Divergence from conserved residues underlies intracellular retention of mammalian odorant receptors

Kentaro Ikegami [a, b], Claire A. de March [a], Maira H. Nagai [a, c], Soumadwip Ghosh [e], Matthew Do [a], Ruchira Sharma [a], Elise S. Bruguera [a], Yueyang Eric Lu [a], Yosuke Fukutani [a, b], Nagarajan Vaidehi [e] , Masafumi Yohda [b], and Hiroaki Matsunami [a, d *]

[a] Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710.

[b] Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan

[c] Universidade de Sao Paulo, Sao Paulo, Brazil

[d] Department of Neurobiology, Duke Institute for Brain Sciences, Duke University Medical Center, Durham, NC 27710

[e] Department of Computational and Quantitative Medicine, Beckman Research Institute of the City of Hope, Duarte, CA 91006

* hiroaki.matsunami@duke.edu

Abstract

Mammalian odorant receptors are a diverse group of G-protein coupled receptors abundantly expressed in the olfactory cilia membrane. The vast majority of odorant receptors show little or no cell surface expression in non-olfactory cells due to endoplasmic reticulum retention, which has slowed down biochemical and functional studies. Here we show that engineered odorant receptors based on consensus amino acids exhibit high levels of cell surface expression that surpass any native odorant receptors. Using a combination of large-scale screening of odorant receptors for cell surface expression in heterologous cells, amino acid sequence analysis and machine learning techniques, we identified amino acid positions scattered throughout the receptor structure that are critical for odorant receptor trafficking. We demonstrated the importance of conserved residues by synthesizing "consensus" odorant receptors that show similar levels of cell surface expression with conventional G protein-coupled receptors. Furthermore, we associated *in silico* structural instability with poor cell surface expression using molecular dynamics simulations. Our results highlight the importance of conserved residues in complex regulation of odorant receptor trafficking.

Main

Mammalian olfactory receptors (ORs) are a diverse, rapidly evolving and form the largest family of G-protein coupled receptors (GPCRs) [1,2]. ORs are expressed on the cell surface of the olfactory sensory neurons (OSNs) to detect and discriminate the vast number of odors in the environment [3,4].

ORs are usually retained in the endoplasmic reticulum (ER) when expressed alone in non-olfactory cells, including neurons [5-11]. Receptor transporting protein (RTP) 1 and RTP2, which enhance the cell surface expression of ORs when co-transfected with ORs in heterologous cells [12-15]. In *RTP1* and *RTP2* double knockout mice (RTP DKO), the majority of ORs are significantly underrepresented (uORs) due to the absence of mature OSNs expressing them, suggesting that these ORs require RTP1 and RTP2 in order to function [16]. Interestingly, a small subset of ORs are overrepresented (oORs), suggesting that a minor subset of ORs function without RTP1 and RTP2. Accordingly, some oORs show cell surface expression when expressed in heterologous cells [16].

The underlying causes of OR retention in the ER in cells except for mature OSNs are not well understood. Previous structure-functional analysis using model ORs based on the assumption of OR-specific ER retention signals did not identify common residues that are involved in the cell surface expression of ORs [7,17,18].

In this study, we approach the mechanistic understanding of OR trafficking with the goal of engineering ORs with increased expression in heterologous cells comparable to that of other GPCRs, which will enable structure-function studies of ORs. Towards this goal, we have used inter-disciplinary strategies. First, we used a pair of closely related ORs that show differential cell surface expression in heterologous cells to identify specific amino acid residues that regulate cell surface expression. We performed molecular dynamics simulations on a set of ORs and mutants with differential cell surface expression to estimate protein stability and its possible relationship to expression. Second, we conducted a large-scale analysis of the cell surface expression of 210 ORs and used the dataset to statistically identify critical amino acid residues that predict cell surface expression. Third, we synthesized consensus ORs based on the most frequently used amino acid residues for a given site and directly tested the role of conserved residues in OR trafficking. Fourth, a stabilization strategies commonly used on GPCRs and other proteins[19-22] were applied to the OR family. We improved the stability of the most promising consensus ORs by inserting salt bridges in their structure and obtained mutated consensus ORs that show surface expression levels comparable to a non-olfactory class A GPCR. Together, our data suggest that divergence from conserved residues results in the retention of ORs inside the cells, which may be caused by structural instability.

RESULTS

A TM4 residue, $G^{4.53}$, is crucial for cell surface trafficking of model ORs

We chose Olfr539 and Olfr541 with 90% amino acid identity as a model system to study OR trafficking. Olfr539 is an oOR, an indication of high representation in the absence of RTP1 and RTP2 *in vivo*, and shows robust cell surface expression in human embryonic kidney (HEK) 293T cells. In contrast, Olfr541 is an uOR that requires RTP1 and RTP2 for its cell surface expression *in vivo*, and shows no detectable cell surface expression in HEK293T cells (Fig. 1A, 1B). We generated a series of chimeric ORs by replacing a part of the amino acid sequence of Olfr539 with that of Olfr541 (Fig. 1C). Cell surface expression levels of the chimeric ORs were measured by flow cytometry. Mutant ORs which had the middle region (from 152th residue to 247th residue) of Olfr539 showed high surface expression levels, indicating an important domain for cell surface expression.

We next generated single amino acid mutants of each middle region residue of Olfr539 by substituting each amino acid of Olfr539 with that of Olfr541 (Fig. 1D). Olfr539 $G161C^{4.53}$ (4.53 refers to the Ballesteros-Weinstein residue numbering system used for GPCRs[23], see Fig. 1B) and $V216G^{5.47}$ showed abolished or diminished Olfr539 cell surface expression. Accordingly, we expressed the reciprocal single mutants Olfr541 $C154G^{4.53}$ and Olfr541 $G209V^{5.47}$, and the double mutant Olfr541 $C154G^{4.53}/G209V^{5.47}$ to test if these residues are sufficient for promoting cell surface expression. Olfr541 $C154G^{4.53}$ showed a moderate level of cell surface expression although Olfr541 $G209V^{5.47}$ did not show any improvements (Fig. 1E). Olfr541 $C154G^{4.53}/G209V^{5.47}$ showed a more enhanced cell surface expression than $C154G^{4.53}$ alone, suggesting a synergistic interaction between the two critical residues in facilitating cell surface

trafficking. This indicated that these two residues are critical in determining cell surface expression levels. Other residues had little or no effect on cell surface expression.

We further investigated which properties of $G^{4.53}$ were tied with controlling the cell surface trafficking, generating mutants by substituting $G161^{4.53}$ of Olfr539 for less frequently used amino acids (S, T, A or N), which were altogether present in almost 30% of mouse ORs (Fig. 1F). None of these mutants exhibited cell surface expression, indicating a requirement of $G^{4.53}$ for the cell surface trafficking of Olfr539. We next tested whether $G^{4.53}$ controlled the trafficking of other ORs that are trafficked to the cell surface. We chose five ORs (Olfr1362, Olfr1508, Olfr449, Olfr168 and Olfr1239) that show cell surface expression in HEK293T cells (RTP-independent ORs, see below for details). All the $G^{4.53}$ mutants showed abolished or diminished cell surface expression (Fig. 1G) supporting the importance of this residue in facilitating cell surface trafficking of ORs.

