



HAL
open science

Molecular modeling of chemosensory receptors

Jody Pacalon

► **To cite this version:**

Jody Pacalon. Molecular modeling of chemosensory receptors. Cheminformatics. Université Côte d'Azur, 2022. English. NNT : 2022COAZ4084 . tel-04026629

HAL Id: tel-04026629

<https://theses.hal.science/tel-04026629v1>

Submitted on 13 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT

Modélisation moléculaire des récepteurs chimiosensoriels

Jody PACALON

Institut de Chimie de Nice, UMR CNRS 7272

**Présentée en vue de l'obtention
du grade de docteur en Chimie
d'Université Côte d'Azur**
Dirigée par : Dr. Sébastien Fiorucci
Soutenue le : 18 Novembre 2022

Devant le jury, composé de :
Dr. Sébastien Fiorucci, directeur de thèse,
Université Côte d'Azur, France
Dr. Lucie Delemotte, rapporteur, KTH
Royal Institute of Technology, Suède
Dr. Nicolas Floquet, rapporteur, Institut
des Biomolécules Max Mousseron, France
Pr. Isabelle Mus-Veteau, examinatrice,
Institut de Pharmacologie Moléculaire et
Cellulaire, France
Dr. Steffen Wolf, examinateur, University
of Freiburg, Allemagne

Abstract

Molecular modeling of chemosensory receptors

The perpetual struggle of living organisms to maintain their homeostasis despite an ever-changing environment has pushed evolution towards ever greater complexity. Even early in evolution, organisms were able to analyze their chemical environment through chemoperception and respond accordingly with specific behavior. The wide variety of chemicals has given rise to an equally diverse array of chemoreceptors to detect them. This evolution has resulted in the creation of specific and complex sensory organs as diverse as the antenna of *Drosophila*, the vomeronasal organ of the mouse, or the taste buds and olfactory epithelium of humans. There is a general organization of the olfactory system found in the animal kingdom, but the olfaction of insects and mammals is totally different at the level of receptors. In insects, ion channels are responsible for initiating signal transduction, whereas metabotropic G protein coupled receptors play this role in mammals. This work focuses on understanding the molecular basis of chemoreception at the level of olfactory receptors (ORs) in insects and mammals.

Humans possess about 400 subtypes of ORs able to sense a virtually infinite number of odorants, and 6 trace amine-associated receptors (TAARs) that bind specifically to volatile amines. Deciphering the combinatorial code of odorants is the first step in understanding olfaction and predicting the odor of a molecule based on its structure, but data are scarce. First, to accelerate the deorphanization process of mammalian olfaction, we implement machine learning models powered by *in vitro* and structural data and found 66 novel odorant-receptor pairs. Today, more than 50% of human ORs are deorphanized, allowing a finer understanding of the combinatorial code. Second, we predict the impact of a mutation in the activation process of the human receptor TAAR5, responsible for the detection of the trimethylamine rotten fish odor. This demonstrates how a joint approach combining molecular dynamic simulations combined and *in vitro* functional assays can decipher OR structure-function relationships. We then apply a similar protocol to get new insights into the importance of OR extracellular loops 2 and 3. We finally describe the ligand diffusion

pathway from the extracellular medium into the insect olfactory co-receptor (Orco) binding site. This work paves the way for the rational design of broad-spectrum insect repellents.

This thesis illustrates that computational approaches coupled to experimental ones, are powerful tools to study the sequence-structure-function relationships of olfactory receptors.

Keywords: Olfaction, GPCR, Molecular modeling, Molecular dynamics, Structure/function relationships.

Modélisation moléculaire des récepteurs chimiosensoriels

La lutte perpétuelle des organismes vivants pour maintenir leur homéostasie malgré un environnement en perpétuelle transformation a poussé l'évolution vers une complexité toujours plus grande. Dès le début de l'évolution, les organismes étaient capables d'analyser leur environnement chimique grâce à la chémoperception et de réagir en conséquence par un comportement spécifique. La grande variété de substances chimiques a donné lieu à un éventail tout aussi diversifié de chémorécepteurs pour les détecter. Cette évolution a abouti à la création d'organes sensoriels spécifiques et complexes aussi divers que l'antenne de la drosophile, l'organe voméronasal de la souris, ou les papilles gustatives et l'épithélium olfactif de l'homme.

