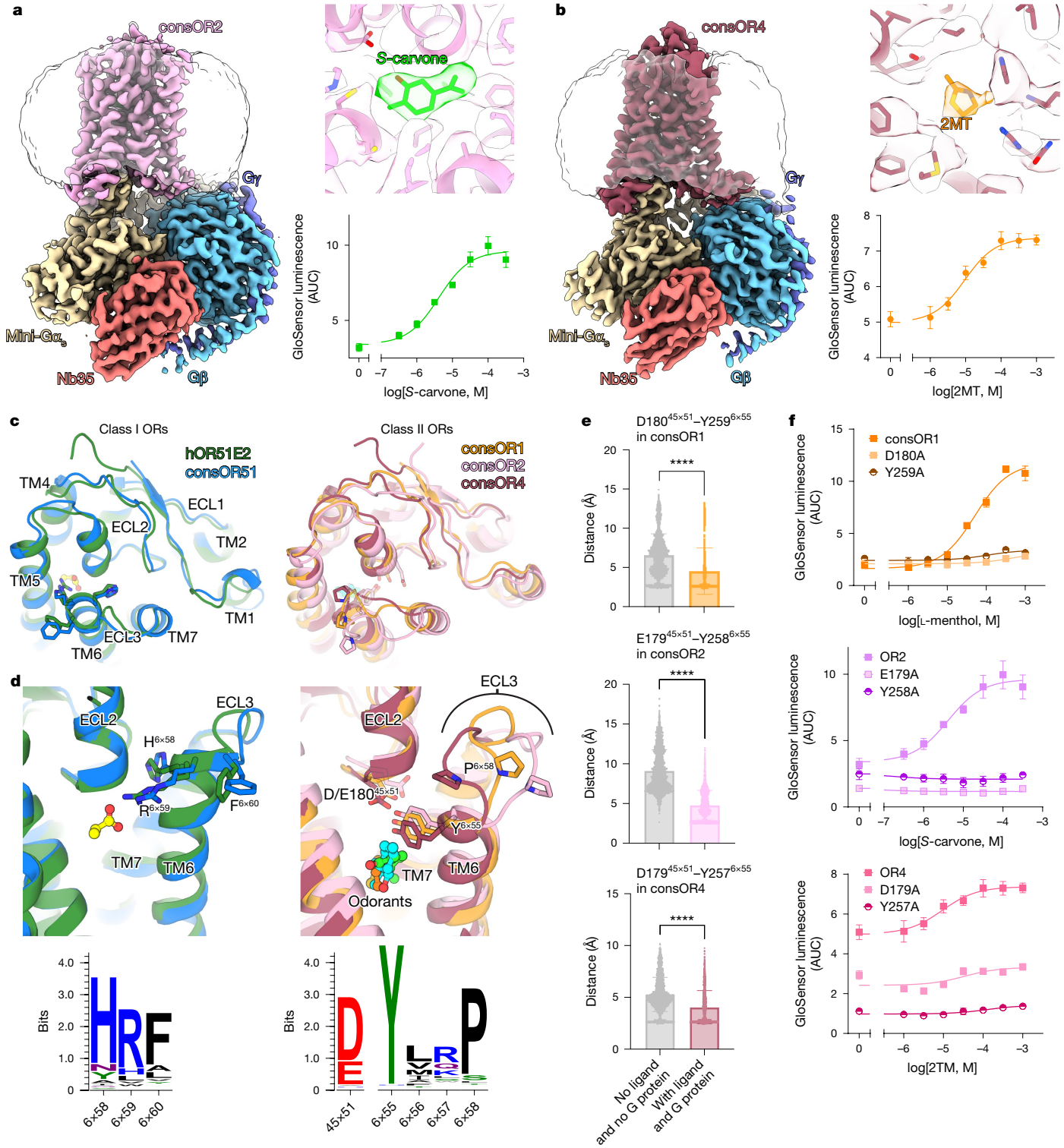


Fig. 5 | Structures of consOR2 and consOR4 reveal common features of OR function. a, Cryo-EM map of a consOR2–G_s complex bound to the activating odorant *S*-carvone. A close-up view of ligand density is shown on the right, along with dose–response in an activation assay for consOR2. **b**, Cryo-EM map of a consOR4-G_s complex bound to the activating odorant 2-MT. A close-up view of ligand density is shown on the right, along with dose–response in an activation assay for consOR4. **c**, Comparison of class I and class II OR structures in the extracellular region. consOR structures of class II ORs show variability in the ECL3 conformation. **d**, Close-up view of ligand-binding sites in class I and class II ORs. Class I ORs recognize carboxylic acids through the $R^{6\times59}$ residue in the extracellular portion of TM6. Class II ORs bind ligands through a highly conserved $Y^{6\times55}$ residue that further engages a conserved acidic residue in ECL2 ($D/E^{45\times51}$). WebLogo plots show $R^{6\times59}$ is conserved in 71% of class I ORs. $Y^{6\times55}$ and $D/E^{45\times51}$ are conserved in 96% and 95% of class II ORs, respectively. **e**, The interaction between $D/E^{45\times51}$ and $Y^{6\times55}$ is maintained in simulations of consOR1, consOR2 and consOR4 bound to their agonist and mini-$G\alpha_s$. Removal of mini-$G\alpha_s$ and agonist leads to an increased distance between these two positions ($n = 2{,}500$ snapshots from 5 independent simulations for each condition). $^{****}P < 0.0001$ calculated using repeated measures two-way analysis of variance (ANOVA). **f**, Mutation of $D^{45\times51}$ and $Y^{6\times55}$ in consOR1, consOR2 and consOR4 reduces OR responses to odorant in a cAMP production assay. For all cell assay, data points are the mean ± s.e.m. from $n = 3$ replicates.

class I and class II ORs. We therefore applied the consensus strategy to other human class II OR subfamilies: the OR2 family (68 members) and the OR4 family (51 members) (Extended Data Fig. 1 and Supplementary Table. 1). We previously identified odorants that activate consOR2 and consOR4 (ref. 3), and selected those with high potency and water solubility for structural studies: *S*-carvone for consOR2 and 2-methyl thiazoline (2MT) for consOR4 (Fig. 5a,b). We determined cryo-EM structures of consOR2 bound to *S*-carvone and consOR4 bound to 2MT at 3.2 Å and 3.5 Å, respectively (Extended Data Figs. 8 and 9 and Extended Data Table 1). For both receptors, we identified cryo-EM densities for the odorant molecules (Fig. 5a,b and Extended Data Fig. 10). The placement of *S*-carvone was relatively clear based on the cryo-EM density, whereas the position of 2MT was more ambiguous. Similar to consOR1, our simulations of consOR2 and consOR4 revealed significant flexibility in the binding pose of odorants at these receptors (Extended Data Fig. 10)

With this set of OR structures, we aimed to identify structural features that are specific to class I and class II ORs. As expected, the intracellular regions of class I and class II ORs were conserved both in sequence and structure because these regions are crucial for G protein coupling in response to odorant binding (Extended Data Fig. 11). The overall fold of the extracellular region was similar between class I and class II ORs (Fig. 5c).

Despite these similarities, our structural analysis highlighted a common motif in the extracellular region of class II ORs that was distinct from class I ORs and is probably important for receptor activation (Fig. 5d). For OR51E2, we previously demonstrated that a highly conserved arginine residue in class I ORs ($R^{6\times59}$) at the extracellular tip of TM6 engages the carboxylic acid group of fatty acids. This interaction restrains an otherwise dynamic ECL3, which is associated with receptor activation (Extended Data Fig. 7). In structures of class II consORs, position 6×59 is not conserved. Instead, we identified a highly conserved acidic residue in ECL2 ($D/E^{45\times51}$) that makes a hydrogen bond with another class-II-specific conserved tyrosine residue in class II ORs ($Y^{6\times55}$). This conserved contact is in proximity to the odorant-binding site, which suggests that it might have an important role in connecting odorant binding to receptor activation. To explore this possibility, we used MD simulations to examine this conserved class II OR contact. With the odorant and G protein bound, this contact was maintained in most simulations across all three class II consORs. In simulations without the odorant and G protein, the interaction between $D/E^{45\times51}$ and $Y^{6\times55}$ was less stable, with significantly greater distances over the simulation time frames (Fig. 5e). Indeed, for all three consORs and three native class II ORs (OR1A1, OR2W1 and OR5P3; Extended Data Fig. 7), disruption of this interaction by alanine mutation markedly reduced odorant-induced activity (Fig. 5f). We therefore conclude that the conserved interaction between $D/E^{45\times51}$ and $Y^{6\times55}$ is an important mechanism for odorant-induced activation of class II ORs.

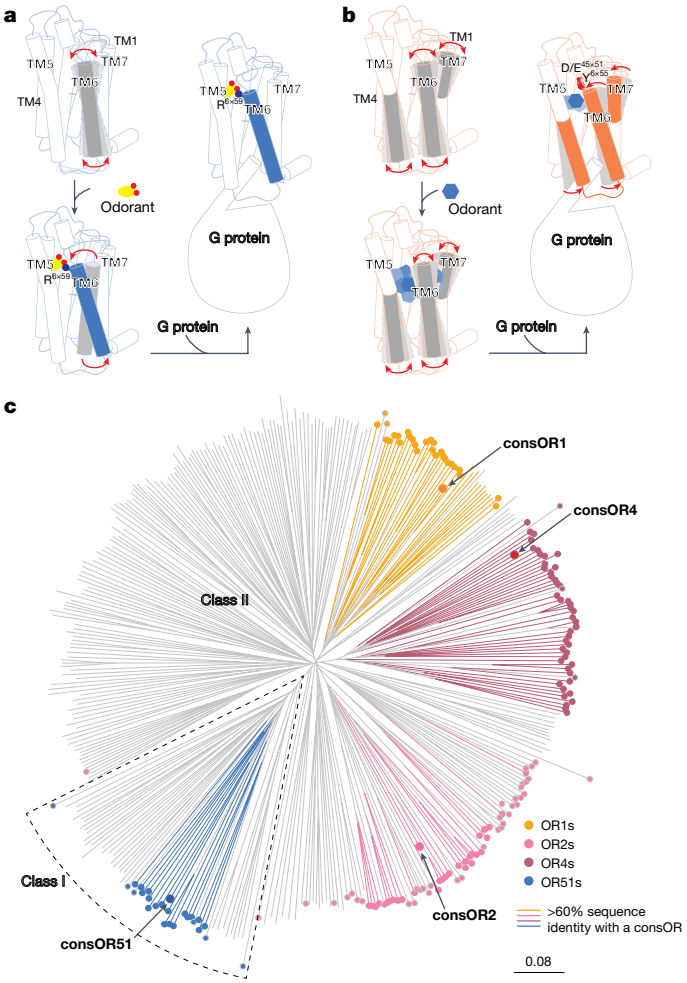


Fig. 6 | Accessing class I and class II OR mechanisms and structures. **a**, Model for class I OR activation. TM6 is conformationally flexible in the apo receptor. Binding of a carboxylic acid odorant stabilizes TM6 by engaging the class I-specific conserved $R^{6\times59}$, which leads to conformational changes that enable the binding of a G protein. **b**, Model for class II OR activation. In the apo state, TM5, TM6 and TM7 are dynamic. Binding of an agonist leads to decreased flexibility. Unlike class I ORs, the odorant–OR complex is dynamic and the odorant explores many areas within the large odorant-binding cavity. G protein-binding further stabilizes the activated OR, characterized by an interaction between the class II-specific residues $D/E^{45\times51}$ and $Y^{6\times55}$. **c**, A revised OR phylogenetic tree that includes the structurally elucidated consORs. ORs that belong to a consOR subfamily are highlighted with rounded tips. ORs with >60% sequence identity are shown in coloured lines. If used as a cut-off for accurate homology modelling, consOR structures would enable modelling of 34% of all native human ORs. The scale represents the square root of the pairwise distance between sequences.

Discussion

Our studies of several OR structures and their dynamic movements produced an emerging general model for odorant recognition. Class I ORs recognize carboxylic acids through a conserved arginine residue in TM6 ($R^{6\times59}$). The structure of constitutively active consOR51 captured without an odorant underscores that this residue occupies a conserved position in the binding pocket of activated class I ORs. Predictive homology modelling of OR51E1 based on consOR51 further supports the following model for class I OR odorant recognition: conserved binding pocket residues that engage the carboxylic acid combined with more divergent binding pocket residues that tune the response profile for fatty acids of varying aliphatic length. Together, these interactions stably position an odorant in the binding site. Although odorants bind class II ORs in a similar location as class I ORs, our studies highlighted several distinct mechanisms of odorant recognition between class I and class II ORs. First, class II ORs do not harbour a conserved interaction partner analogous to $R^{6\times59}$ in class I ORs. Second, odorants make a diffuse set of van der Waals contacts in the class II OR binding pocket, often with a single hydrogen-bonding interaction. For broadly tuned class II ORs, different odorants are likely to occupy different subpockets of the odorant-binding site, thereby leading to distinct sets of interactions important for their activity. Third, our studies of consOR1 showed that odorants bind with significant flexibility in class II ORs compared with class I ORs. This difference probably arises from the more limited set of strong ionic or hydrogen bond contacts in most volatile odorants that activate class II ORs compared with the charged water-soluble odorants that activate class I ORs. An additional factor is probably the increased flexibility of the OR binding pocket in class II ORs. Recognition of odorants by class II ORs is therefore also distinct

from TAARs, which recognize aminergic odorants through conserved ionic interactions[46].

More broadly, the majority of small-molecule-binding class A GPCRs use specific hydrogen bonding or ionic interactions to drive specificity in ligand binding. Class II ORs, by contrast, recognize odorants primarily through van der Waals contacts and with limited hydrogen-bonding interactions. Our model of odorant recognition in vertebrate ORs recapitulates recent structural biology studies that have identified a flexible binding mode for odorants at a broadly tuned ionotropic insect olfactory receptor, MhOR5 (refs. 47,48). In both cases, odorant binding is not confined to a single pose. Despite this flexibility, distinct interactions made between the odorant and OR binding pocket can still result in different odorant activity, as outlined by our studies of *R*-carvone and L-menthol acting at OR1A1. These distinct sets of interactions drive odorant discrimination. Although our studies start to explain some features of molecular recognition in class II ORs, a more complete understanding of how the large diversity of odorants is recognized by this set of ORs will require significant further structural interrogation of both broadly and narrowly tuned receptors.