Contribution of $G^{4.53}$ in the *in silico* structural stability of Olfr539 and Olfr541

In order to understand how the two critical residues, $G^{4.53}$ and $V^{5.47}$, control the cell surface trafficking of Olfr539 and Olfr541, we localized the two residues in a 3D structural models of Olfr539 and Olfr541. OR models were constructed using homology modeling method based on known GPCR crystal structures in the inactive state [24,24,25]. $G^{4.53}$ and $V^{5.47}$ are in the middle of the 4th and the 5th transmembrane domains (TM), respectively (Fig. 2A). We hypothesized that diminished structural stability of ORs might cause ER retention due to general quality control mechanisms and that $G^{4.53}$ and $V^{5.47}$ are important for the stabilization of ORs. To start testing this possibility, we built Olfr539, Olfr541, Olfr539 $G161C^{4.53}$, Olfr539 $V216G^{5.47}$, Olfr541

C154G$^{4.53}$, and Olfr541 C154G$^{4.53}$/G209V$^{5.47}$ homology models. We also included a control mutant Olfr539 L162A$^{4.54}$ which shows a similar cell surface expression to Olfr539. We ran six independent molecular dynamics (MD) simulations of 500ns for all the systems embedded in an explicit lipid bilayer. All the six simulations of Olfr539 showed well-packed structures while Olfr539 G161C$^{4.53}$ showed flexibility (Fig. 2B and C) as quantified by the root mean square deviations (RMSD) of atomic positions of TM domains for each MD simulation. Olfr539 and L162A$^{4.54}$ showed similar RMSD across multiple simulations, all converging though equivalent equilibrated structures. However, Olfr539G161C$^{4.53}$ and V216G$^{5.47}$, which are poorly expressed, showed variability in structures. This suggests that poorly trafficked ORs have more flexible structures and hence unstable when inserted into a cell membrane (Fig. 2C, S1). In contrast, Olfr541 showed a wide range of RMSD, whereas its mutants C154G$^{4.53}$ and C154G/G209V showed less variation in RMSD (Fig. 2C). To further examine whether structural stability is associated with cell surface expression level, we plotted the cell surface expression levels and variances of RMSDs for both Olfr539 and Olfr541 systems. Indeed, the mutants with lower cell surface expression showed a larger variance of RMSDs. Conversely, the mutants with higher cell surface expression showed a smaller variance of RMSDs (Fig. 2D). This trend is more pronounced in RMSDs calculated for the extracellular (EC) side of the TM domains, suggesting that stabilities of this part may be especially important for cell surface expression (Fig. S2). These data support the hypothesis that structural stability contributes to the cell surface expression of ORs.

A comprehensive evaluation of cell surface expression levels of ORs in heterologous cells

So far, we showed that the TM4 residue (4.53) plays a crucial role in OR trafficking. However, it is unlikely that this position is the sole determinant, as $G^{4.53}$ is present in 66% of mouse ORs, yet the vast majority of ORs do not show cell surface expression in heterologous cells. To gain a more comprehensive understanding of OR cell surface expression, we selected 210 ORs (76 oORs and 134 uORs) and tested which ORs exhibit cell surface expression in heterologous cells. Consistent with previous findings [16], oORs, as a group, showed more robust OR surface expression than uORs ($p<0.05$, U test) (Fig. 3A). Using the approximation that surface expression levels is the overlapping of two normal distributions of positive ORs and negative ORs, we defined the ORs with expression levels of more than 0.144 (A.U.) as positive ORs (cut off = top 0.001 of negative ORs). 34/210 ORs (14.0%) were positive under this criterion. Consistent with the more robust cell surface expression of oORs, 26/76 (34.2%) of oORs, and only 8/134 (6.0%) of uORs, showed positive cell surface expression. We defined the 26 ORs that are oORs and cell surface expression positive, as RTP-independent ORs and the 126 ORs that are uORs and cell surface expression negative, as RTP-dependent ORs.

Critical residues predict cell surface expression of ORs

We hypothesized that residues specific to RTP-independent or RTP-dependent ORs are critical for OR trafficking in the absence of RTP1 and RTP2. To start investigating whether overall amino acid sequence similarities are associated with cell surface trafficking, we asked whether RTP-independent and/or RTP-dependent ORs are clustered in a phylogenetic tree. Both are distributed on multiple branches, indicating that overall similarities do not determine their ability to be trafficked to the cell surface (Fig. 3B, S3).

Next, we hypothesized that amino acid residues at particular positions create a network that control cell surface trafficking. To investigate this idea, we aligned 26 RTP independent and 126 RTP dependent ORs (total of 152 ORs) and calculated Grantham distances of amino acid properties at individual sites. We identified 66 positions with lower Grantham distances between amino acids for RTP-independent ORs than for all the 152 tested ORs (cut off $p<0.05$, t test with Bonferroni correction). As expected, (4.53) is one of these 66 sites; 80.8% of RTP-independent ORs possess a G residue at this position against only 61.1% in the RTP-dependent ORs. Contrary to the initial assumption that specific domains control OR cell surface expression, the 66 sites are scattered throughout the OR sequence. Moreover, there was no site that is exclusively present in one of the groups, suggesting that there are no trafficking promotion or inhibition signals that are shared among all ORs (Fig. 3C). To investigate whether these 66 sites can predict the RTP dependence of ORs, we classified tested ORs by support vector machine based classifiers in ten-fold cross-validation. The support vector machine model generated by the 66 sites of amino acid residues discriminated RTP-independent ORs (1.70×10^{-92}, Wilcoxon test; AUC = 0.893). However, those generated by 66 randomly selected sites ($p=0.999$, Wilcoxon test; AUC = 0.425) and those generated by all sites ($p=0.999$, Wilcoxon test; AUC = 0.414) failed to discriminate RTP-independent ORs. This demonstrates that these 66 sites robustly predict whether an OR shows cell surface expression in heterologous cells (Fig. 3D).

What properties of these 66 residues are associated with cell surface expression? When we looked at the degree of conservation of these residues, many of these sites were conserved among ORs (Fig. 3E, S4). 29% of the sites (19 / 66) are conserved in more than 90% of the mouse ORs whereas only 13% of all amino acid sites (41/307) are conserved in more than 90% of full length OR ($p=0.0048$, Fisher's exact test). RTP-independent ORs have the most common amino acid

residues much more frequently present than RTP-dependent ORs (58 out of the 66 sites, $p=6.35x10^{-6}$, chi-square test), suggesting that ORs that are in line with consensus amino acids in these positions are more likely to show cell surface expression.

Engineered consensus ORs robustly express on the cell surface in heterologous cells

The above results suggest the importance of the most frequently-used amino acid residues at a given site in cell surface expression. This observation led us to predict that ORs that are designed based on "consensus" amino acids for each site would be efficiently trafficked to the cell surface. The consensus strategy has already been applied to proteins or codons to improve their thermostability or function in other proteins [19,20,26,27]. The success of this strategy relies on the number of proteins available to build the consensus sequence and their sequence similarity.[19,20] Here, we used the unique diversity of the OR family among GPCRs to apply the consensus strategy, aiming to obtain stable ORs. We aligned amino acid sequences of human OR families and determined the consensus sequences as the mostly frequently occurring amino acid residue at each position (Table S1). We first chose the OR10 family to measure the cell surface expression levels of the consensus OR in HEK293T cells in comparison to each member of OR10 family. Strikingly, the consensus OR10 robustly expressed on the cell surface, more than any of the individual OR10 family members tested (Fig. 4A). We generated consensus human ORs for 8 other families (OR1, OR2, OR4, OR5, OR6, OR8, OR51 and OR52) and measured their cell surface expression levels. 6 out of 9 consensus ORs show more robust cell surface expression than any tested natural ORs (Fig. 4B).

To exclude the possibility that this effect is specific only to HEK293T cells, we expressed the consensus ORs in NIH3T3 cells, which are derived from mouse fibroblasts. Most of the consensus ORs again showed robust cell surface expression as determined by flowcytometry analysis (Fig. 4B). We evaluated the level of conservation of the 66 amino acid sites that we showed to be important for RTP independent OR expression (Fig. 3E) between our consensus ORs and the mouse OR repertoire. Consensus ORs have the most common amino acid residues much more frequently represented than natural ORs (59 out of the 66 sites, $p=4.70x10^{-94}$, chi-square test) (Fig. 4C). We built phylogenetic trees using parsimony criterion for each OR family including the corresponding consensus OR and found that the consensus OR is always located at the origin of the tree (Fig. 4D, S5).