Il existe une organisation générale des systèmes olfactifs que l'on retrouve dans le règne animal, mais l'olfaction des insectes et des mammifères est totalement différente au niveau des récepteurs. Chez les insectes, des canaux ioniques sont responsables de l'initiation de la transduction du signal, alors que des récepteurs métabotropiques couplés aux protéines G jouent ce rôle chez les mammifères. Ce travail vise à comprendre les bases moléculaires de la chémoréception au niveau des récepteurs olfactifs (RO) chez les insectes et les mammifères.

L'homme possède environ 400 sous-types de ROs capables de détecter un nombre virtuellement infini d'odorants, et 6 récepteurs associés aux amines traces (TAARs) qui se lient spécifiquement aux amines volatiles. Déchiffrer le code combinatoire des odorants est la première étape pour comprendre l'olfaction et prédire l'odeur d'une molécule à partir de sa structure, mais les données sont rares. Dans un premier temps, pour accélérer le processus de déorphanisation de l'olfaction des mammifères, nous mettons en œuvre des modèles d'apprentissage automatique alimentés par des données *in vitro* et structurales et découvrons 66 nouvelles paires odorant-récepteur. Aujourd'hui, plus de 50% des ORs humains sont déorphanisés, permettant une compréhension plus fine du code combinatoire. Deuxièmement, nous prédisons l'impact d'une mutation dans le processus d'activation du récepteur humain TAAR5, responsable de la détection de l'odeur de poisson pourri de la triméthylamine. Ceci démontre comment une approche conjointe combinant des simulations de dynamique moléculaire et des essais fonctionnels *in vitro* peut déchiffrer les relations structure-fonction des ROs. Nous appliquons ensuite un protocole similaire pour obtenir de nouvelles informations sur l'importance des boucles extracellulaires 2 et 3 dans la fonction des ROs.

Nous décrivons enfin le chemin de diffusion du ligand depuis le milieu extracellulaire jusqu'au site de liaison du corécepteur olfactif (Orco) des insectes. Ce travail ouvre la voie à la conception rationnelle de répulsifs pour insectes à large spectre.

Cette thèse illustre que les approches computationnelles, couplées aux approches expérimentales, sont des outils puissants pour étudier les relations séquence-structure-fonction des récepteurs olfactifs.

Mots-clefs : Olfaction, RCPG, Modélisation moléculaire, Dynamique moléculaire, Relations structure/fonction.

Acknowledgments

First, my supervisors. To Jérôme Golebioswki. "It's always the good ones who leave first". The animator and mentor of Chemosim left too soon, but the precepts you taught me are forever in my scientific heart. I try to apply them daily in the lab, like "keep it simple" or "quick and dirty". Although I abandoned the second one after trying to apply it in my romantic life for reasons I won't go into here. To Jérémie Topin. You were right. I could stop here, but I want to add that I don't know any more balaise scientific advisor in the featherweight category, none of them comes close to you! Thank you for being my hippocampus and don't forget to tell me when my storage capacities will be exceeded with my current plan. To Sébastien Fiorucci. You accepted me as a PhD student when I needed it, not even holding a grudge for all the times I corrected you at Unreal Tournament 2003. Thank you for all the times we have actively "generated discussion" together. And like for Jérôme I apply your precepts religiously; the proof is that I am going bald. To Serge Antonczak, Nicolas Baldovini and Nathanaël Guigo, your support has helped me a lot in some rough moments, and to say that I am grateful is an understatement.

Second, my jury. I would like to thank Dr. Lucie Delemotte, Dr. Nicolas Floquet, Pr. Isabelle Mus-Veteau and Dr. Steffen Wolf for agreeing to come and review my work of the last four years. "Hope is a virtue of slaves" said Cioran. But I hope you'll enjoy my work as much as you'll enjoy the sunshine on the French Riviera.

Because love of science and fresh water only last for so long, I am grateful to the GIRACT Foundation and the Roudniska Foundation and the Université Côte D'Azur for the financial support they have given me during these 4 years. I am particularly grateful for the scientific exchanges we had Gilles Sicard. Dolphins and sea snakes will totally lose their sense of smell long before you lose your interest in science, and I find that especially inspiring.