Our structural analysis also sheds light on a unifying mechanism of class I and class II OR activation by chemically diverse odorants. Although the specific motifs that engage odorants are distinct between class I and class II ORs, a highly conserved interaction between TM6 and the odorant or odorant-binding pocket stabilizes an inward movement of the extracellular side of TM6. For class I ORs, this interaction is driven by the odorant engaging $R^{6\times59}$ (Fig. 6a). For class II ORs, odorants stabilize an interaction between $D/E^{45\times51}$ and $Y^{6\times55}$ (Fig. 6b). Odorant binding in both class I and class II ORs causes an inward movement of the extracellular region in TM6 (refs. 15,30). This movement is accompanied by an outward movement of the intracellular side of TM6, which creates a cavity for engaging a G protein. Odorants can be structurally flexible while bound to class II ORs. Full activation of the OR with odorant and G protein restrains some of this flexibility. Although a more accurate model will require an experimental structure of an inactive OR, our proposed model provides a shared activation mechanism for the broader OR family.

A key advancement of this study is the broad utility of a consensus engineering approach to understand OR function[3]. The majority of ORs, in both vertebrate and invertebrate species, remain intractable for biochemical and structural studies. With the consensus approach, we obtained four cryo-EM structures of consORs with high sequence identity to a subset of native human ORs. Our comparison of consOR51 to native human OR51E2 highlighted that consOR structures not only share virtually identical backbone structures to native OR family members but also key residue positions in the structures relevant for odorant recognition. Although AlphaFold has transformed protein structure prediction[49], OR regions important for odorant recognition (for example, ECL3) remain poorly predicted, probably because they diverge so widely across the OR family[15]. We therefore propose that consOR structures will enable the generation of higher quality models of many ORs, used either as templates for AlphaFold or in more classical homology modelling approaches. If we use a threshold of 60% sequence identity as a metric for high-quality templates for such modelling[50], the four consOR structures described here would facilitate the determination of high-quality models for 34% of all human ORs (Fig. 6c and Supplementary Table 3). Additional consOR structures derived from the other major OR families will further expand this number[30]. The ability to capture structures of odorants bound to consORs will probably continue to provide fundamental insights into how vertebrate ORs cope with the chemical diversity of odorous molecules.

The success of consORs provides potential insight into the evolution of the OR family. We previously proposed that the stability of consORs indicates that ancestral OR sequences were probably more stable than the majority of extant OR sequences[3]. Diversification of OR sequences is enabled by evolutionary capacitance provided by OR-specific chaperones[51–53]. The high structural similarity between a consOR and a native OR suggests that evolution drives OR diversity within a family primarily by altering contacts with odorants as opposed to substantial variation in the overall fold of the OR. This diversity re-tunes odorant specificity. Furthermore, the fact that the consensus strategy produces stable ORs for multiple OR subfamilies suggests that the common ancestor of each major human OR subfamily was probably a more stable receptor, and that evolution drives diversification for odorant recognition function at the cost of stability. We anticipate that future studies of visualizing OR structures and odorant recognition will provide deeper insight into the importance of such trade-offs.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08126-0.

1. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
2. Glusman, G., Yanai, I., Rubin, I. & Lancet, D. The complete human olfactory subgenome. *Genome Res.* **11**, 685–702 (2001).
3. Ikegami, K. et al. Structural instability and divergence from conserved residues underlie intracellular retention of mammalian odorant receptors. *Proc. Natl Acad. Sci. USA* **117**, 2957–2967 (2020).
4. Malnic, B., Godfrey, P. A. & Buck, L. B. The human olfactory receptor gene family. *Proc. Natl Acad Sci. USA* **101**, 2584–2589 (2004).
5. Bjarnadóttir, T. K. et al. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics* **88**, 263–273 (2006).
6. Liberles, S. D. & Buck, L. B. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**, 645–650 (2006).
7. Olender, T., Jones, T. E. M., Bruford, E. & Lancet, D. A unified nomenclature for vertebrate olfactory receptors. *BMC Evol. Biol.* **20**, 42 (2020).
8. Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
9. Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J. D. Odor coding by a mammalian receptor repertoire. *Sci. Signal.* **2**, ra9 (2009).
10. Cichy, A., Shah, A., Dewan, A., Kaye, S. & Bozza, T. Genetic depletion of class I odorant receptors impacts perception of carboxylic acids. *Curr. Biol.* **29**, 2687–2697.e4 (2019).
11. Dewan, A., Pacifico, R., Zhan, R., Rinberg, D. & Bozza, T. Non-redundant coding of aversive odours in the main olfactory pathway. *Nature* **497**, 486–489 (2013).
12. Niimura, Y. On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biol. Evol.* **1**, 34–44 (2009).
13. Bear, D. M., Lassance, J.-M., Hoekstra, H. E. & Datta, S. R. The evolving neural and genetic architecture of vertebrate olfaction. *Curr. Biol.* **26**, R1039–R1049 (2016).
14. Freitag, J., Krieger, J., Strotmann, J. & Breer, H. Two classes of olfactory receptors in *Xenopus laevis*. *Neuron* **15**, 1383–1392 (1995).
15. Billesbølle, C. B. et al. Structural basis of odorant recognition by a human odorant receptor. *Nature* **615**, 742–749 (2023).
16. Guo, L. et al. Structural basis of amine odorant perception by a mammal olfactory receptor. *Nature* **618**, 193–200 (2023).
17. Shang, P. et al. Structural and signaling mechanisms of TAAR1 enabled preferential agonist design. *Cell* **186**, 5347–5362.e24 (2023).
18. Xu, Z. et al. Ligand recognition and G-protein coupling of trace amine receptor TAAR1. *Nature* **624**, 672–681 (2023).
19. Liu, H. et al. Recognition of methamphetamine and other amines by trace amine receptor TAAR1. *Nature* **624**, 663–671 (2023).
20. Gusach, A. et al. Molecular recognition of an odorant by the murine trace amine-associated receptor TAAR7f. *Nat. Commun.* **15**, 7555 (2024).
21. Lu, M., Echeverri, F. & Moyer, B. D. Endoplasmic reticulum retention, degradation, and aggregation of olfactory G-protein coupled receptors. *Traffic* **4**, 416–433 (2003).
22. Saito, H., Kubota, M., Roberts, R. W., Chi, Q. & Matsunami, H. RTP family members induce functional expression of mammalian odorant receptors. *Cell* **119**, 679–691 (2004).
23. Zhuang, H. & Matsunami, H. Evaluating cell-surface expression and measuring activation of mammalian odorant receptors in heterologous cells. *Nat. Protoc.* **3**, 1402–1413 (2008).
24. Noe, F. et al. IL-6-HaloTag® enables live-cell plasma membrane staining, flow cytometry, functional expression, and de-orphaning of recombinant odorant receptors. *J. Biol. Methods* **4**, e81 (2017).
25. Sternke, M., Tripp, K. W. & Barrick, D. Consensus sequence design as a general strategy to create hyperstable, biologically active proteins. *Proc. Natl Acad. Sci. USA* **116**, 11275–11284 (2019).
26. Desjarlais, J. R. & Berg, J. M. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc. Natl Acad. Sci. USA* **90**, 2256–2260 (1993).
27. Porebski, B. T. & Buckle, A. M. Consensus protein design. *Protein Eng. Des. Sel.* **29**, 245–251 (2016).

28. Steipe, B., Schiller, B., Plückthun, A. & Steinbacher, S. Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* **240**, 188–192 (1994).
29. Lehmann, M. et al. From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. *Protein Eng.* **13**, 49–57 (2000).
30. Choi, C. et al. Understanding the molecular mechanisms of odorant binding and activation of the human OR52 family. *Nat. Commun.* **14**, 8105 (2023).
31. Nehmé, R. et al. Mini-G proteins: novel tools for studying GPCRs in their active conformation. *PLoS ONE* **12**, e0175642 (2017).
32. Ballesteros, J. A. & Weinstein, H. in *Methods in Neurosciences* Vol. 25 (ed. Sealfon, S. C.) 366–428 (Academic Press, 1995).
33. de March, C. A., Kim, S.-K., Antonczak, S., Goddard, W. A. 3rd & Golebiowski, J. G protein-coupled odorant receptors: from sequence to structure. *Protein Sci.* **24**, 1543–1548 (2015).
34. Isberg, V. et al. Generic GPCR residue numbers—aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **36**, 22–31 (2015).
35. de March, C. A. et al. Conserved residues control activation of mammalian G protein-coupled odorant receptors. *J. Am. Chem. Soc.* **137**, 8611–8616 (2015).
36. Pluznick, J. L. et al. Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. *Proc. Natl Acad. Sci. USA* **110**, 4410–4415 (2013).
37. Shayya, H. J. et al. ER stress transforms random olfactory receptor choice into axon targeting precision. *Cell* **185**, 3896–3912.e22 (2022).
38. Mainland, J. D., Li, Y. R., Zhou, T., Liu, W. L. L. & Matsunami, H. Human olfactory receptor responses to odorants. *Sci. Data* **2**, 150002 (2015).
39. Kajiya, K. et al. Molecular bases of odor discrimination: reconstitution of olfactory receptors that recognize overlapping sets of odorants. *J. Neurosci.* **21**, 6018–6025 (2001).
40. Grosmaitre, X. et al. SR1, a mouse odorant receptor with an unusually broad response profile. *J. Neurosci.* **29**, 14545–14552 (2009).
41. Schmiedeberg, K. et al. Structural determinants of odorant recognition by the human olfactory receptors OR1A1 and OR1A2. *J. Struct. Biol.* **159**, 400–412 (2007).
42. Geithe, C., Noe, F., Kreissl, J. & Krautwurst, D. The broadly tuned odorant receptor OR1A1 is highly selective for 3-methyl-2,4-nonanedione, a key food odorant in aged wines, tea, and other foods. *Chem. Senses* **42**, 181–193 (2017).
43. Ma, N., Lee, S. & Vaidehi, N. Activation microswitches in adenosine receptor A_{2A} function as rheostats in the cell membrane. *Biochemistry* **59**, 4059–4071 (2020).
44. Dror, R. O. et al. Activation mechanism of the β_2-adrenergic receptor. *Proc. Natl Acad. Sci. USA* **108**, 18684–18689 (2011).
45. Lee, S., Nivedha, A. K., Tate, C. G. & Vaidehi, N. Dynamic role of the G protein in stabilizing the active state of the adenosine A_{2A} receptor. *Structure* **27**, 703–712.e3 (2019).
46. Li, Q. et al. Non-classical amine recognition evolved in a large clade of olfactory receptors. *eLife* **4**, e10441 (2015).
47. Del Mármol, J., Yedlin, M. A. & Ruta, V. The structural basis of odorant recognition in insect olfactory receptors. *Nature* **597**, 126–131 (2021).
48. Butterwick, J. A. et al. Cryo-EM structure of the insect olfactory receptor Orco. *Nature* **560**, 447–452 (2018).
49. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
50. Bender, B. J., Marlow, B. & Meiler, J. Improving homology modeling from low-sequence identity templates in Rosetta: a case study in GPCRs. *PLoS Comput. Biol.* **16**, e1007597 (2020).
51. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
52. Wyganowski, K. T., Kaltenbach, M. & Tokuriki, N. GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. *J. Mol. Biol.* **425**, 3403–3414 (2013).
53. Agozzino, L. & Dill, K. A. Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *Proc. Natl Acad. Sci. USA* **115**, 9092–9097 (2018).

Methods

Expression and purification of consOR–mini-G_s protein complexes

Expression and purification of consOR constructs was done in a similar manner to OR51E2–mini-G_s (ref. 15). In brief, consensus OR sequences[3] were cloned into pCDNA-Zeo-TetO with an amino-terminal influenza haemagglutinin signal sequence and a Flag (DYKDDDDK) epitope. The construct included the mini-G_{s399} protein[31] fused to the C terminus with a human rhinovirus 3C protease cleavage site. The resulting constructs were transfected into inducible Expi293F-TetR cells (ThermoFisher; untested for mycoplasma contamination) using an ExpiFectamine 293 Transfection kit per the manufacturer's instructions. After 16 h, protein expression was induced with 1 μg ml^{-1} doxycycline hyclate and the culture was incubated for 36 h in a shaking incubator maintained at 37 °C and a 5% CO_2 atmosphere before cell collection by centrifugation. The resulting pellet was stored at −80 °C until purification.