Finally, we verified that the consensus receptors are functional in their response to odorants. We used cAMP-mediated luciferase reporter gene assay [13] to screen active ligands from a set of 320 diverse odorants at 150μM. We identified robust ligands for OR1, OR6, OR10 and OR52, each of which shows responses to specific subsets of the tested odorants (Fig 4E and F). Our data shows that these consensus ORs are indeed functional, which suggests that consensus ORs are properly folded.

Stabilization of consensus ORs structure by introduction of salt bridges

The consensus ORs already show a clear increase of cell surface expression in comparison to naturally-occurring ORs. Since we observed that the expression of an OR seems to be correlated with its stability and rigidity when inserted in a membrane using molecular dynamics simulations (Fig. 2B, C and D) , we attempted to further improve the expression level of the consensus ORs by engineering salt bridges in the structures guided by 3D homology models (Fig. 5A).

Thermostabilization studies on β1-adrenergic receptor[28] and chemokine receptor[21] showed that tightening of interactions between intracellular loop1 (ICL1) and the helix8 improved the GPCR stabilization. Another study[29] showed that enriching the number of basic residues in the intracellular loops of an OR enhances its expression. Based on these premises, we inserted triple Arg mutations in intracellular loop 1 (ICL1) to promote salt bridge interactions between ICL1 and helix8 in four human consensus ORs, namely OR1, OR10, OR51 and OR52 (Fig. 5B). We evaluated the cell surface expression of the consensus ORs with their corresponding triple Arg mutants, including a well-studied non-olfactory class A GPCR, the muscarinic 3 receptor (M_3) (Fig. 5C). We observed enhanced expression of OR10 and OR52 mutants which are comparable to the expression level of the M_3 receptor even when we changed the amount of DNA used in transfection by 100-fold. However, we did not observe any enhancement in expression level of mutant OR1 and OR51. The insertion of such stabilizing interactions improves the rigidity of the OR structure that might aid expression[24,30]. We tested their functionality in comparison to their corresponding OR consensus (Fig. 5D). Mut-OR10 and Mut-OR52 both responded to their agonists (β-ionone and nonanoic acid respectively). Mut-OR52 showed activity similar to the consensus OR52, but the response of Mut-OR10 was diminished compared to OR10. For both OR10/Mut-hOR10 and OR52/Mut-OR52, decreasing the amount of transfected DNA by 10-100 fold compared with optimized amount for natural ORs resulted in robust responses against tested odorants, showing the capacity of these consensus ORs in supporting high levels of functional expression.

Discussion

Mammalian ORs are notoriously difficult to work *in vitro* due to poor expression. In this study, we investigated the underlying mechanisms by which OR trafficking is regulated in heterologous cells. Using chimeras and point mutations with a pair of highly similar ORs, we identified critical residues that regulate cell surface expression. MD simulations suggest that these residues may affect the flexibility and stability of ORs. We also conducted a large-scale cell-based screening to comprehensively identify ORs that express on the cell surface, which lead to the identification of specific residues associated with their cell surface expression. As cell surface expression positive ORs tend to have conserved residues at the critical sites, we investigated whether consensus ORs express on the cell surface in heterologous cells and demonstrated that most of the consensus ORs show robust cell surface expression.

Which residues or domains make OR trafficking difficult? Previous studies suggest different residues, domains, or features of ORs are responsible for intracellular retention [7,18]. Among them, a study proposed that fewer charged residues and more hydrophobic residues distributed throughout ORs might underlie intracellular retention [17]. We tested whether RTP-independent ORs possess fewer charged and more hydrophobic residues compared to RTP-dependent ORs, as suggested by Bubnell et al[17]. However, there were no significant differences between tested ORs ($p=0.74$ and $p=0.71$, respectively), suggesting that these features do not explain the cell surface trafficking of ORs as a group. Our current study identified two sites ($G^{4.53}$ and $V^{5.47}$) contributing to cell surface expression in model ORs. Since these residues were not identified in previous studies, it seems that different residues regulate the trafficking of different ORs. This raised a question whether there are any common mechanisms that regulate OR trafficking. Our statistical

analysis based on a large-scale cell surface expression analysis of hundreds of ORs identified 66 residues, scattered throughout the receptor play a critical role on in OR cell surface expression. Though there was no single residue or domain that solely determined cell surface expression, we succeeded in building a machine learning model that reliably predicts OR cell surface expression based on these residues. This suggests that these 66 residues differentially contribute to trafficking efficiency of individual ORs.

What features of ORs cause their ER retention when expressed in non-olfactory cells? It was long hypothesized that ORs may possess specific ER retention signals [6,7,12,17]. However, a previous report shows that mutating the most highly conserved OR-specific amino acids in the one of the trafficking non-olfactory GPCR superfamily did not enhance the cell-surface expression of a model OR [17]. Our data is also inconsistent with the idea that common OR-specific residues or domains direct ER retention, supporting the absence of OR-specific ER retention signals. In stark contrast, our study demonstrated that consensus ORs show robust cell surface expression, suggesting that such ER retention signals, if any, would not be shared by OR members. How do ORs (aside from a minor subset) possess the common feature of being retained inside the cells? Our study supports the model that the intracellular retention of ORs is caused by the structural instability of ORs, which is caused by divergence from conserved residues at the critical sites. In other words, the majority of ORs do not fold correctly in heterologous cells due to a divergence from conserved residues, thus they are trapped by the general protein quality control in the ER and cannot be trafficked to the plasma membrane. This model explains why many consensus ORs exhibit robust cell surface expression, while the vast majority of natural ORs do not show detectable cell surface expression in heterologous cells [31].

Why are ORs functional in the OSNs but not in heterologous cells when expressed alone? Our study supports the idea that accessory proteins or chaperones expressed in the OSNs assist the folding of structurally unstable ORs. In addition to RTP1 and RTP2, a previous study showed that an Hsp70 homolog enhanced the cell surface expression of an OR, suggesting the importance of chaperones for the cell surface expression of ORs [32]. A possibility is that unstable nature of ORs may be integral to proper OSN development [33] [16]. Initiation of OR expression induces unfolded protein response (UPR) in developing OSNs, suggesting inefficient folding and ER accumulation of ORs in these cells [33,34]. RTP1 and RTP2 are induced by UPR signaling in developing OSNs, presumably allowing OR proteins to exit the ER and downregulate UPR [33,34] [16].

Despite some successes [8,9], biochemical and structural studies of ORs have been challenging using the currently available expression systems, due to the relatively poor expression of ORs when compared to canonical GPCRs. Our study shows the success of the application of protein engineering strategies to enhance OR stability and cell surface expression. We have identified a set of engineered ORs that show robust cell surface expression comparable to M_3 muscarinic acetylcholine receptor, a canonical GPCR. The consensus ORs will serve as candidate ORs for future biochemical studies such as the large-scale production of ORs and purification. In addition, the residues regulating trafficking identified in this study will point to further improvements in protein production, which will accelerate our effort towards determining OR structures. Altogether, our study provides novel insights into how various residues in OR regulate the receptor expression on the cell membrane.

EXPERIMENTAL PROCEDURES

DNA and vector preparation.

Open reading frames of OR genes were subcloned into pCI (Promega) with a Rho tag at the N terminal. DNA fragments of OR genes were amplified by Phusion polymerase (Thermo Fisher Scienctific). To generate chimeras and mutants of ORs, DNA fragments of OR genes were amplified by Phusion polymerase (Thermo Fisher Scienctific). The fragments were mixed and amplified by PCR reaction to obtain full sequences. All plasmid sequences were verified using Sanger sequencing (3100 Genetic Analyzer, Applied Biosystems).

Cell culture.