Because prediction backed by the harsh *in vitro* and *in vivo* reality is far more stylish than postdiction in the void, I want to thank the work of our collaborators without which this work could not have been. The Grenoble people (IBS), Christophe Moreau and Guillaume Audic, who kindly managed not to vomit when I tried to present them the active site of Orco in

animated gif (I finally understood it was a false good idea). The Dijon people from CSGA, Pr. Loïc Briand and Dr. Christine Belloir. It's a pleasure to work with biologists who are both interested in molecular modeling and Berlin beers. Pr. Hiroaki Matsunami and Dr. Claire De March who kindly invited me to step away from my computer screen for a few weeks at Duke University, USA, and everything was perfect (except the American food). Thanks to Yiseul Kim and Meera Rhyu from the KFRI (Korea); Yiqun Yu, Zhenjie Ma, Lun Xu, Weihao Li, Rui Xu, Xuewen Li and Wenwen Ren from Shanghai, and Hongmeng Yu from Beijing. Finally, the Koreans from Daegu Won Cheol Kim, Professor Cheil Moon. and Tammy Shim who invited us to a memorable visit of their laboratory.

Next are all the people I interacted with during my thesis. Tammy again dear, I miss you. We finally managed to glycosylate that spaghetti peptide together for the beauty of it, and God it was beautiful. A few years later, we have the ECL3 article ready to publish. I do not doubt that you will manage to finalize your thesis with as much brilliance. Cédric Bouysset, brologue. You have put up with me all these years, and that is a sure sign of pugnacity: you will go far. You've always been there to help me, whether it's to clean up my filthy scripts or to teach me how to use all the possibilities of my virtual reality headset, I owe you a debt of gratitude. Forever. Xiaojing Cong who regaled us both with exotic Chinese food and bad humor. Were these exotic shaped shooters really necessary? To the Matejmatician, we have been through so many unforgettable journeys together, such as the rat attack or the installation of Alpha Fold with singularity. You single-handedly saved us from statistical hell. Stay as a balaize and don't forget to judge you and Jana, Slovakian people are too good for their own good! Maxence Lalis the pirate, my little scientific brother from ISDD. We will all bury you (soon), but your unbridled good humor and equally bad theories will remain in our hearts forever. Xhino Meleqi, thank you for bringing a little sweetness into our rough lives in the form of beautifully virtual watermelon pieces. You're a good guy, but don't try to mine bitcoin with Azzurra, I saw you coming from afar. Jean-Baptiste Coffin. If you can become half as much articulated in ML than in legislation, Nobel prize is the only possibility.

Benjamin, the radiologist. I shouldn't have started this doctoral race with you, fake brother, but you only won by a little. I'll be the first to have a Pr. in front of my name anyway. Benjamin, the radiochemist. When I arrived in Nice, I found you lost among all these girls. Glad you found THE one. I greatly appreciated your brilliant mind, your unfailing sociability,

and your shitty humor despite your dubious origins, rosbif. I look forward to having a ricard with you and your alter ego Simon Bayle the carac'! Thanks to my best Peruvian poto Vanessa who taught me that I was more suited for savate than salsa, the harsh way. Pierre Matricon, I owe you so much beer with all the help you gave me, I hope your GABA_A receptors are desensitized when we hook up again if you want to survive. Looking forward to it, maybe in Uppsala.

To Timothée Dejour and Camille Coireau, my depraved brothers. You are both too weird to live but too shlag to die. Too many souvenirs, fortunately, souvenirs often dissolve in solvent. Camille I just pray one day you learn decency.

To Martine Dunoyer de Segonzac. I remember our first meeting and the treasures of patience you showed me, inversely proportional to my understanding of French administration. This was appreciated. To Uwe Meierhenrich. You promised me a quality thesis with Chemosim team. After a few years/papers later, I can say with confidence that you did not lie to me. Thank you for your unwavering support.

To my family. Mom. You are present at each of my battles. I have the humble claim to understand a little better the molecular basis of olfactory perception. But the molecular basis of a mom's love it's beyond me. To my little brother. I think my goal in my professional life is to recapture the fun we had playing together as kids (and getting paid for it). I think I'm close to it, and I wish you the same. To my little sister and her infinite patience with her big brother's characterized toxic masculinity. I'm proud of the woman you are (even if for me you will always be my little sister with chubby cheeks). For my father, who passed on his passion for medicine to me at a very young age. It has led me far, in multiple unexpected ways. If we had known that one day, I would be able to give you lessons on olfaction... The 10 first are free by the way.

And finally, to Yonie. Thank you for all the laughter, joy, patience, and unwavering support you have had with me over the years. It has helped me more than you know.