OR purification was performed as previously described[15]. Cell pellets were thawed with hypotonic lysis buffer in 20 mM HEPES, pH 7.5, 1 mM EDTA, 100 μM tris(2-carboxyethyl)phosphine (TCEP; Fisher Scientific) and 1 EDTA-free protease inhibitor tablet (Pierce; Thermo Scientific) for 10 min at 4 °C. The lysis buffer was supplemented with odorants to stabilize the consOR constructs: 3 mM L-menthol, 1 mM *S*-carvone and 30 mM 2MT were used for consOR1–mini-G_{s399}, consOR2–mini-G_{s399} and consOR4–mini-G_{s399} purification, respectively. Lysed cells were collected by centrifugation at 16,000*g* for 15 min and immediately dounce-homogenized in ice-cold solubilization buffer comprising 50 mM HEPES, pH 7.5, 300 mM NaCl, 1% (w/v) lauryl maltose neopentyl glycol (L-MNG; Anatrace), 0.1% (w/v) cholesteryl hemisuccinate (CHS; Steraloids), 5 mM ATP (Fisher Scientific), 2 mM $MgCl_2$ and 100 μM TCEP. For consOR2–mini-G_{s399} and consOR4–mini-G_{s399}, the solubilization buffer was supplemented with 1 mM *S*-carvone and 30 mM 2MT, respectively. Owing to the low solubility of L-menthol in aqueous buffers, we generated L-menthol-doped L-MNG micelles for consOR1–mini-G_{s399} purification with a ratio of 0.4 mol% L-menthol in 1% w/v L-MNG. This solution was used in place of 1% L-MNG during the purification steps. Solubilized cells were stirred for 1 h at 4 °C, and the detergent-solubilized fraction was clarified by centrifugation at 16,000*g* for 30 min. The detergent-solubilized sample was supplemented with 5 mM $CaCl_2$ and incubated in batch with home-made M1-Flag-antibody-conjugated CNBr-Sepharose under slow rotation for 1.5 h at 4 °C. The Sepharose resin was transferred to a glass column and washed with 15 column volumes of ice-cold buffer comprising 50 mM HEPES, pH 7.5, 300 mM NaCl, 0.05% (w/v) L-MNG or 0.05% (w/v) L-MNG with 0.02 mol% L-menthol, 0.001% (w/v) CHS, 2.5 mM ATP, 4 mM $CaCl_2$, 2 mM $MgCl_2$, 100 μM TCEP and the corresponding odorant. This was followed by 10 column volumes of ice-cold 50 mM HEPES, pH 7.5, 150 mM NaCl, 0.0075% (w/v) L-MNG or 0.0075% (w/v) L-MNG with 0.003 mol% L-menthol, 0.0025% glyco-diosgenin (GDN; Anatrace), 0.001% (w/v) CHS, 4 mM $CaCl_2$, 100 μM TCEP and corresponding odorant. Receptor-containing fractions were eluted with ice-cold 50 mM HEPES, pH 7.5, 150 mM NaCl, 0.0075% (w/v) L-MNG or 0.0075% (w/v) L-MNG with 0.003 mol% L-menthol, 0.0025% (w/v) GDN, 0.001% (w/v) CHS, 5 mM EDTA, 100 μM TCEP, corresponding odorant and 0.2 mg ml^{-1} Flag peptide. Fractions containing the consOR–mini-G_{s399} fusion protein were concentrated in a 50-kDa MWCO spin filter (Amicon) and further purified over a Superdex 200 Increase 10/300 GL (Cytiva) size-exclusion chromatography (SEC) column, which was equilibrated with 20 mM HEPES, pH 7.5, 150 mM NaCl, 0.0075% (w/v) L-MNG or 0.0075% (w/v) L-MNG with 0.003 mol% L-menthol, 0.0025% (w/v) GDN, 0.001% (w/v) CHS, 100 μM TCEP and corresponding odorant. Fractions containing monodisperse consOR–mini-G_{s399} complexes were combined and concentrated in a 50-kDa MWCO spin filter.

Other components of the G protein complex, including $G\beta_1\gamma_2$ and Nb35, were purified as previously described[15,54]. To prepare active-state complexes for cryo-EM, a 3-fold molar excess or 6-fold molar excess of $G\beta_1\gamma_2$ and Nb35 was added to concentrated consOR4–mini-G_{s399} or consOR1–mini-G_{s399} and consOR2–mini-G_{s399} samples, respectively. The resulting preparation was incubated overnight on ice. Complexed samples were purified using a Superdex 200 Increase 10/300 GL SEC column in a buffer comprising 20 mM HEPES, pH 7.5, 150 mM NaCl, 0.0075% (w/v) L-MNG or 0.0075% (w/v) L-MNG with 0.003 mol% L-menthol, 0.0025% GDN, 0.001% w/v CHS and 100 μM TCEP and corresponding odorant. Fractions containing the consOR–G protein complex were collected and concentrated on a 100 kDa MWCO spin filter immediately before cryo-EM grid preparation.

Cryo-EM vitrification, data collection and processing

The purified OR–G_s complex was applied to glow-discharged 300 mesh R1.2/1.3 UltrAuFoil Holey gold support films (Quantifoil). Support films were plunge-frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher) with a 10-s hold period, blot force of 0 and blotting time varying between 1.5 and 3 s while maintaining 100% humidity and 4 °C. Vitrified grids were clipped with Autogrid sample carrier assemblies (Thermo Fisher) immediately before imaging. Movies were recorded using a Titan Krios Gi3 (Thermo Fisher) with a BioQuantum Energy Filter (Gatan) and a K3 Direct Electron Detector (Gatan). Data were collected using SerialEM (v.3.8)[55] running a 3 × 3 image shift pattern at 0° stage tilt. A nominal magnification of ×105,000 with a 100-μm objective was used in super-resolution mode with a physical pixel size as indicated in Supplementary Table 2. Movies were recorded using dose-fractionated illumination with a total exposure of 50 e^- $Å^{-2}$ over 60 frames producing 0.833 e^- $Å^{-2}$ per frame. Movies were motion-corrected and Fourier-cropped to physical pixel size using UCSF MotionCor2 (ref. 56). Dose-weighted micrographs were imported into cryoSPARC (v.4.0.3; Structural Biotechnology)[57], and contrast transfer functions (CTFs) were calculated using the patch CTF estimation tool. Where indicated (Extended Data Figs. 3, 4, 8 and 9), a threshold of CTF fit resolution was used to exclude low-quality micrographs. Particles were template picked using a 20 Å low-pass-filtered model that was generated ab initio from data collected on the consOR51 sample (Extended Data Fig. 3). Particles were extracted with a box size of 288 pixels, binned and sorted by 3D classification with alignment using the heterogeneous refinement tool. Template volumes for each of the four classes were low-pass filtered to 20 Å. The resulting particles were re-extracted with a box size of 288 pixels binned to 144 pixels and where indicated sorted by heterogeneous refinement. Particles from the highest resolution reconstruction were extracted with an unbinned box size of 288 pixels and were subjected to homogeneous refinement followed by non-uniform refinement. Where indicated, particles were exported using csparc2star.py from the pyem (v.0.5) script package[58], and an inclusion mask covering the seven transmembrane domain was generated using the Segger tool in UCSF ChimeraX (v.1.25)[59] and the mask.py tool in pyem (v.0.5). Particles and mask were imported into Relion (v.4.0)[60] and sorted by several rounds of 3D classification without image alignment, in which the number of classes and tau factor were allowed to vary. The resulting particles were brought back into cryoSPARC and subjected to non-uniform refinement. Finally, for all datasets a local refinement using an inclusion mask covering the seven transmembrane domain was performed. Pose and shift Gaussian priors were used with standard deviation of rotational and shift magnitudes set as indicated in Extended Data Figs. 3, 4, 8 and 9.

Site-directed mutagenesis

Generation of OR mutants was performed as previously described[61]. Forward and reverse primers coding for the mutation of interest were obtained from Integrated DNA Technologies. Two successive rounds of PCR using Phusion polymerase (F-549L, Thermo Fisher Scientific) were performed to amplify ORs with mutations. The first round of PCR generated two fragments, one containing the 5′ region upstream of the mutation site and the other containing the 3′ downstream region. The

second PCR amplification joined these two fragments to produce a full open reading frame of the OR. PCR products with desired length were gel purified and cloned into the MluI and NotI sites of the mammalian expression vector pCI (Promega) that contains rho-tag. Plasmids were purified using a ZymoPure miniprep kit (D4212).

cAMP signalling assays

GloSensor cAMP assays (Promega) were used to determine real-time cAMP levels downstream of OR activation in HEK293T cells, as previously described[62]. HEK293T cells (authenticated by short tandem repeat profiling and tested negative for mycoplasma contamination) were cultured in minimum essential medium (MEM; Corning) supplemented with 10% FBS (Gibco), 0.5% penicillin–streptomycin (Gibco) and 0.5% amphotericin B (Gibco). Cultured HEK293T cells were plated the day before transfection at 1/10 of 100% confluence from a 100-mm plate into 96-well plates coated with poly-D-lysine (Corning) or tissue-culture coated 96-well plates with 0.001% poly-D-lysine (Sigma). For each 96-well plate, 1,000 ng pGloSensor-20F plasmid (Promega), 500 ng of RTP1S plasmid (only for OR1A1 and its mutants) and 7,500 ng of rho-tagged OR (250 ng for OR51E1(M158A), 750 ng for consOR1, 75 ng for consOR51, consOR2 and consOR4) in the pCI mammalian expression vector (Promega) were transfected 18–24 h before odorant stimulation using Lipofectamine 2000 (11668019, Invitrogen) in MEM supplemented with 10% FBS. On stimulation day, plates were injected with 25 µl of GloSensor substrate (Promega) and incubated for 2 h in the dark at room temperature and in an odour-free environment. Odorants were diluted to the desired concentration in CD293 medium (Gibco) supplemented with copper (30 µM $CuCl_2$; Sigma-Aldrich) and 2 mM L-glutamine (Gibco) and the pH adjusted to 7.0 with a 150 mM solution of sodium hydroxide (Sigma-Aldrich). After injecting 25 µl of odorants in CD293 medium into each well, GloSensor luminescence was immediately recorded for 10 or 20 cycles of monitoring over a total period of 15 or 30 min using a BMG Labtech POLARStar Optima plate reader. The resulting luminescence activity was normalized to an empty vector negative control, and the OR response was obtained by calculating the AUC by summing the response from all cycles. Dose-dependent responses of ORs were analysed by fitting a least squares function to the data and by generating EC_{50} values and efficacy using GraphPrism 10. The area under the dose–response curve was then calculated by summing the response from each concentrations.

Evaluating cell surface expression

Flow cytometry was used to evaluate cell surface expression of ORs as previously described[23]. HEK293T cells were seeded onto 35-mm plates (Greiner Bio-One) with approximately 3.5×10^5 cells (25% confluency). The cells were cultured overnight. After 18–24 h, 1,200 ng of ORs (120 ng for consOR1, OR51 subfamily members and their mutants) tagged with the first 20 amino acids of human rhodopsin (rho-tag) at the N-terminal ends in pCI mammalian expression vector (Promega), 300 ng of RTP1s (only for OR1A1 and its mutants) and 30 ng eGFP were transfected using Lipofectamine 2000 (11668019, Invitrogen). At 18–24 h after transfection, the cells were detached and resuspended using Cell stripper (Corning) and then transferred into 5-ml round-bottom polystyrene tubes (Falcon) on ice. The cells were spun down at 4 °C and resuspended in PBS (Gibco) containing 15 mM NaN_3 (Sigma-Aldrich) and 2% FBS. (Gibco). They were stained with 1/400 (v/v) of primary antibody mouse anti-rhodopsin clone 4D2 (MABN15, Sigma-Aldrich) and incubated for 30 min, then washed with PBS containing 15 mM NaN_3 and 2% FBS. The cells were spun again and then stained with 1/200 (v/v) of the phycoerythrin-conjugated donkey anti-mouse F(ab')2 fragment antibody (715-116-150, Jackson Immunologicals) and incubated for 30 min in the dark. To label dead cells, 1/500 (v/v) of 7-amino-actinomycin D (129935, Calbiochem) was added. The cells were then immediately analysed using a BD FACSCanto II flow cytometer with gating allowing for GFP-positive, single, spherical, viable cells, and the measured phycoerythrin fluorescence intensities were analysed and visualized using FlowJo (v.10.8.1). Empty plasmid pCI is used as negative control.

Homology model of OR1A1 and docking studies

The OR1A1 homology model was generated using the consOR1 bound to L-menthol and G protein cryo-EM structure as a template with Schrödinger Maestro (v.2022-2). The consOR1 cryo-EM structure was prepared using the protein preparation wizard, which involved adding missing side chains and hydrogen atoms. Subsequently, the model was refined through hydrogen bond assignment and energy minimization. A pairwise alignment of consOR1 and OR1A1 sequences was then performed, revealing 64% sequence identity and identifying a gap at position 194 in consOR1 compared to OR1A1. Last, the knowledge-based method within the build homology model module was used to create the OR1A1 homology model, and the corresponding homology model was further energy minimized. The OR51E1 homology model was made using a similar procedure using the consOR51 structure.

For docking studies of *R*-carvone and L-menthol into the OR1A1 homology model, we followed the Schrödinger induced fit docking protocol with the following steps. (1) Constrained minimization of the receptor with a RMSD cut-off of 0.18 Å. (2) A RMSD alignment of the consOR1 EM structure onto the OR1A1 homology model was performed, followed by definition of a 25 × 25 × 25 Å docking grid box centred on the position of L-menthol in consOR1. This step was followed by initial Glide docking of *R*-carvone and L-menthol using a softened potential and removal of side chains that are within 5 Å of L-menthol. (3) A prime side-chain prediction for each receptor–ligand pose to rebuild the side chain conformation. (4) A prime minimization on the receptor–ligand complex. (5) After removal of the ligand, a rigid Glide redocking was performed to re-dock the ligand back into the ligand binding site. (6) Estimation of the binding energy.

MD simulations

Simulations were performed similar to previous methods[15] using the GROMACS package (v. 2022)[63] and the CHARMM36m force field[64]. The following simulation systems were constructed: consOR1-apo, consOR1–L-menthol bound, consOR1–L-menthol–mini-$G\alpha_s$ subunit bound, consOR2–*S*-carvone–mini-$G\alpha_s$ subunit bound, consOR2-apo, consOR4–2MT–mini-$G\alpha_s$ subunit bound and consOR4-apo. For the mini-$G\alpha_s$ subunit bound simulations, the $G\beta\gamma$ subunit was removed from the cryo-EM structure to reduce computational time. All ligands were parameterized using ParaChem (v.3.0)[65]. The GPCR structures were prepared using the Maestro Schrödinger (v.2022-2) protein preparation wizard module. Missing side chains and hydrogen atoms were added, protein termini were capped with neutral acetyl and methylamide groups, and histidine states were assigned. The complex was then minimized. The simulation box was created using CHARMM-GUI[66,67]. The GPCRs were aligned in the bilayer using the PPM 2.0 function of the Orientation of Proteins in Membranes (OPM) tool[68], and the bilayer was filled with 75% palmitoyl-oleoyl-phosphatidylcholine (POPC) and 25% cholesteryl hemisuccinate deprotonated (CHSD). The initial positions of CHSD were taken from our previous study of OR51E2 (ref. 15). TIP3P water molecules were used for solvation, whereas 0.15 M potassium chloride ions were added to neutralize the system box. The final system dimensions were approximately 85 × 85 × 110 Å without the Gα subunit and 100 × 100 × 150 Å with the mini-$G\alpha_s$ subunit.