HEK293T and Hana 3A cells[12] were grown in Minimal Essential Medium (MEM) containing 10% FBS (vol/vol) with penicillin-streptomycin and amphotericin B. Hana 3A cells were authenticated using polymorphic short tandem repeat (STR) at the Duke DNA Analysis Facility using GenePrint 10 (Promega) and shown to share profiles with the reference (ATCC). NIH/3T3 cells were grown in high glucose Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% CS, penicillin-streptomycin and amphotericin B. All cell lines were incubated at 37°C, saturating humidity and 5% CO2. No mycoplasma infection was detected in all cell cultures.

Flowcytometry analyses.

HEK293T cells were grown to confluency, resuspended and seeded onto 35 mm plates at 25% confluency. The cells were cultured overnight. A Rho tagged OR in the plasmid pCI and GFP expression vector were transfected using Lipofectamine 2000. After 18-24 hours, the cells were resuspended by cell stripper and then kept in 5 mL round bottom polystyrene (PS) tubes (Falcon

2052) on ice. The cells were spun down at 4°C and resuspended in PBS containing 15 mM NaN3, and 2% FBS to wash the cell stripper. They were incubated in primary antibody (mouse anti Rho4D2 [35]) and then washed, stained with phycoerythrin (PE)-conjugated donkey anti-mouse antibody (Jackson Immunologicals) in the dark. To stain dead cells, 7-Amino-actinomycin D (Calbiochem) was added. The cells were analyzed using BD FACSCanto II FACS with gating allowing for GFP positive, single, spherical, viable cells, and the measured PE fluorescence intensities were analyzed and visualized using Flowjo [36]. We normalized the surface expression levels by cells expressing Olfr539, which was robustly expressed on the cell surface, and cells expressing Olfr541, which showed no detectable cell surface expression.

Homology model building

The protocol follows a previously published method.[37] Aligned protein sequences of mouse 1092 ORs are manually aligned to pre-aligned protein sequences of 11 GPCRs including bovine rhodopsin (PDB: 1U19), human chemokine receptors CXCR4 (PDB: 3ODU) and CXCR1 (PDB: 2LNL), and human adenosine a2A receptor (PDB: 2YDV) using Jalview.[24]

Four experimental GPCR structures (1U19, 3ODU, 2YDV and 2LNL) are used as templates to build Olfr539 and its mutants (G154C, V209G, and L155A) and Olfr541 and its mutants (C154G and C154G/G209V) by homology modeling with Modeller. Five models are obtained and the one fulfilling several constraints (binding cavity sufficiently large, no large folded structure in extra-cellular loops, all TMs folded as α-helices, a small α-helix structure between TM3 and TM4) is kept for further molecular dynamics simulations.

Molecular dynamics simulations

The models were embedded in a model membrane made up of POPC lipids solvated by TIP3P water molecules using Maestro. The total system is made up of ~48,650 atoms in a periodic box of 91*89*98 $Å^3$.

Molecular dynamics simulations are performed with sander and pmemd.cuda modules of AMBER12 with the ff03 force-field for the protein and the gaff.lipid for the membrane. Hydrogen atoms are constrained by SHAKE algorithm and long-range electrostatics interactions are handled with Particle Mesh Ewald (PME). The cutoff for non-bonded interactions is set at 8 Å. Temperature and pressure are maintained constant with a Langevin thermostat with a collision frequency of 2 ps^{-1}. In addition, a weak coupling anisotropic algorithm with a relaxation time of 1 ps^{-1} is applied. Snapshots are saved every 20 ps.

Two energy minimizations are performed during 10,000 steps with the 5,000 first steps using a conjugate gradient algorithm. The first one is run with a restraint of 200 $kcal.mol^{-1}$ applied on all atoms of the membrane and water and the second one with the same restraint on all atoms of the receptor. This last constraint is kept for the heating phase of 20 ps (NTP, 100K to 310K, Langevin thermostat with collision frequency of 5 ps^{-1}) and equilibration of 15 ns (NTP, 310K). Restraints are then reduced by 5 $kcal.mol^{-1}Å^{-2}$ and another cycle of minimization-equilibration is performed. The systems (Olfr539 models (wt, G154C, V209G and L155A) and Olfr541 models (wt, C154G and C154G/G209V)) are replicated six times and 525 ns-long production molecular dynamics are performed after an equilibration period of 50 ns. RMSDs of seven transmembrane domains were calculated using CPPTRAJ in AmberTools. The RMSDs are between initial positions and each frame in the production step. 3D structures were visualized using VMD.

OR protein sequence analyses.

Protein sequences of 1092 mouse ORs were aligned by Clustal Omega with default parameters. Conservation degree of amino acid residues was visualized by WebLogo.[38]

To identify the amino acid residues involved in RTP dependence, Grantham distances were calculated for all pairs of ORs at each position. The positions where Grantham distances of RTP-independent ORs were significantly shorter than those of all ORs were searched by one-sided t-tests followed by Bonferroni correction.

Designing consensus ORs

Protein sequences of human ORs were downloaded from The Human Olfactory Data Explorer (HORDE) webpage (https://genome.weizmann.ac.il/horde/). The protein sequences were aligned using MAFFT. The most frequently used amino acid residues were defined as consensus residues at each position. The consensus amino acid sequences were translated into DNA sequences using Codon Optimization Tool on Integrated DNA Technologies (IDT) webpage.

Luciferase assay in Hana3A cells

The Dual-Glo Luciferase Assay (Promega) was used to determine the activities of firefly and Renilla luciferase in Hana3A cells as previously described[37]. Briefly, firefly luciferase, driven by a cAMP response element promoter (CRE-Luc; Stratagene), was used to determine OR activation levels. For each well of a 96-well plate, 5 ng SV40-RL, 10 ng CRE-Luc, 5 ng mouse RTP1s, 2.5 ng M3 receptor3, and 5 ng of Rho-tagged receptor plasmid DNA were transfected. Normalized activity for each well was further calculated as (Luc-400)/(Rluc-400) where Luc = luminescence of firefly luciferase and Rluc = Renilla luminescence. The basal activity of an OR was averaged

from six wells in the absence of odorants and further corrected by subtracting that of the control empty vector. An odorant-induced activity was averaged from at least three wells and further corrected by subtracting the basal activity of that receptor. Odorant-induced responses were normalized to that of wt.

Acknowledgements

We thank Mengjue Jessica Ni for expert technical assistance and Conan Juan, Sahar Kaleem and Yen Dinh for manuscript editing. This work was supported by grants from NIH (HM: DC014423 and DC016224, NV: NIH R01-GM097261). KI stayed at Duke University with financial support from Tokyo University of Agriculture and Technology as a student of the program for leading graduate schools in Japan. YF stayed at Duke University with financial support from JSPS Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers.

Author Contribution

HM conceived and designed the project. KI, RS, ESB, YEL and HM performed research concerning the large-scale expressions screening of OR and Olfr539 and Olfr541 mutants. KI, MD, and CADM performed the molecular modeling of Olfr539 and Olfr541 plus their mutants. MHN, YF, CADM and HM performed research on the consensus receptors. NV, SG, CADM and HM performed the salt bridge insertion. KI, CADM, MY and HM carried out the analysis and wrote the paper with inputs from all authors. HM supervised the project.

Competing Interests statement

H.M., K.I., M.H.N. and C.A.D.M. filed a provisional patent application relevant to this work. The remaining authors declare no competing interests.