The system was minimized with position restraints (10 kcal mol^{-1} $Å^{-2}$) on all heavy atoms of the protein, ligand and head group atoms of lipids, followed by a 1 ns heating step that raised the temperature from 0 K to 310 K in the NVT ensemble using the Nosé–Hoover thermostat. Next, an 80-ns-long equilibration in the NPT ensemble was performed. During the heating step and the long equilibration, the same position restraints of 10 kcal mol^{-1} $Å^{-2}$ were applied for the first 1 ns, then reduced to 5 kcal mol^{-1} $Å^{-2}$, and gradually to 0 kcal mol^{-1} $Å^{-2}$ in steps

of 1 kcal mol^{-1} $Å^{-2}$, with 5 ns of simulations per equilibration window. Afterward, a 50 ns non-restrained equilibration was conducted.

The final snapshot of the equilibration step served as the initial conformation for five production runs, which were initiated with randomly generated velocities. Pressure was coupled to a 1 bar pressure bath and controlled using the Parrinello–Rahman method[69]. Throughout all simulations, the LINCS algorithm was applied to all bonds and angles of water molecules, with a 2 fs time step used for integration. Additionally, a 12 Å cut-off was used for non-bonded interactions, and the particle mesh Ewald method[70] treated long-range L–J interactions. MD snapshots were saved every 20 ps, and all MD analyses were conducted on the aggregated trajectories for each system from the 5 runs (totalling 5 × 1,000 ns = 5000 ns) using VMD (v.1.9.4), PyMOL (v.2.5), GROMACS modules (v.2019–2022) and Python scripts.

Ligand–receptor interactions analysis. Ligand–receptor contact frequencies were determined using the get_contacts script (https://getcontacts.github.io/). Measurements were carried out on trajectories that included solvents. All types of contacts were taken into account, encompassing water-mediated contacts as well. Contact frequencies were visualized as heatmaps using the matplotlib library.

Ligand-binding site volume calculations. To calculate the volume of the ligand-binding site, we first performed protein conformational clustering using the GROMACS cluster module. Clustering was conducted on the Cα atoms of proteins, adjusting the RMSD cut-off to between 1.5 Å and 1.9 Å until the top 5–10 clusters encompassed more than 60% of all sampled points. The centroid structure of each of the top clusters was then used for ligand-binding site volume calculation. Volume calculations were executed using the Maestro SiteMap module, defining the ligand-binding pocket as being within 6 Å of the ligand. For structures obtained from apo simulations, docking was first performed to insert L-menthol into the pocket. Subsequently, these docked structures underwent the same volume calculation protocol as the others. In the volume calculation, a more restrictive definition of hydrophobicity and fine grid was applied, and the ligand-binding site map was cropped at 4 Å from the nearest site point. The calculated volumes from the top cluster structures were utilized to compute the average and standard deviation of the ligand-binding site volume.

Ligand flexibility analysis. Ligand RMSD values were calculated using the MDAnalysis script, which initially aligned the structure based on protein Cα atoms. Then, for each simulation frame, the RMSD matrix was computed using the coordinates of all ligand heavy atoms. Both rotational and translational movements of the ligand were taken into account. The resulting RMSD values were used to calculate the average RMSD for ligands.

$D/E^{45\times51}$–$Y^{6\times55}$ distance analysis. The distance between the D–Y motif was calculated using the MDAnalysis script. This distance was measured as the minimum distance between the carboxylate oxygens of $D/E^{45\times51}$ (OD1, OD2 or OE1, OE2 in the CHARMM force field) and the hydroxyl oxygen (OH in the CHARMM force field) of $Y^{6\times55}$ in consOR1, consOR2 and consOR4 production trajectories, respectively. The time evaluation of distance from a selected velocity was plotted as a moving average and rolling standard deviation using the matplotlib and scipy library. The overall distances from the production trajectory was represented as a violin plot using matplotlib.

Phylogenetic tree, sequence identity and structure comparison
On R (v.4.3.1), alignment reading and matrix of distance between sequences (by sequence identity) calculation were performed using the Biostrings (v.3.19)[71] and seqinr (v.4.2-36)[72] packages. Neighbour-joining tree and tree visualization were realized with packages ape (v.5.8)[73] and ggtree (v.3.12.0)[74], and the tree was plotted unrooted with the daylight method. Sequence identity and RMSD between structures were calculated using the package bio3D (v.2.4-4) and graphs were made with pheatmap (v.1.0.12) and gtools (v.4.8.0) packages. Conserved positions in aligned sequences of class I and class II ORs were visualized with WebLogo3 (ref. 75).

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Coordinates for consOR51, consOR1, consOR2 and consOR4 have been deposited into the RCSB's Protein Data Bank under accession codes 8UXV, 8UXY, 8UY0 and 8UYQ, respectively. EM density maps for consOR51, consOR1, consOR2 and consOR4 have been deposited into the Electron Microscopy Data Bank under accession codes EMD-42786, EMD-42789, EMD-42791 and EMD-42817, respectively. The locally refined maps for the seven transmembrane domain for consOR51, consOR1, consOR2 and consOR4 have been deposited into the Electron Microscopy Data Bank under accession codes EMD-45881, EMD-45876, EMD-45879 and EMD-45880, respectively. The MD simulation trajectories have been deposited into the GPCRmd database under access codes 2056 (consOR1 with ligand), 2069 (consOR1 with ligand and mini-Gs), 2058 consOR2 apo), 2059 (consOR2 with ligand and mini-Gs), 2060 (consOR4-apo), 2061 (consOR4 with ligand and mini-Gs), 1245 (OR51E2 with ligand) and 2062 (OR51E2 with ligand and mini-Gs).

54. Faust, B. et al. Autoantibody mimicry of hormone action at the thyrotropin receptor. *Nature* **609**, 846–853 (2022).
55. Mastronarde, D. N. SerialEM: a program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microsc. Microanal.* **9**, 1182–1183 (2003).
56. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
57. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
58. Asarnow, D., Palovcak, E. & Cheng, Y. asarnow/pyem: UCSF Pyem v0.5. *Zenodo* https://doi.org/10.5281/zenodo.3576630 (2019).
59. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
60. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
61. Bushdid, C., de March, C. A., Matsunami, H. & Golebiowski, J. Numerical models and in vitro assays to study odorant receptors. *Methods Mol. Biol.* **1820**, 77–93 (2018).
62. Zhang, Y., Pan, Y., Matsunami, H. & Zhuang, H. Live-cell measurement of odorant receptor activation using a real-time cAMP assay. *J. Vis. Exp.* **128**, 55831 (2017).
63. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).
64. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
65. Vanommeslaeghe, K. et al. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
66. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
67. Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **97**, 50–58 (2009).
68. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
69. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
70. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an *N*·log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
71. Pagès, H., Aboyoun, P., Gentleman, R. & DebRoy, S. Biostrings: Efficient manipulation of biological strings. R package version 2.72.1 https://bioconductor.org/packages/Biostrings (2022).
72. Charif, D. & Lobry, J. R. in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations* (eds Bastolla, U. et al.) 207–232 (Springer Berlin Heidelberg, 2007).
73. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
74. Xu, S. et al. *Ggtree*: a serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta* **1**, e56 (2022).

75. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
76. Dang, S. et al. Cryo-EM structures of the TMEM16A calcium-activated chloride channel. *Nature* **552**, 426–429 (2017).

Acknowledgements This work was supported by National Institutes of Health (NIH) grants R01DC020353 (H.M., N.V. and A.M.) and K99DC018333 (C.A.d.M.), and the Agence Nationale de la Recherche grant ANR-23-CE44-0004 (C.A.d.M.). Cryo-EM equipment at UCSF is partially supported by NIH grants S10OD020054 and S10OD021741. Some of this work was performed at the Stanford-SLAC Cryo-EM Center (S2C2), which is supported by the National Institutes of Health Common Fund Transformative High-Resolution Cryo-Electron Microscopy program (U24 GM129541). This project was funded by the UCSF Program for Breakthrough Biomedical Research, funded in part by the Sandler Foundation. H.M. acknowledges support from National Science Foundation grant 2014217. A.M. acknowledges support from the Edward Mallinckrodt, Jr Foundation and the Vallee Foundation. A.M. is a Chan Zuckerberg Biohub San Francisco Investigator.

Author contributions C.A.d.M. and H.M. designed the consOR strategy, and with A.M., outlined a structure determination strategy for consORs. The study was also designed by C.B.B., N.M., W.J.C.v.d.V. and N.V. Consensus constructs were designed and cloned by C.A.d.M. with input from A.M. C.A.d.M generated the phylogenetic trees. C.B.B. led the effort for structure determination, including cloning constructs, preparing baculoviruses, expressing and purifying G protein complexing reagents. With C.L.d.T., C.B.B. established conditions to biochemically purify and stabilize consOR1, consOR2 and consOR4 complexes, identified optimal cryo-EM grid preparation procedures and collected cryo-EM datasets for structure determination. L.L. assisted with protein purifications. A.M. purified the consOR51 complex and collected cryo-EM data with help from B.F. C.B.B. determined high-resolution cryo-EM maps by extensive image processing with input from A.M. and C.L.d.T. A.M. and C.B.B. built and refined models of consOR complexes. C.A.d.M. and J.T. analysed OR models and sequences to design and clone OR mutants. C.A.d.M., J.T. and I.O. performed GloSensor signalling experiments for OR functional activity, and C.A.d.M., J.T., I.T and I.O. generated OR cell surface expression data by flow cytometry with input from H.M. C.A.d.M. and J.T. analysed and prepared figures and tables for signalling and flow cytometry data. N.M. set up MD simulations, ligand docking and performed binding pocket volume calculations. W.J.C.v.d.V. created the homology model of OR1A1. N.M. and W.J.C.v.d.V. analysed simulation trajectories and prepared figures describing simulation data. N.M., W.J.C.v.d.V. and N.V. provided mechanistic insight from simulation data. C.A.d.M. performed the comparative structure analysis. C.A.d.M., N.M. and A.M. wrote an initial draft of the manuscript and generated figures with contributions from all authors. Further edits to the manuscript were provided by W.J.C.v.d.V., N.M., N.V. and H.M. The overall project was supervised and funded by C.A.d.M., N.V., H.M. and A.M.

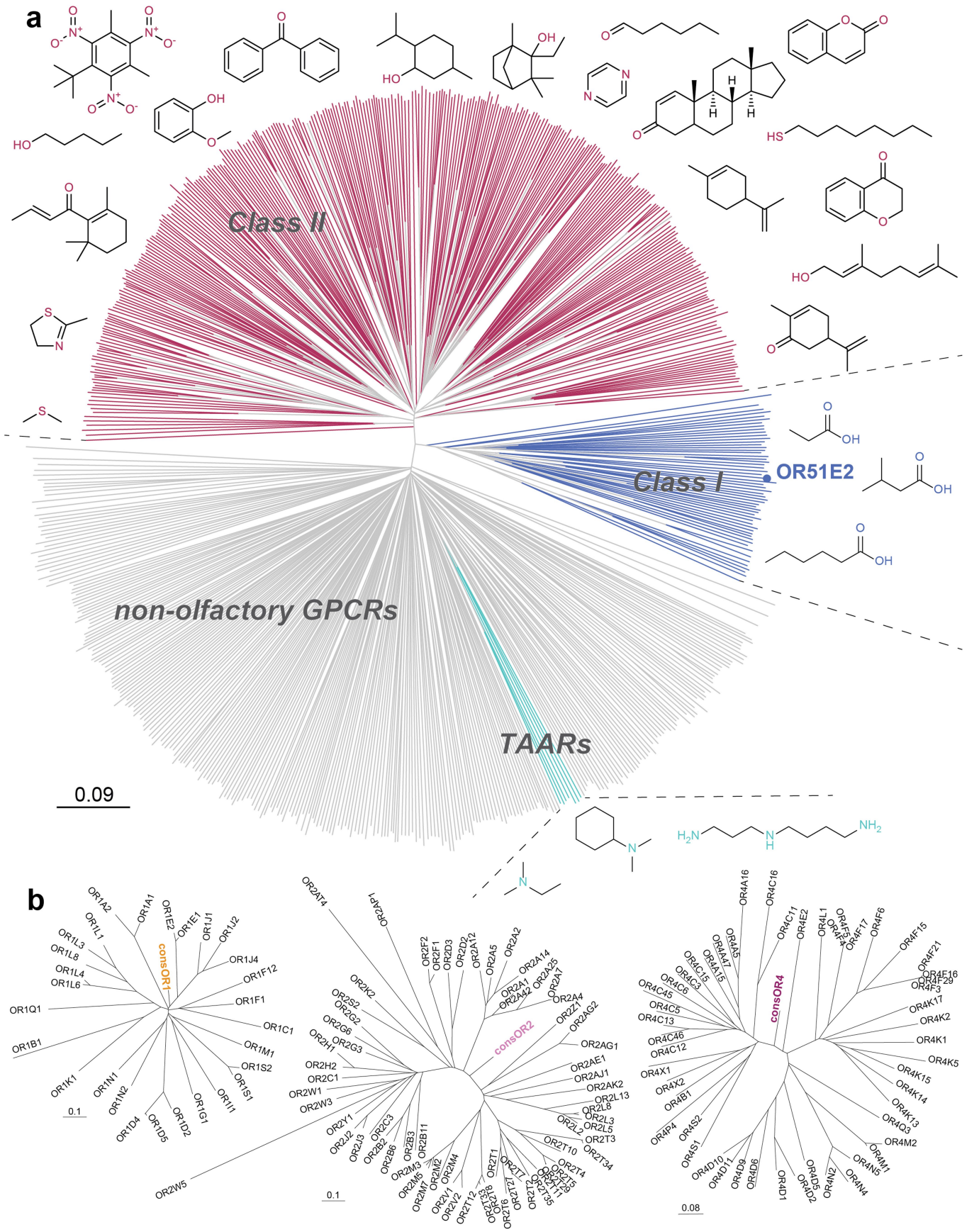
Competing interests H.M. has received royalties from Chemcom, research grants from Givaudan and consultant fees from Kao. A.M. is a founder of Epiodyne and Stipple Bio, consults for Abalone and serves on the scientific advisory board of Septerna.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08126-0.

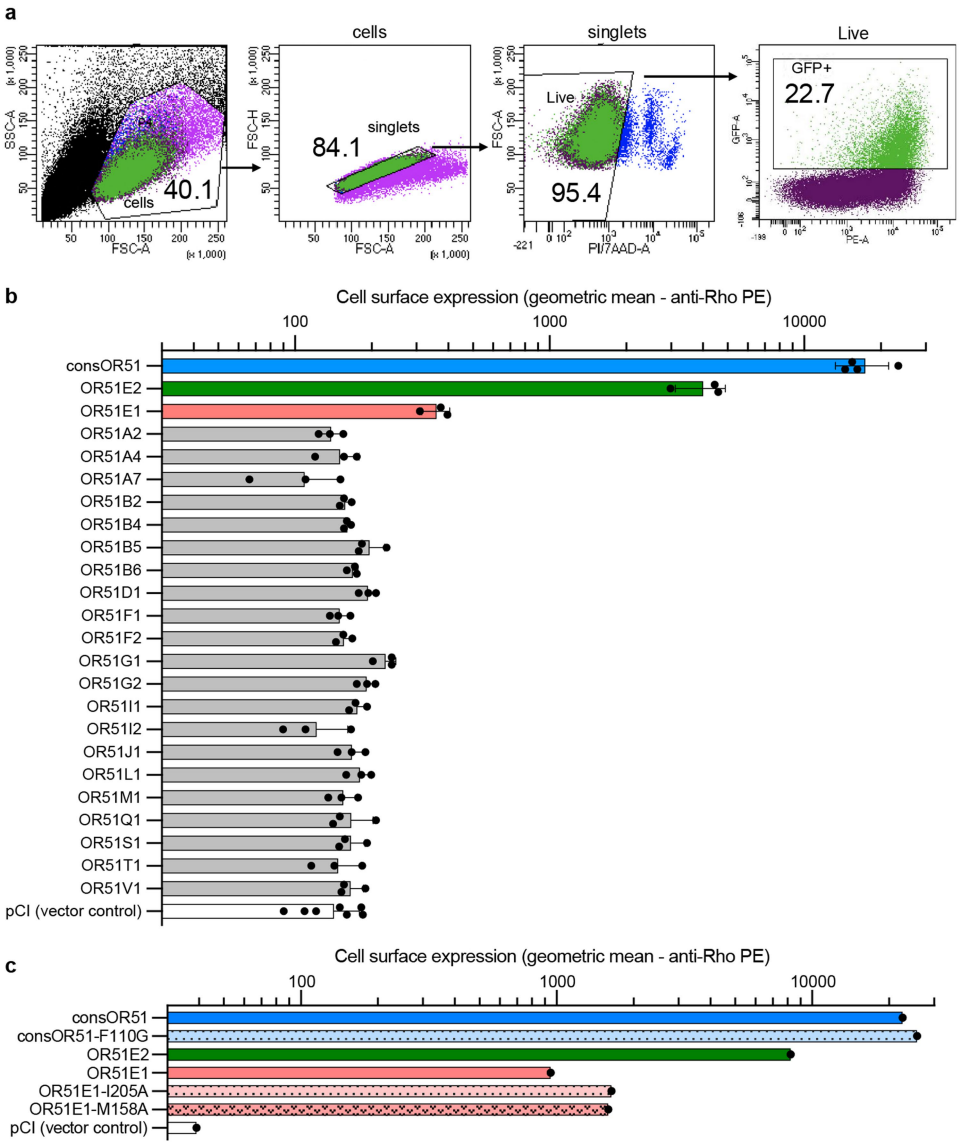
Correspondence and requests for materials should be addressed to Claire A. de March, Nagarajan Vaidehi, Aashish Manglik or Hiroaki Matsunami.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

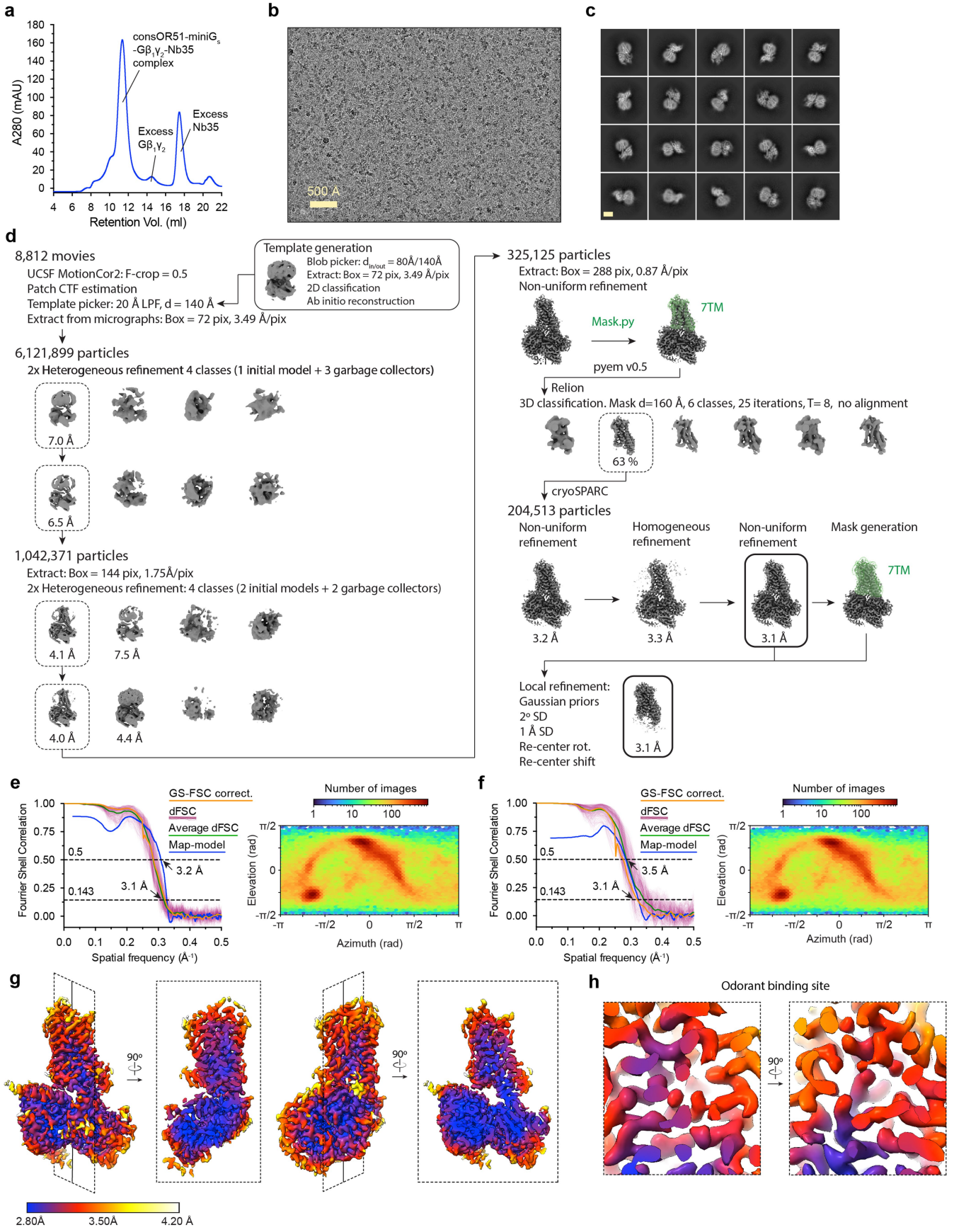


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Phylogenetic tree of human Class A GPCRs and consORs. a, Olfactory GPCRs are highlighted in pink (class II ORs), blue (class I ORs) and cyan (TAARs). Non-olfactory GPCRs are represented with grey lines. A subset of ligands recognized by the olfactory GPCRs are presented, with heteroatoms coloured by the olfactory GPCR families responsible for detection. OR51E2, the only human odorant receptor known structure is highlighted. On the bottom left, the scale represents the amount of amino acid change for a set distance. **b**, Phylogenetic trees for human OR subfamilies OR1, OR2 and OR4. Consensus ORs (consORs) are indicated in colour.

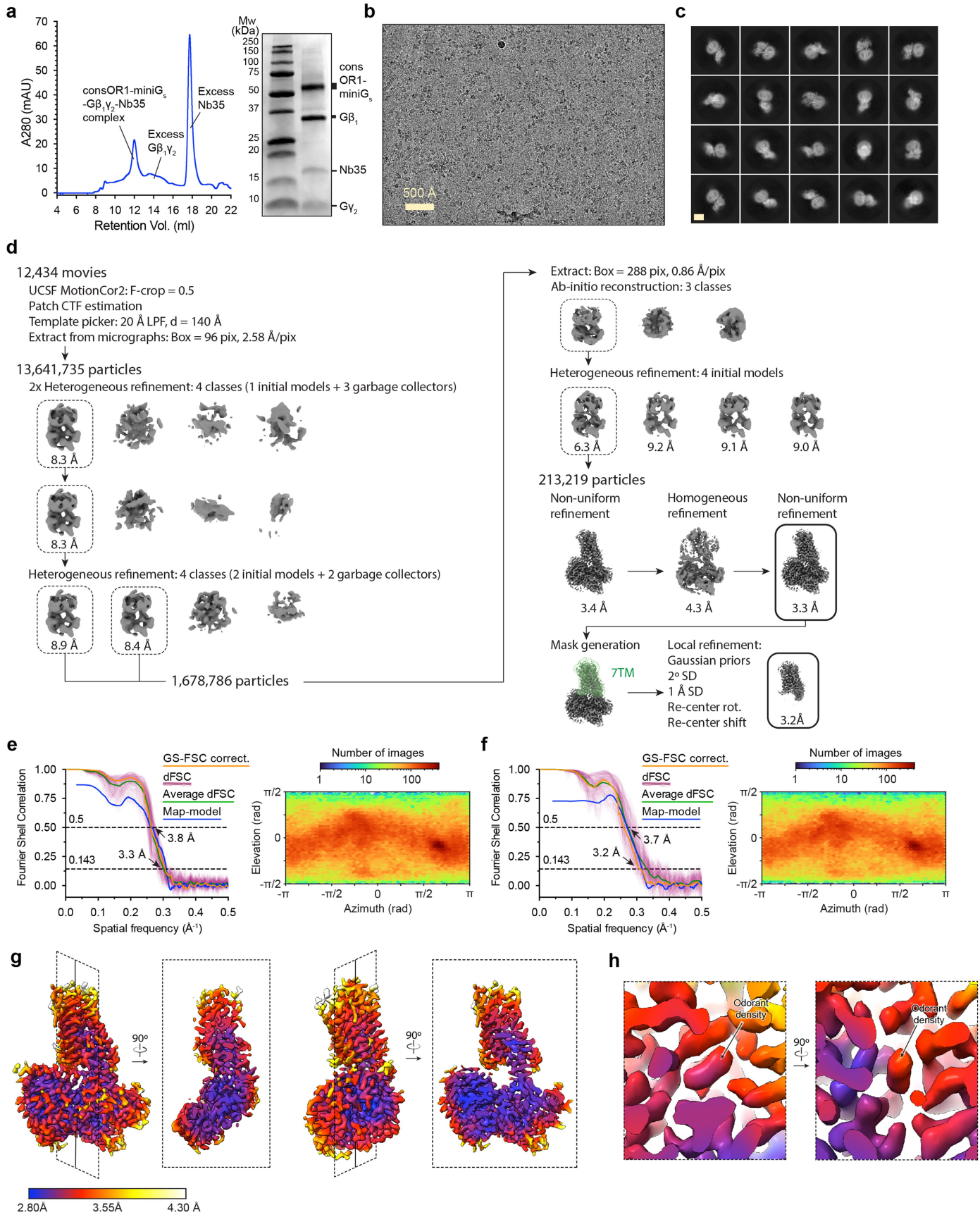


Extended Data Fig. 2 | Expression of OR51 family members. a, Flow cytometry gating strategy on a sample expressing OR51E2. **b**, Cell surface expression of OR51 family members monitored by flow cytometry on 30,000 cells (n = 30,000 cells examined over more than 3 independent experiments; data are presented as mean values +/− SD). **c**, Cell surface expression of consOR51 and F110G mutant as well as OR51E1 and its mutants M158A and I205A monitored by flow cytometry on 10,000 cells.