Figures

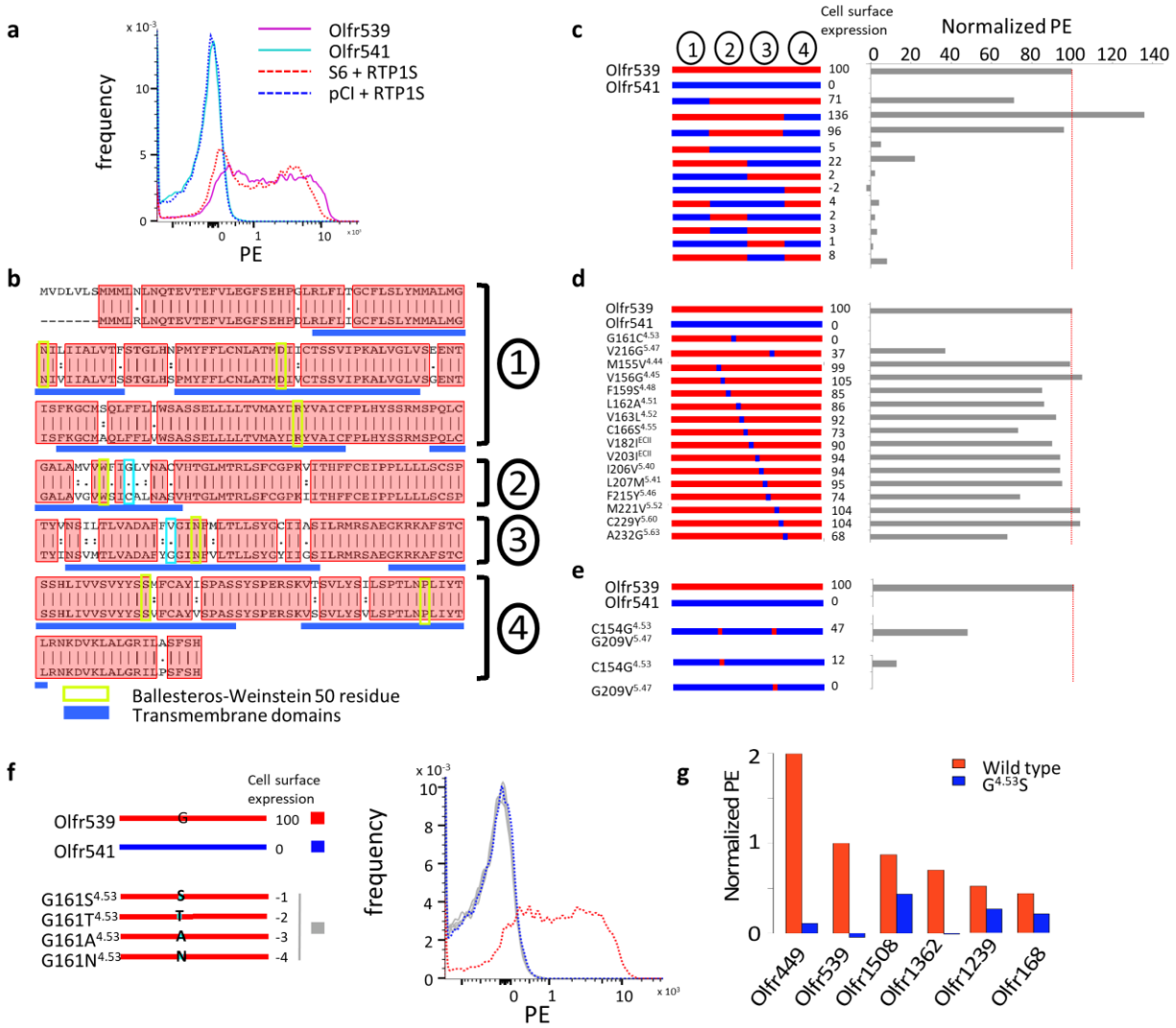

Figure 1.

G161$^{4.53}$ and V216$^{5.47}$ are critical to the cell surface expression of Olfr539 and Olfr541. **a**, Olfr539 robustly but Olfr541 poorly expresses in the cell surface in heterologous cells in the absence of RTP1 and RTP2. Expression is evaluated by flowcytometry with OR S6 and pCI as a positive and negative controls, respectively, by recording the frequency of PE fluorescence (see Methods section). **b**, Alignment of protein sequences of Olfr539 (top) and Olfr541 (bottom). 90% of the amino acid residues are shared between the two ORs and are shown in red boxes.

Ballosteros-Weinstein 50 residue for each transmembrane domain is boxed in green [23]. G/C161$^{4.53}$ and G/V216$^{5.47}$ are highlighted in cyan boxes. . **c, d, e, f**, Designs of chimeric ORs are shown on the left and cell surface expression results (Normalized PE) on the right. **c**, Chimeric ORs created by replacing parts of Olfr539 (red) with those of Olfr541 (blue) **d**, Single amino acid mutants created by substituting amino acids of Olfr539 in region 2 and 3 for those of Olfr541. **e**, Reciprocal Olfr541 mutants created by substituting single or double amino acids of Olfr541 with those of Olfr539. **f**, Olfr539 mutants created by substituting G161$^{4.53}$ with S, T, A or N, which are altogether conserved in almost 30% of the mouse ORs, lose cell surface expression. **g**, G$^{4.53}$S mutants of RTP-independent ORs (Olfr449, Olfr539, Olfr1508, Olfr1362, Olfr1239 and Olfr168) (blue) show less cell surface expression levels compared with the wild types (red) in the absence of RTP1 and RTP2.

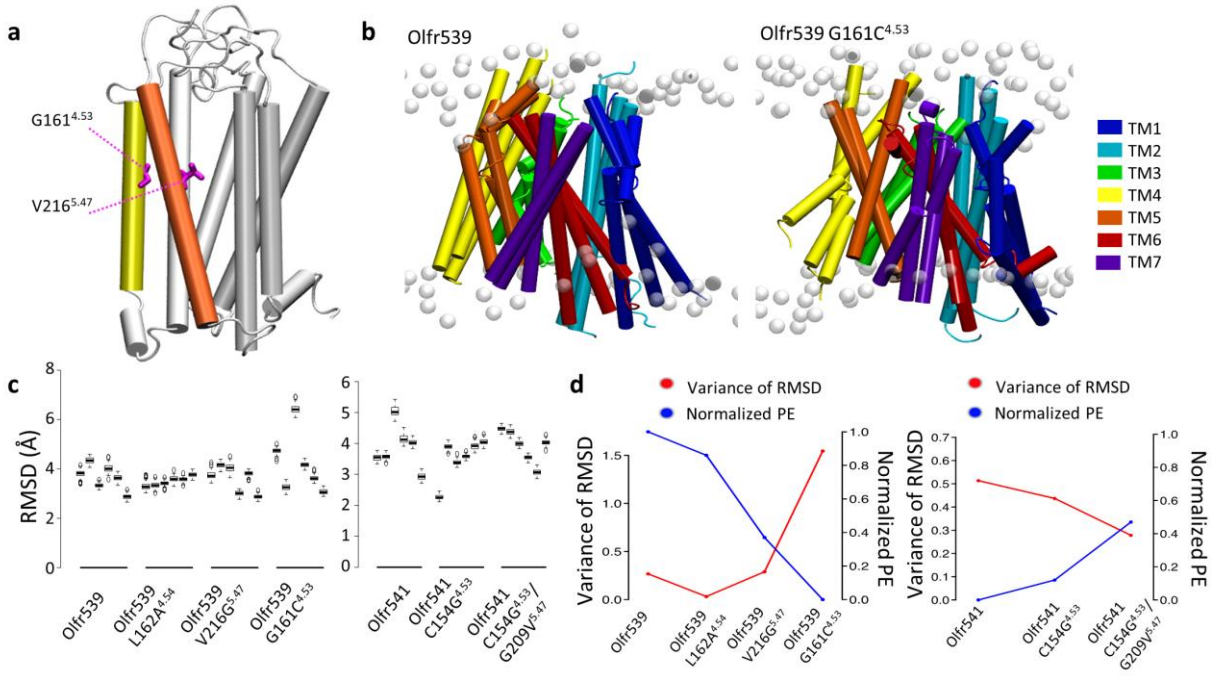

Figure 2.

a, 3D homology model of Olfr539. TM4 (yellow) and TM5 (orange) are represented in colored tubes and the remaining structure is in white. G154$^{4.53}$ and V209$^{5.47}$ are developed in pink licorice. **b**, Superposed images of triplicate wild type Olfr539 (left) and Olfr539 G161C$^{4.53}$ (right) models after 500ns molecular dynamics (MD) simulations in an explicit model of plasma membrane (white). **c**, RMSDs of 6 individual MD simulations of Olfr539 systems (wild type, L162A$^{4.54}$ G161C$^{4.53}$ and V216G$^{5.47}$) (left), and Olfr541 systems (wild type, C154G$^{4.53}$ and C154G$^{4.53}$/G209V$^{5.47}$). The models are placed in descending order based on cell surface expression levels for Olfr539 systems and in ascending order for Olfr541 systems. **d**, Plots of variance of mean RMSDs (left axis, red plots) and cell surface expression levels (right axis, blue plots) of Olfr539 systems (left) and Olfr541 systems (right).