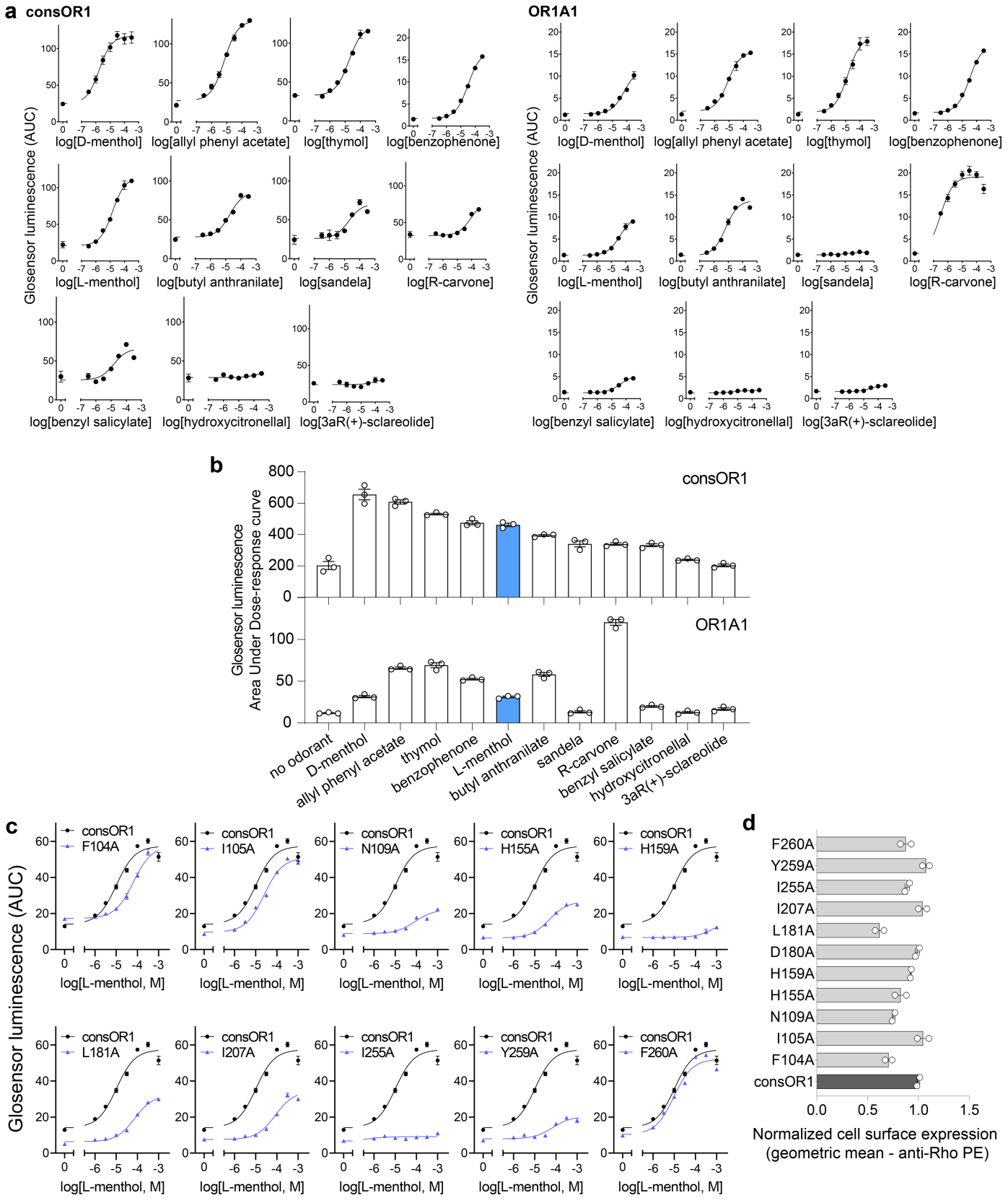
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Cryo-EM data processing for consOR51-G_s. a, Size exclusion chromatogram of consOR1-G_s sample. **b**, A representative cryo-EM micrograph from the curated consOR51-G_s dataset (n = 8,812) obtained from a Titan Krios microscope. **c**, A subset of highly populated, reference-free 2D-class averages are shown. Scale bar is 50 Å. **d**, Schematic showing the image processing workflow for consOR51-G_s. Initial processing was performed using UCSF MotionCor2 and cryoSPARC. Particles were then transferred using the pyem script package[58] to RELION for alignment-free 3D classification. Finally, particles were processed in cryoSPARC using the non-uniform and local refinement tools. Dashed boxes indicate selected classes, and 3D volumes of classes and refinements are shown along with global Gold-standard Fourier Shell Coefficient (GSFSC) resolutions. **e**,**f**, Map validation for the consOR51-G_s (**e**) globally refined, and (**f**) locally refined cryo-EM maps. Gold-standard Fourier shell correlation (FSC) curves are calculated in cryoSPARC, and shown together with directional FSC (dFSC) curves generated with dfsc.0.0.1.py as previously described[76]. Map-model correlations calculated in the Phenix suite are also shown. Arrows indicate map and map-model resolution estimates at 0.143 and 0.5 correlation respectively. Euler angle distributions calculated in cryoSPARC are also provided for each map. **g**,**h**, Local resolution estimation for the consOR51-G_s (**g**) globally refined, and (**h**) locally refined cryo-EM maps.



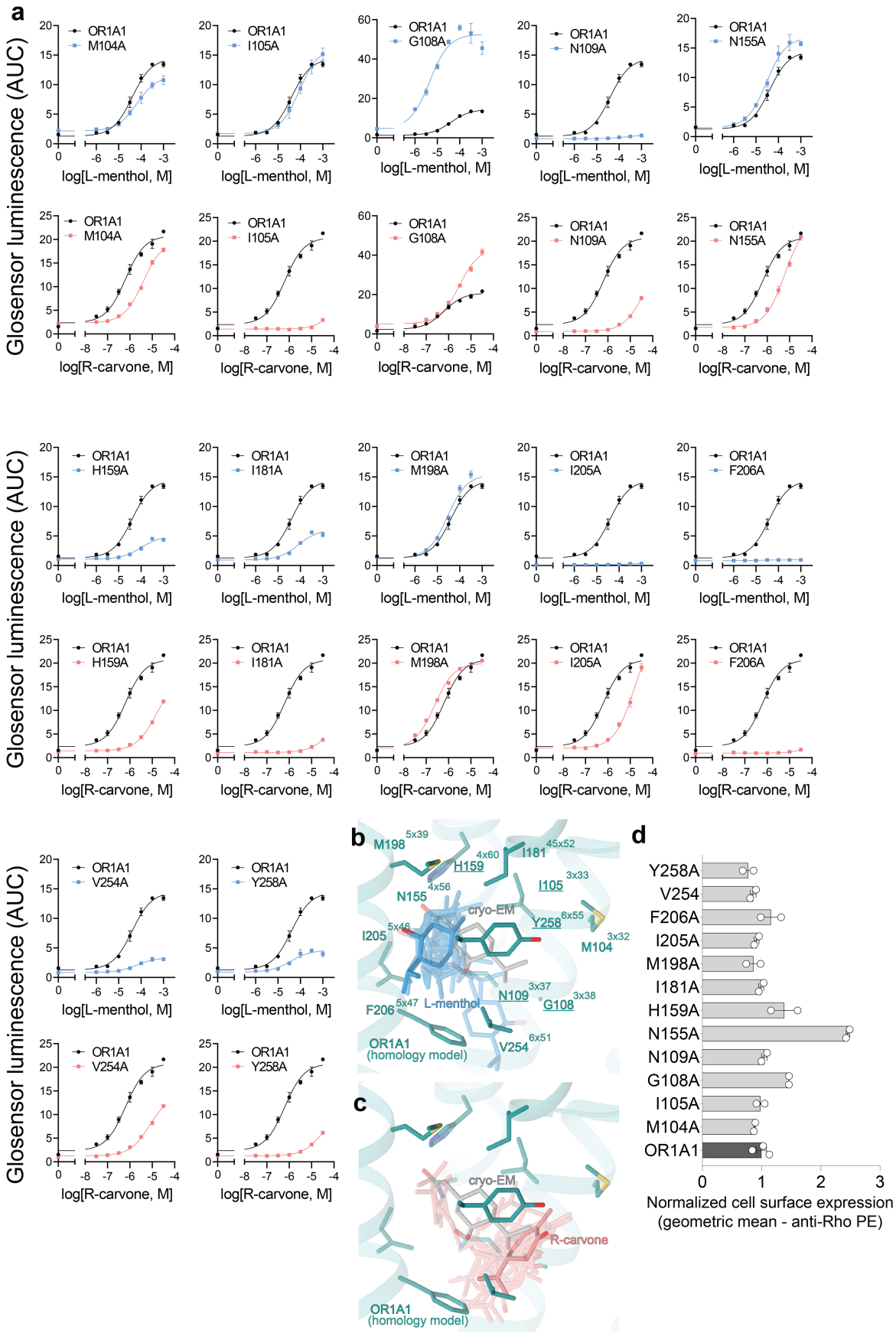
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Cryo-EM data processing for consOR1-G_s. a, Size exclusion chromatogram and SDS-PAGE of consOR1-G_s sample. **b**, A representative cryo-EM micrograph from the curated consOR1-G_s dataset (n = 12,434) obtained from a Titan Krios microscope. **c**, A subset of highly populated, reference-free 2D-class averages are shown. Scale bar is 50 Å. **d**, Schematic showing the image processing workflow for consOR1-G_s. Initial processing was performed using UCSF MotionCor2 and cryoSPARC. Particles were then transferred using the pyem script package[58] to RELION for alignment-free 3D classification. Finally, particles were processed in cryoSPARC using the non-uniform and local refinement tools. Dashed boxes indicate selected classes, and 3D volumes of classes and refinements are shown along with global Gold-standard Fourier Shell Coefficient (GSFSC) resolutions. **e**,**f**, Map validation for the consOR1-G_s (**e**) globally refined, and (**f**) locally refined cryo-EM maps. Gold-standard Fourier shell correlation (FSC) curves are calculated in cryoSPARC, and shown together with directional FSC (dFSC) curves generated with dfsc.0.0.1.py as previously described[76]. Map-model correlations calculated in the Phenix suite are also shown. Arrows indicate map and map-model resolution estimates at 0.143 and 0.5 correlation respectively. Euler angle distributions calculated in cryoSPARC are also provided for each map. **g**,**h**, Local resolution estimation for the consOR1-G_s (**g**) globally refined, and (**h**) locally refined cryo-EM maps.



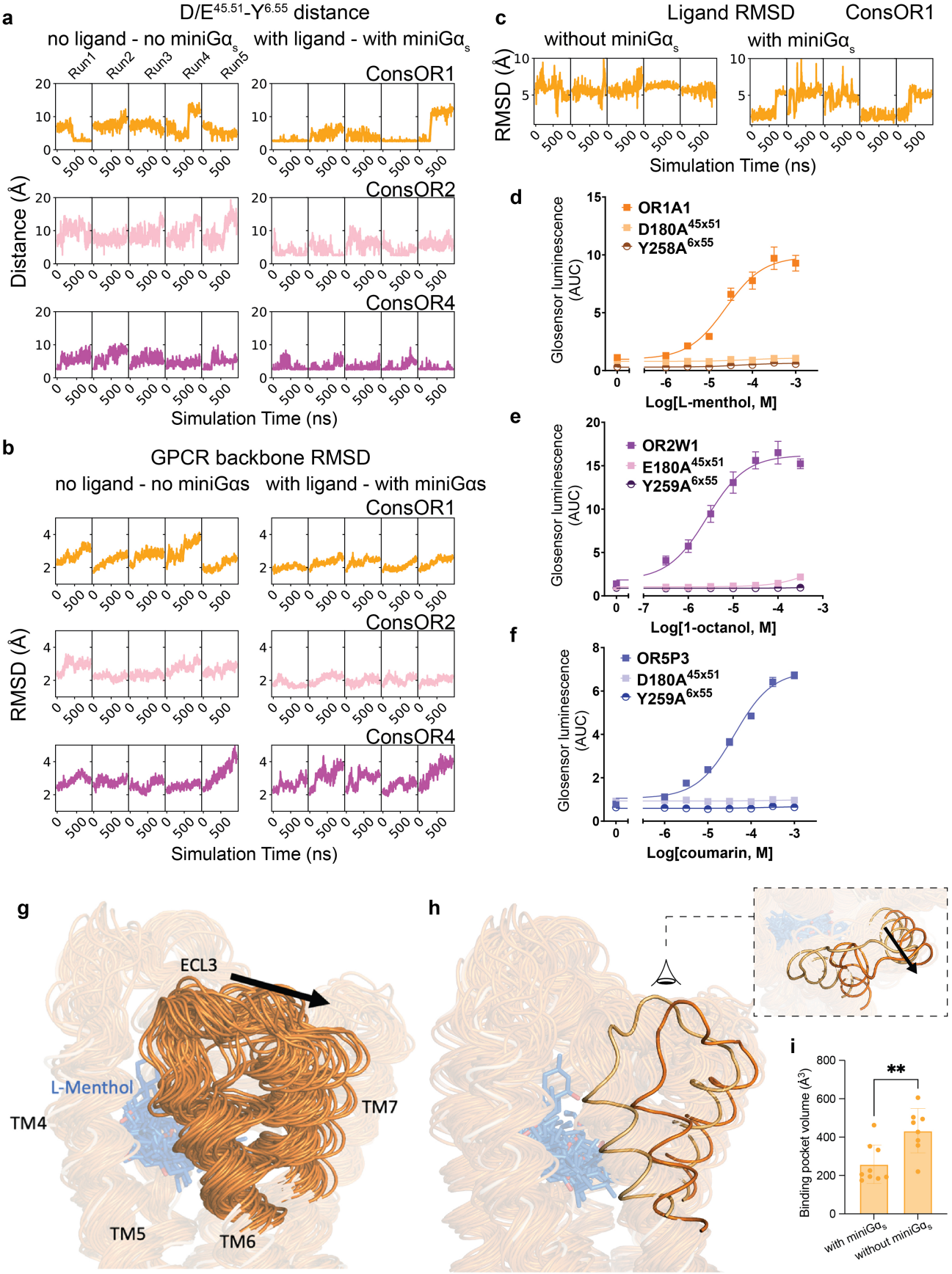
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | consOR1 and OR1A1 activity against odorants and consOR1 mutation screening. a, consOR1 and OR1A1 response in cAMP accumulation assay to 11 odorant stimulations in dose response (data normalized by the empty vector and the Area Under Dose-response curve of a repeated consOR1 response to L-menthol and pCI, concentration in mol.L^{-1}, n = 3 biologically independent samples). **b**, consOR1 and OR1A1 are activated by diverse odorants as measured by a Glosensor cAMP production assay. Area under the dose response curve (AUC) was calculated from response normalized to the negative control and the AUC of consOR1 response to L-menthol repeated in each plate (n = 3). **c**, consOR1 and its mutants response in cAMP accumulation assay to L-menthol (data normalized by the Area Under Dose-response curve of a repeated consOR1 response to L-menthol and pCI, concentration in mol.L^{-1}, n = 3 biologically independent samples). **d**, Cell surface expression of consOR1 and its mutants evaluated in flow cytometry and normalized by consOR1 (120 ng of DNA, n = 10,000 cells per experiment, examined over 2 independent experiments). For **a** to **d**, data are presented as mean values +/− SEM.