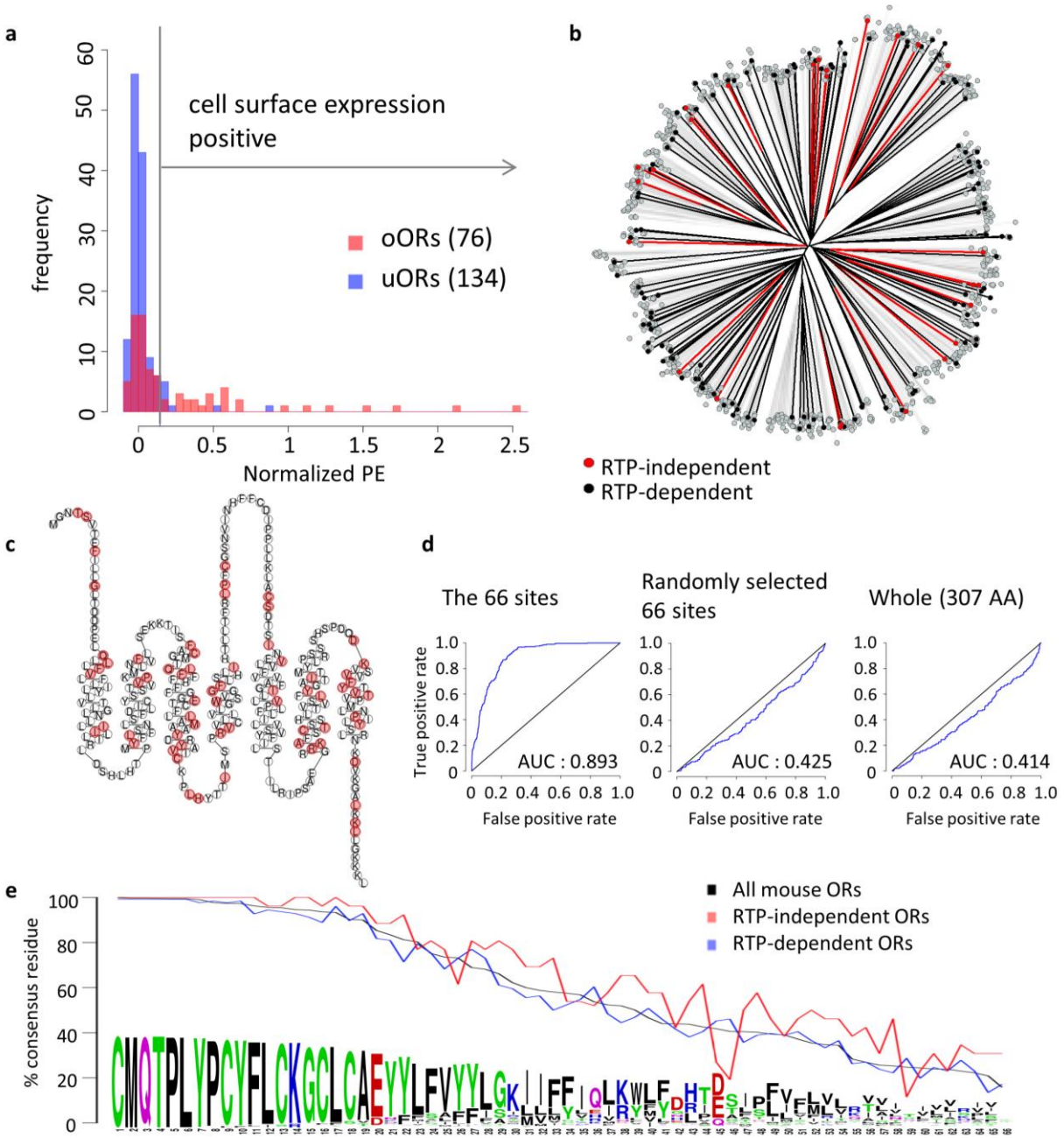

Figure 3.

Cell surface expression analyses for a large repertoire of mouse ORs in heterologous cells in the absence of RTP1 and RTP2. **a**, Histogram of cell surface expression levels of 76 oORs (red) and 134 uORs (blue). 0.144 (vertical line) is a cut-off for positive/negative cell surface expression. **b**, Phylogenetic tree of protein sequences of mouse ORs. Red, black and gray indicate RTP-

independent ORs, RTP-dependent ORs and the others ORs respectively. **c**, Snake plot of the consensus protein sequence of mouse ORs. The 66 sites with less diverse residues in RTP-independent ORs ($p<0.05$, Bonferroni corrected) are colored in red. **d**, RTP dependence of tested ORs is predicted using SVM-based classifiers based on amino acid properties in ten-fold cross-validation, and the accuracy is validated by the ROC curves. The SVM-based classifier has been built using the 66 sites of interest, 66 random amino acid positions and the entire 307 amino acid positions in the tested ORs. **e**, Usage rate of consensus residues at the 66 sites. RTP-independent ORs more frequently use consensus residues than RTP-dependent ORs at 58 out of 66 sites.

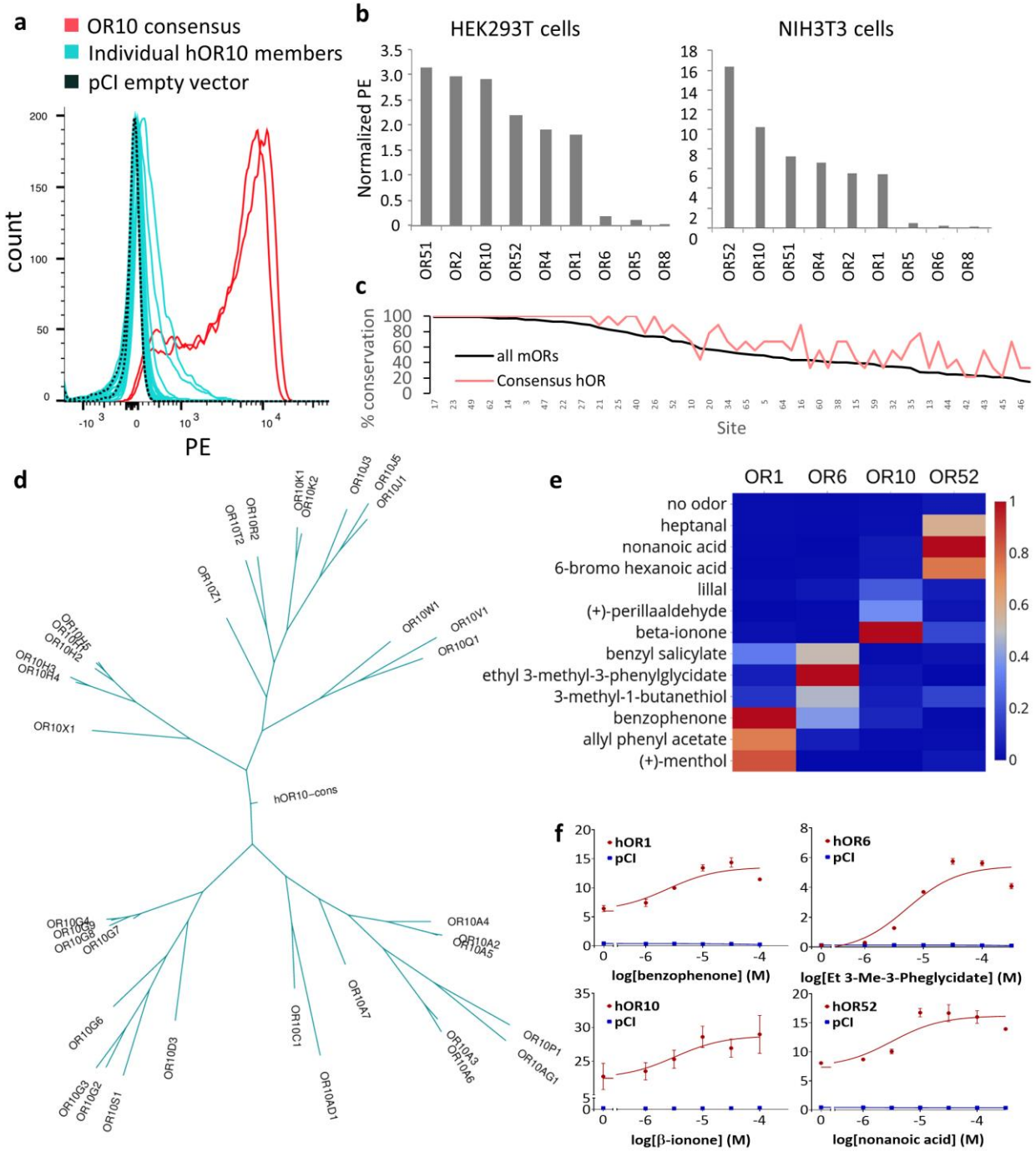

Figure 4.

Potential of consensus ORs for robust cell surface expression in the absence of RTP1S. **a**, 32 hOR10 subfamily members (cyan) and the consensus OR (red, duplicate) were transfected into

HEK293T cells and their cell surface expression levels were measured by flowcytometry. **b**, Cell surface expression levels of 9 consensus ORs (OR1, OR2, OR4, OR5, OR6, OR8, OR10, OR51 and OR52) were evaluated in HEK293T or NIH/3T3 cells. The PE fluorescence is normalized by setting Olfr539 response to 1 and Olfr541 to 0. **c**, Usage rate of consensus residues at the 66 sites for consensus ORs (red) and all mouse ORs (black). **d**, Ancestral tree of OR10 family member including OR10-consensus. **e**, Heatmap of ORs responses to 30 μM of odorants selected from a previously screened panel of 320 compounds. Luciferase activity was normalized for each OR by setting as 1.0 the highest response value, and 0 as the lowest response value. Consensus OR1, OR6, OR10 and OR52 are functionally expressed in Hana3A cells. **f**, Dose-response curve of consensus ORs to candidate ligands ranging from 0 to 316 μM or 1 mM. Cells transfected with pCI vector were used as a negative control. n=3, error bars indicate SEM.

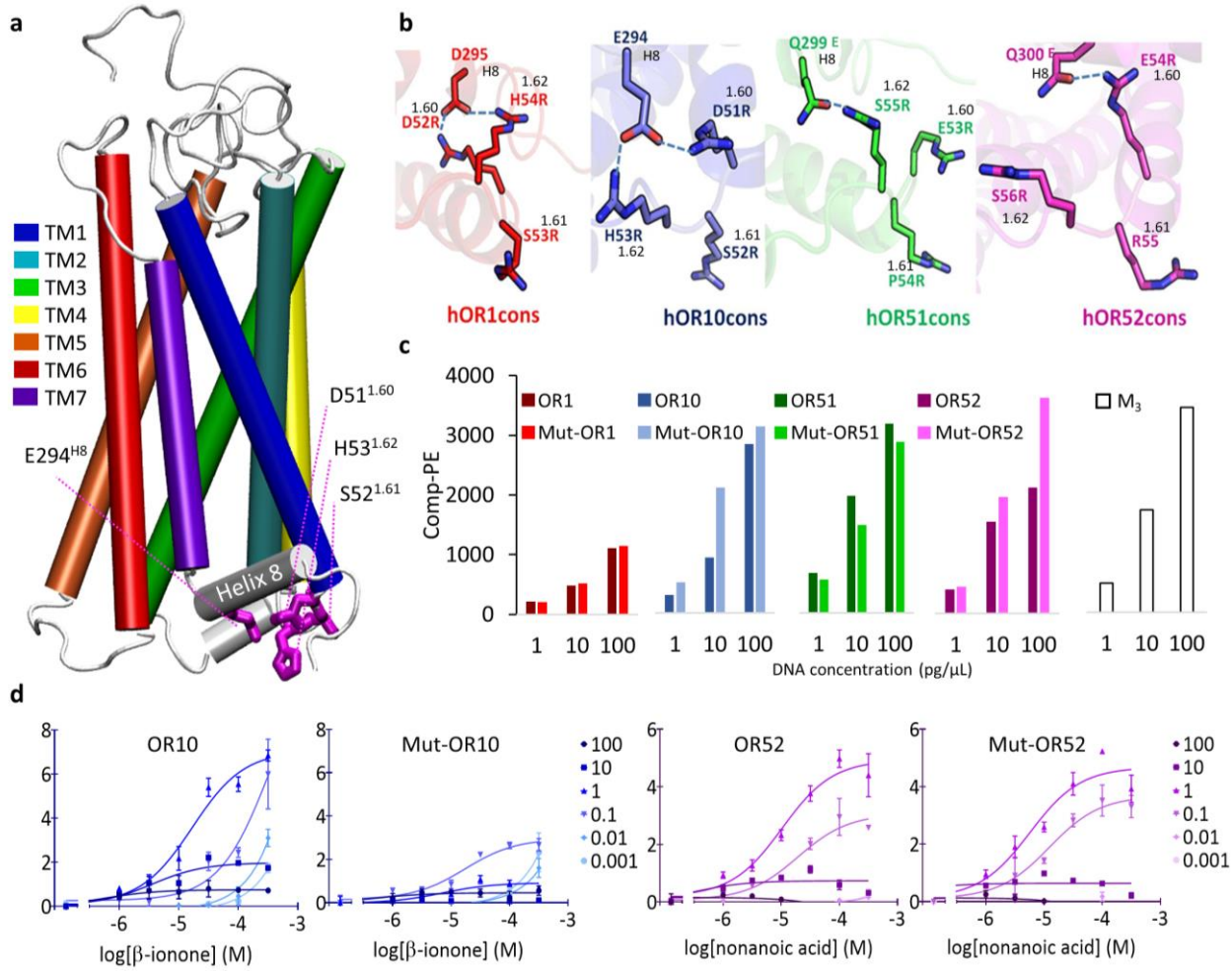

Figure 5

Improvement of OR expression by mutations in TM1 and Helix 8. **a**, Homology model of hOR10. Each TM is highlighted in a colored tubes and residues D51$^{1.60}$, S52$^{1.61}$, H53$^{1.62}$ and E294^{H8} are represented in licorice (pink). **b**, Zoom on the residues 1.60, 1.61, 1.62, H8 mutated for hOR1 (red), hOR10 (blue), hOR51 (green) and hOR52 (pink). Ionic interactions between the residues are shown by dotted lines. **c**, Expression analysis of hOR1, hOR10, hOR51, hOR52 and their mutant and the muscarinic receptor 3 (M3) at 1, 10 and 100 pg/µL of DNA in the transfection mix. **d**, Dose-response curve of hOR10 and hOR52 and their mutant at different DNA concentration (from 0.001 to 100 in pg/µL of transfection mix) to β-ionone and nonanoic

acid, respectively. The y-axis represent the luciferase luminescence normalized to the basal activity of each DNA concentration.