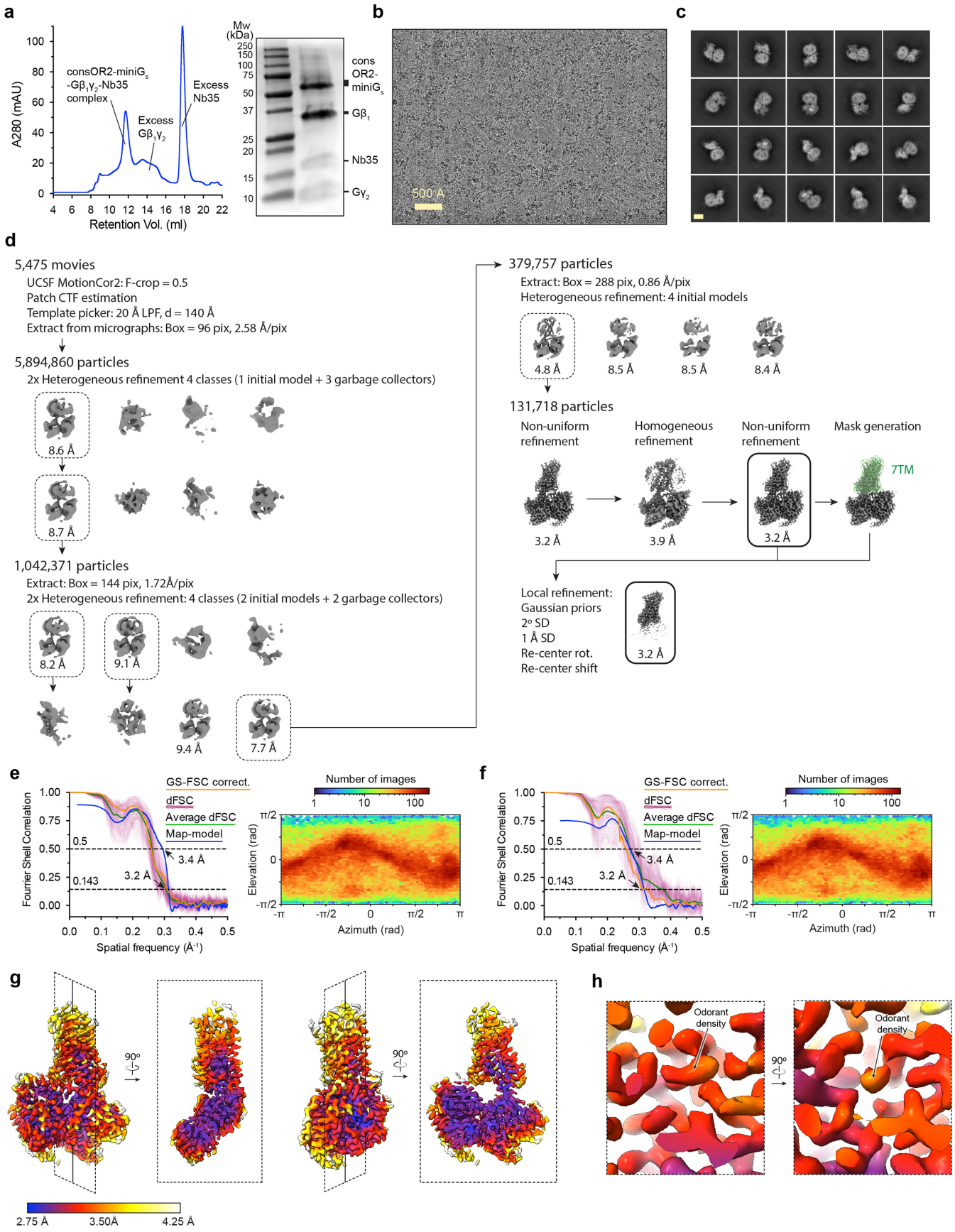
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | OR1A1 mutants activity against odorants. a, OR1A1 and its mutants response in cAMP accumulation assay to L-menthol and R-carvone (data normalized by the Area Under Dose-response curve of OR1A1 response to L-menthol and R-carvone, respectively, concentration in mol.L^{-1}, n = 3 biologically independent samples). **b**,**c**, Docking of L-menthol (**b**) and R-carvone (**c**) to the homology model of OR1A1. Top scoring docking results are shown for both odorants as transparent sticks. The best scoring pose is shown as solid sticks. **d**, Cell surface expression of OR1A1 and its mutants evaluated in flow cytometry and normalized to OR1A1 (1200 ng of DNA, n = 10,000 cells per experiment, examined over 2 independent experiments). For **a** and **d**, data are presented as mean values +/− SEM.



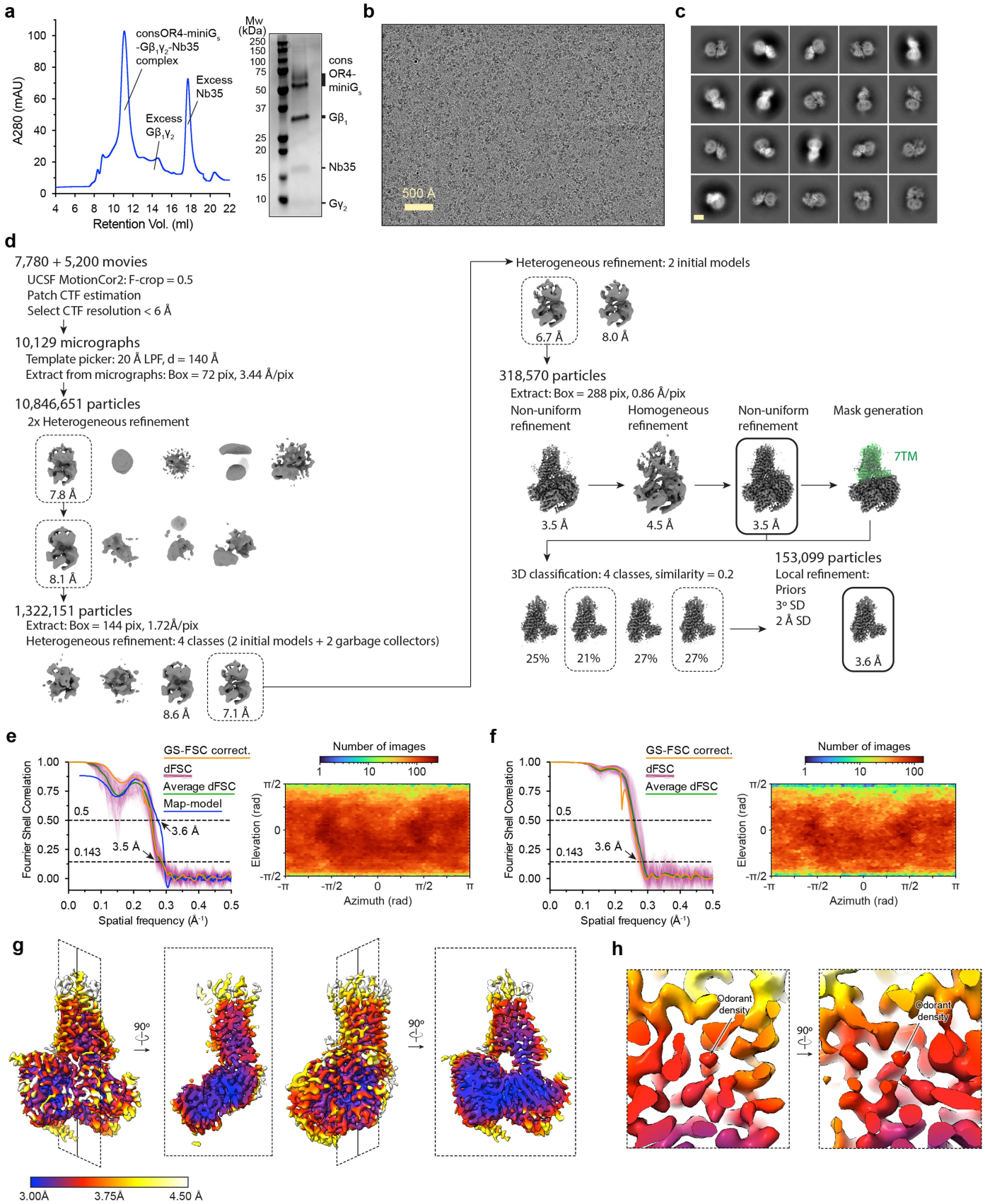
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | D/E$^{45\times51}$-Y$^{6\times55}$ motif. Left, time evolution of D/E$^{45\times51}$-Y$^{6\times55}$ distance and GPCR/Ligand RMSD from MD simulations. All measurements were performed on MD trajectories by skipping every 100 frames. The raw data points were plotted transparent. A fitted curve was plotted as a solid line by using a smoothing window of 50 sampling points. **a**, The distance between D/E$^{45\times51}$ and Y$^{6\times55}$ from MD simulations without ligand/miniGα_s and with ligand/miniGα_s states for consOR1, consOR2, consOR4. The atoms involved in this measurement are described in the Methods. **b**, The GPCR backbone atoms RMSD from MD simulations without ligand/miniGα_s and with ligand/miniGα_s states for consOR1, consOR2, consOR4. **c**, The ligand RMSD from consOR1 bound to L-menthol MD simulations without miniGα_s, and with miniGα_s. Right, effect of mutations at positions 45×51 and 6×55. **d**–**f**, Three native ORs, **d**) OR1A1, **e**) OR2W1, **f**) OR5P3, showed a decrease of their response to agonist when positions 45×51 and 6×55 are mutated in alanine (n = 3 biologically independent samples, data are presented as mean values +/− SEM). **g**, Sampling of conformations in consOR1 bound to menthol without miniGα_s MD simulations shows a motion of the extracellular parts of TM6 and TM7 as well as ECL3 away from the receptor bundle. **h**, Side view of extreme conformations; insert represents top view. **i**, consOR1 binding pocket volumes during MD simulations are bigger when the ligand-bound receptor is not bound to miniGα_s than when it is fully complexed.(n = 9 representative snapshots from simulation, data are presented as mean values +/− SEM, p = 0.00453, ** indicates p < 0.01).



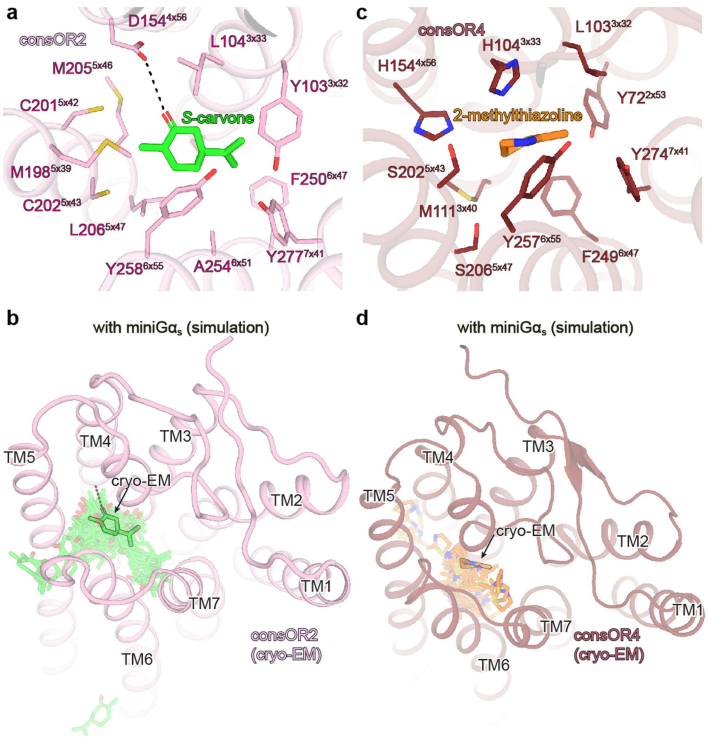
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Cryo-EM data processing for consOR2-G_s. a, Size exclusion chromatogram and SDS-PAGE of consOR2-G_s sample. **b**, A representative cryo-EM micrograph from the curated consOR2-G_s dataset (n = 5,475) obtained from a Titan Krios microscope. **c**, A subset of highly populated, reference-free 2D-class averages are shown. Scale bar is 50 Å. **d**, Schematic showing the image processing workflow for consOR2-G_s. Initial processing was performed using UCSF MotionCor2 and cryoSPARC. Particles were then transferred using the pyem script package[58] to RELION for alignment-free 3D classification. Finally, particles were processed in cryoSPARC using the non-uniform and local refinement tools. Dashed boxes indicate selected classes, and 3D volumes of classes and refinements are shown along with global Gold-standard Fourier Shell Coefficient (GSFSC) resolutions. **e**,**f**, Map validation for the consOR2-G_s (**e**) globally refined, and (**f**) locally refined cryo-EM maps. Gold-standard Fourier shell correlation (FSC) curves are calculated in cryoSPARC, and shown together with directional FSC (dFSC) curves generated with dfsc.0.0.1.py as previously described[76]. Map-model correlations calculated in the Phenix suite are also shown. Arrows indicate map and map-model resolution estimates at 0.143 and 0.5 correlation respectively. Euler angle distributions calculated in cryoSPARC are also provided for each map. **g**,**h**, Local resolution estimation for the consOR2-G_s (**g**) globally refined, and (**h**) locally refined cryo-EM maps.