References

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 Niimura, Y., Matsui, A. & Touhara, K. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res* **24**, 1485-1496, doi:10.1101/gr.169532.113 (2014).

3 Firestein, S. How the olfactory system makes sense of scents. *Nature* **413**, 211 (2001).

4 Saraiva, L. R. *et al.* Combinatorial effects of odorants on mouse behavior. *Proceedings of the National Academy of Sciences* **113**, E3300-E3306 (2016).

5 McClintock, T. S. *et al.* Functional expression of olfactory-adrenergic receptor chimeras and intracellular retention of heterologously expressed olfactory receptors. *Brain Res Mol Brain Res* **48**, 270-278 (1997).

6 Lu, M., Echeverri, F. & Moyer, B. D. Endoplasmic Reticulum Retention, Degradation, and Aggregation of Olfactory G-Protein Coupled Receptors. *Traffic* **4**, 416-433, doi:10.1034/j.1600-0854.2003.00097.x (2003).

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

8 Kaiser, L. *et al.* Efficient cell-free production of olfactory receptors: detergent optimization, structure, and ligand binding analyses. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 15726-15731, doi:10.1073/pnas.0804766105 (2008).

9 Belloir, C., Miller-Leseigneur, M. L., Neiers, F., Briand, L. & Le Bon, A. M. Biophysical and functional characterization of the human olfactory receptor OR1A1 expressed in a mammalian inducible cell line. *Protein Expr Purif* **129**, 31-43, doi:10.1016/j.pep.2016.09.006 (2017).

10 Von Dannecker, L. E. C., Mercadante, A. F. & Malnic, B. Ric-8B promotes functional expression of odorant receptors. *Proceedings of the National Academy of Sciences* **103**, 9310-9314 (2006).

11 Noe, F. *et al.* IL-6–HaloTag® enables live-cell plasma membrane staining, flow cytometry, functional expression, and de-orphaning of recombinant odorant receptors. *Journal of Biological Methods* **4**, e81 (2017).

12 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).

13 Zhuang, H. & Matsunami, H. Evaluating cell-surface expression and measuring activation of mammalian odorant receptors in heterologous cells. *Nature protocols* **3**, 1402-1413, doi:10.1038/nprot.2008.120 (2008).

14 Wu, L., Pan, Y., Chen, G. Q., Matsunami, H. & Zhuang, H. Receptor-transporting protein 1 short (RTP1S) mediates translocation and activation of odorant receptors by acting through multiple steps. *The Journal of biological chemistry* **287**, 22287-22294, doi:10.1074/jbc.M112.345884 (2012).

15 Dunham, J. H. & Hall, R. A. Enhancement of the surface expression of G protein-coupled receptors. *Trends in biotechnology* **27**, 541-545 (2009).

16 Sharma, R. *et al.* Olfactory receptor accessory proteins play crucial roles in receptor function and gene choice. *Elife* **6**, doi:10.7554/eLife.21895 (2017).

17 Bubnell, J. *et al.* In Vitro Mutational and Bioinformatics Analysis of the M71 Odorant Receptor and Its Superfamily. *PLoS One* **10**, e0141712, doi:10.1371/journal.pone.0141712 (2015).

18 Jamet, S. *et al.* In Vitro Mutational Analysis of the β 2 Adrenergic Receptor, an In Vivo Surrogate Odorant Receptor. *PloS one* **10**, e0141696 (2015).

19 Lehmann, M., Pasamontes, L., Lassen, S. F. & Wyss, M. The consensus concept for thermostability engineering of proteins. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1543**, 408-415, doi:https://doi.org/10.1016/S0167-4838(00)00238-7 (2000).

20 Porebski, B. T. & Buckle, A. M. Consensus protein design. *Protein Engineering, Design and Selection* **29**, 245-251 (2016).

21 Ghosh, S. *et al.* Engineering Salt Bridge Networks between Transmembrane Helices Confers Thermostability in G-Protein-Coupled Receptors. *Journal of Chemical Theory and Computation* **14**, 6574-6585 (2018).

22 Tate, C. G. A crystal clear solution for determining G-protein-coupled receptor structures. *Trends in biochemical sciences* **37**, 343-352 (2012).

23 Ballesteros, J. A. & Weinstein, H. Analysis and refinement of criteria for predicting the structure and relative orientations of transmembranal helical domains. *Biophys J* **62**, 107-109, doi:10.1016/S0006-3495(92)81794-0 (1992).

24 de March, C. A., Kim, S.-K., Antonczak, S., Goddard, W. A. & Golebiowski, J. G protein-coupled odorant receptors: From sequence to structure. *Protein Science* **24**, 1543-1548, doi:10.1002/pro.2717 (2015).

25 Park, J. H. *et al.* Opsin, a structural model for olfactory receptors? *Angewandte Chemie International Edition* **52**, 11021-11024 (2013).

26 Rauscher, F. J., Morris, J. F., Tournay, O. E., Cook, D. M. & Curran, T. Binding of the Wilms' tumor locus zinc finger protein to the EGR-1 consensus sequence. *Science* **250**, 1259-1262 (1990).

27 Blatt, L. M., Davis, J. M., Klein, S. B. & Taylor, M. W. The biologic activity and molecular characterization of a novel synthetic interferon-alpha species, consensus interferon. *Journal of interferon & cytokine research* **16**, 489-499 (1996).

28 Warne, T., Edwards, P. C., Leslie, A. G. & Tate, C. G. Crystal structures of a stabilized β1-adrenoceptor bound to the biased agonists bucindolol and carvedilol. *Structure* **20**, 841-849 (2012).

29 Kiefer, H. *et al.* Expression of an olfactory receptor in Escherichia coli: purification, reconstitution, and ligand binding. *Biochemistry* **35**, 16077-16084 (1996).

30 de March, C. A. *et al.* Odorant receptor 7D4 activation dynamics. *Angewandte Chemie* **130**, 4644-4648 (2018).

31 McCaffrey, K. & Braakman, I. Protein quality control at the endoplasmic reticulum. *Essays Biochem* **60**, 227-235, doi:10.1042/EBC20160003 (2016).

32 Neuhaus, E. M., Mashukova, A., Zhang, W., Barbour, J. & Hatt, H. A specific heat shock protein enhances the expression of mammalian olfactory receptor proteins. *Chem Senses* **31**, 445-452, doi:10.1093/chemse/bjj049 (2006).

33 Dalton, R. P., Lyons, D. B. & Lomvardas, S. Co-opting the unfolded protein response to elicit olfactory receptor feedback. *Cell* **155**, 321-332, doi:10.1016/j.cell.2013.09.033 (2013).

34 Li, Y. R. & Matsunami, H. Unfolding the mystery of olfactory receptor gene expression. *Dev Cell* **27**, 128-129, doi:10.1016/j.devcel.2013.10.012 (2013).

35 Laird, D. W. & Molday, R. S. Evidence against the role of rhodopsin in rod outer segment binding to RPE cells. *Investigative ophthalmology & visual science* **29**, 419-428 (1988).

36 Dey, S. & Matsunami, H. Calreticulin chaperones regulate functional expression of vomeronasal type 2 pheromone receptors. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 16651-16656, doi:10.1073/pnas.1018140108 (2011).

37 Bushdid, C., de March, C. A., Matsunami, H. & Golebiowski, J. in *Olfactory Receptors: Methods and Protocols* (eds Fabio Marques Simoes de Souza & Gabriela Antunes) 77-93 (Springer New York, 2018).

38 Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome research* **14**, 1188-1190 (2004).