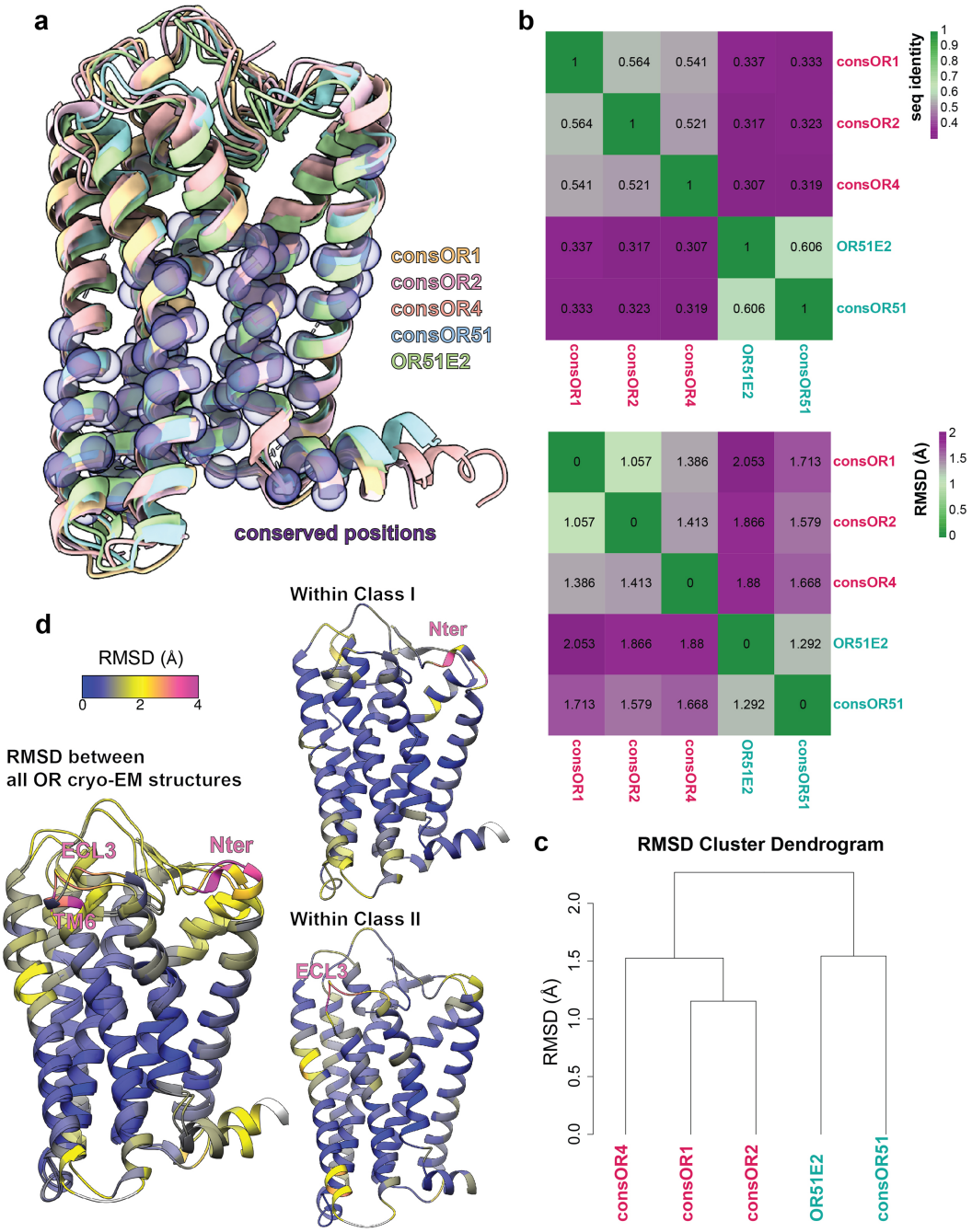


Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Cryo-EM data processing for consOR4-G_s. a, Size exclusion chromatogram and SDS-PAGE of consOR4-G_s sample. **b**, A representative cryo-EM micrograph from the curated consOR4-G_s dataset (n = 10,129) obtained from a Titan Krios microscope. **c**, A subset of highly populated, reference-free 2D-class averages are shown. Scale bar is 50 Å. **d**, Schematic showing the image processing workflow for consOR4-G_s. Initial processing was performed using UCSF MotionCor2 and cryoSPARC. Particles were then transferred using the pyem script package[58] to RELION for alignment-free 3D classification. Finally, particles were processed in cryoSPARC using the non-uniform and local refinement tools. Dashed boxes indicate selected classes, and 3D volumes of classes and refinements are shown along with global Gold-standard Fourier Shell Coefficient (GSFSC) resolutions. **e**,**f**, Map validation for the consOR4-G_s (**e**) globally refined, and (**f**) locally refined cryo-EM maps. Gold-standard Fourier shell correlation (FSC) curves are calculated in cryoSPARC, and shown together with directional FSC (dFSC) curves generated with dfsc.0.0.1.py as previously described[76]. Map-model correlations calculated in the Phenix suite are also shown. Arrows indicate map and map-model resolution estimates at 0.143 and 0.5 correlation respectively. Euler angle distributions calculated in cryoSPARC are also provided for each map. **g**,**h**, Local resolution estimation for the consOR4-G_s (**g**) globally refined, and (**h**) locally refined cryo-EM maps.



Extended Data Fig. 10 | Odorant binding to consOR2 and consOR4. a, The EM model of ligand in consOR2 with the ligand S-carvone contacting residues highlighted. **b**, The ligand positions during the consOR2 simulation. **c**, The EM model of ligand in consOR4 with the ligand 2-methyl thiazoline contacting residues highlighted. In both **a** and **b**, the ligand contacting residues are found by utilizing UCSF Chimera Find Clashes/Contacts module with default parameter. **d**, The ligand positions during the consOR4 simulation. In both **c** and **d**, the transparent ligands represent the simulation snapshots taken by skipping every 100 ns whereas the non-transparent ligand is the cryo-EM pose.



e

Receptor	Maximum sequence identity for native OR in family	Minimum sequence identity for native OR in family	Average sequence identity for all native ORs in family	Maximum sequence similarity for native OR in family	Minimum sequence similarity for native OR in family	Average sequence similarity for all native ORs in family
consOR1	0.71	0.51	0.65	0.82	0.67	0.76
consOR2	0.67	0.34	0.58	0.79	0.49	0.72
consOR4	0.67	0.56	0.62	0.79	0.71	0.75
consOR5	0.70	0.56	0.62	0.81	0.68	0.74
consOR6	0.70	0.47	0.59	0.81	0.64	0.72
consOR8	0.71	0.51	0.66	0.83	0.65	0.77
consOR10	0.67	0.48	0.58	0.76	0.62	0.71
consOR51	0.74	0.45	0.61	0.84	0.57	0.75
consOR52	0.74	0.49	0.62	0.83	0.63	0.75

Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Comparison of all human OR cryo-EM structures. **a**, Superimposition of all cryo-EM structures. The positions most conserved in terms of location are highlighted by purple dots. **b**, Heatmap of sequence identity and RMSD between the cryo-EM structures of consOR1, consOR2, consOR4 (red), consOR51 and OR51E2 (cyan). **c**, Cluster dendrogram made by RMSD. Class II and class I are clustered in two different groups. **d**, CryoEM structures coloured by RMSD for all the structurally elucidated ORs (left, projected on consOR1 and consOR51), only class I ORs (top right, consOR51 and OR51E2 projected on consOR51) and only class II (bottom right, consOR1, consOR2, consOR4 projected on consOR1). **e**, Sequence identity and similarity between consORs and native OR family members.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	consOR51-G_s complex	consOR1 L-menthol G_s complex	consOR2 S-carvone G_s complex	consOR4 2-MT G_s complex
EMDB: Full map	(EMD-42786)	(EMD-42789)	(EMD-42791)	(EMD-42817)
EMDB: Local map (7TM)	(EMD-45881)	(EMD-45876)	(EMD-45879)	(EMD-45880)
RCSB PDB: Model	(PDB 8UXV)	(PDB 8UXY)	(PDB 8UY0)	(PDB 8UYQ)
Data collection and processing				
Magnification	105,000	105,000	105,000	105,000
Voltage (kV)	300	300	300	300
Electron exposure (e–/Å^2)	50	50	50	50
Defocus range (μm)	-2.0 to -0.8	-2.1 to -1.0	-2.1 to -1.0	-2.1 to -1.0
Pixel size (Å)	0.873 (physical)	0.86 (physical)	0.86 (physical)	0.856 (physical)
Symmetry imposed	*C1*	*C1*	*C1*	*C1*
Initial particle images (no.)	6,275,495	15,310,482	5,894,860	10,846,651
Final particle images (no.)	204,513	212,883	131,718	318,570
Map resolution (Å) (masked)	3.2	3.3	3.2	3.5
FSC threshold	0.143	0.143	0.143	0.143
Refinement				
Initial model used (PDB code)	AlphaFold (consOR51) 7LJC (G protein) 3SN6 (Nb35)	AlphaFold (consOR1) 7LJC (G protein) 3SN6 (Nb35)	AlphaFold (consOR2) 7LJC (G protein) 3SN6 (Nb35)	AlphaFold (consOR4) 7LJC (G protein) 3SN6 (Nb35)
Model resolution (Å) (unmasked/masked)	3.4/3.2	3.9/4.0	3.6/3.4	3.8/3.6
FSC threshold	0.5	0.5	0.5	0.5
Map sharpening *B* factor (Å^2)	-122	-156	-133	-134
Model composition				
Non-hydrogen atoms	8074	8153	8121	8193
Protein residues	1041	1035	1031	1036
Ligands	0	1	1	1
B factors (Å^2)				
Protein	37.97	141.99	104.76	51.70
Ligand		129.98	109.86	20.00
R.m.s. deviations				
Bond lengths (Å)	0.003	0.003	0.004	0.013
Bond angles (°)	0.558	0.993	0.588	1.957
Validation				
MolProbity score	1.22	1.04	1.23	0.95
Clashscore	3.11	2.09	4.50	0.67
Poor rotamers (%)	0	0.22	0	0
CaBLAM outliers (%)	0.99	1.29	0.40	1.50
Ramachandran plot				
Favored (%)	97.37	97.75	98.12	97.25
Allowed (%)	2.63	2.25	1.88	2.75
Disallowed (%)	0	0	0	0

Table outlines key statistics for cryo-EM structure determination.

nature research

Corresponding author(s): Aashish Manglik

Last updated by author(s): 8/26/2024

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	SerialEM 3.8
Data analysis	SerialEM 3.8, UCSF MotionCor2, cryoSPARC v4.0.3, pyem v0.5, UCSF ChimeraX v1.25, Relion v4.0, GraphPrism 10, Flowjo v10.8.1, Schrödinger Maestro (version 2022-2), GROMACS package (version 2022), CHARMM36m force field, ParaChem 3.0, Maestro Schrödinger (version 2022-2), CHARMM-GUI, VMD (version 1.9.4), PyMOL (version 2.5), GROMACS modules (versions 2019-2022), R 4.3.1, Biostrings 3.19, seqinr 4.2-36, ape 5.8, ggtree 3.12.0, bio3D 2.4-4, pheatmap 1.0.12, gtools 4.8.0, WebLogo3

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Coordinates for consOR51, consOR1, consOR2, and consOR4 have been deposited in the RCSB PDB under accession codes 8UXV, 8UXY, 8UY0, and 8UYQ, respectively. EM density maps for consOR51, consOR1, consOR2, and consOR4 have been deposited in the Electron Microscopy Data Bank under accession codes EMD-42786, EMD-42789, EMD-42791, and EMD-42817 respectively. The locally refined maps for the 7TM domain for consOR51, consOR1, consOR2, and consOR4 have been deposited in the Electron Microscopy Data Bank under accession codes EMD-45881, EMD-45876, EMD-45879, and EMD-45880 respectively. The MD simulation trajectories have been deposited in the GPCRmd database under access codes 2056 (consOR1 with ligand), 2057 (consOR1 with ligand and miniGs), 2058

consOR2 Apo), 2059 (consOR2 with ligand and miniGs), 2060 (consOR4 Apo), 2061 (consOR4 with ligand and miniGs), 1245 (OR51E2 with ligand), 2062 (OR51E2 with ligand and miniGs)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For cryo-EM studies, the data size was limited by available instrument time and relative particle density on cryo-EM grids. For signaling studies, we used a sample size of 3-4 to enable repeatability and to control for biological variance typical in biochemical assays, with a minimum of two measurements per tested concentration. As is common in the field, no statistical assays were done to power the number of conditions for the signaling assays. For molecular dynamics simulations, five independent replicates were performed to control for variance in initial starting conditions. Similarly, as is common in the field, no statistical tests were used to determine number of simulation replicates.
Data exclusions	No datapoints were excluded from analysis.
Replication	Biochemical assays were replicated 2 or 3 times (as indicated), with identical results. The structural biology approaches represent ensemble averages, and the individual experiments were not repeated, as is common and accepted practice in the field. Data processing approaches were replicated and assessed with well-established approaches as outlined in the Methods section.
Randomization	Randomization was not relevant to the experiments in our study as the assays don't have unknown covariates. For example, when we compare wild-type to mutant ORs in signaling studies, there is no feasible unknown covariate that we can minimize by randomizing experimental units.
Blinding	Blinding was not relevant to the experiments in our study since no subjective allocation was involved.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	M1-FLAG (made with ATCC hybridoma HB-9259), mouse anti-rhodopsin clone 4D2 (MABN15, Sigma-Aldrich), phycoerythrin-conjugated donkey anti-mouse F(ab')2 fragment (715-116-150, Jackson Immunologicals)
Validation	M1-FLAG antibody was purified over its respective antigen peptide, which was also used to validate binding. Validation from the providers: mouse anti-rhodopsin clone 4D2: "Evaluated by Immunohistochemistry in rat eye tissue. Immunohistochemistry Analysis: 1:500 dilution of this antibody detected Rhodopsin in rat eye tissue." phycoerythrin-conjugated donkey anti-mouse F(ab')2 fragment: "Based on immunoelectrophoresis and/or ELISA, the antibody reacts with whole molecule mouse IgG. It also reacts with the light chains of other mouse immunoglobulins. No antibody was detected against non-immunoglobulin serum proteins. The antibody has been tested by ELISA and/or solid-phase adsorbed to ensure minimal cross-reaction with bovine, chicken, goat, guinea pig, Syrian hamster, horse, human, rabbit, and sheep serum proteins, but it may cross-react with

immunoglobulins from other species."

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Sf9 and Hi5 insect cells (Expression Systems), Expi293F cells (ThermoFisher), HEK293T cells (Douglas Alan Marchuk laboratory at Duke University)
Authentication	Cell line authentication of the HEK293T cells: Duke DNA Analysis Facility conducted DNA profiling of HEK293T lab stock for polymorphic short tandem repeat (STR) markers using a GenePrint 10 (Promega) and confirmed shared profiles with the reference (ATCC CRL3216)
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	None