Table of contents

ABSTRACT	I
ACKNOWLEDGMENTS	V
TABLE OF CONTENTS	VIII
ACRONYMS	XI
INTRODUCTION	1
Origin of perception	1
Chemosensory perception	3
Overview of olfactory perception	6
Olfaction across evolution	9
Computational methods applied to olfaction	19
CHAPTER I	
INSECT ODORANT RECEPTOR CO-RECEPTOR (ORCO)	45
Publication 1	48
Ligand of the conserved insect odorant receptor co-receptor reaches its binding site through a dynamic translocation pathway.	
Abstract	49
Introduction	50
Results	52
Discussion	63
Material and Methods	67
Supplementary data	73
CHAPTER II	
HUMAN TRACE AMINE ASSOCIATED RECEPTORS (TAAR)	90
Publication 2	92
TAARs response to amine are largely affected by sequence variants.	
Abstract	93

Introduction	93
Results	94
Supporting information	99
CHAPTER III	
OLFACTORY RECEPTOR EXTRACELLULAR LOOPS (ECL)	118
Publication 3	121
Extracellular loop 2 of G protein-coupled olfactory receptors is critical for odorant recognition.	
Abstract	122
Introduction	122
Results	124
Discussion	131
Material and Methods	132
Supporting information	138
Publication 4	153
The third extracellular loop of mammalian odorant receptors is involved in ligand binding.	
Abstract	154
Introduction	155
Results	156
Discussion	162
Material and Methods	164
Supplementary material	169
CONCLUSIONS	177
Discussion	180
Perspectives	182
ANNEXES	189
Publication 5	190
Functional molecular switches of mammalian G protein-coupled bitter-taste receptors.	
Abstract	191
Introduction	191
Methods	193

Results and discussion	198
Conclusions	205
Supplementary information	214
Publication 6	229
Large-Scale G Protein-Coupled Olfactory Receptor-Ligand Pairing.	
Abstract	230
Introduction	231
Results	233
Discussion	240
Material and Methods	242
Supporting Information	245
Publication 7	265
Smell and taste changes are early indicators of the COVID-19 pandemic and political decision effectiveness.	
Abstract	266
Introduction	266
Results	269
Discussion	273
Methods	277
Supplementary information	286

Acronyms

AF2	AlphaFold2
cAMP	Cyclic adenosine monophosphate
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CG	Coarse grain
Cryo-EM	Cryogenic electron microscopy
CV	Collective variables
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>
ECL	Extracellular loop
<i>E. Coli</i>	<i>Escherichia Coli</i>
FEP	Free energy perturbation techniques
GPCR	G protein coupled receptor
HCA	Hierarchical clustering analysis
ICL	Intracellular loop
IR	Ionotropic glutamate receptor
MD	Molecular dynamic simulations
<i>M. hrabei</i>	<i>Machilis hrabei</i>
ML	Machine learning
MPNN	Message passing neural network
MSA	Multiple sequence alignment
OBP	Odorant binding protein
ODE	Odorant degrading enzyme
OR	Olfactory receptor
Orco	Odorant receptor co-receptor
OSN	Olfactory sensory neuron
PDB	Protein data bank
pLDDT	Predicted local-distance difference test
REMD	Replica exchange molecular dynamics
RMSD	Root-mean-square deviation
SBDD	Structure-based drug discovery
SNP	Single nucleotide polymorphism
<i>S. littoralis</i>	<i>Spodoptera littoralis</i>
SOR	Structure-odor relationships
TAAR	Trace-amine associated receptor
TEVC	Two-electrode voltage-clamp
TM	Transmembrane helix
TPL	TOPLESS-like protein
TRP	Transient receptor potential channel
VNO	Vomer nasal organ
WT	Wild-type

Introduction

Origin of perception

Most basic form of perception?

In the most basic sense, a living organism can be seen as a compartmented system using outside energy to maintain its internal order against entropy. This perpetual fight to stay in homeostasis despite its always changing environment, pushed evolution to more and more sophisticated perceptual systems. Perception can be classified between interoception, when an organism detects its own internal state, and exteroception, when the perception is directed toward the outside environment. Early in evolution, organisms have been able to analyze their environment through exteroception using their senses and react accordingly with precise behavior [1].

Perception comes with a high energy cost and is shaped by selective pressure [2]. But for the organism, it can be central to its survival because it goes with multiple benefits. Indeed, perception increase the ability to identify resources (finding food, communicating, housing, mating) and dangers (avoiding predators) [3].

Evolution of perception

For the most basic of organisms able to perceive, collecting the information and acting upon it is computed in the same cell. For the most complex organisms, millions of neurons can be implicated in the peripheral step of detecting the percept, while billions can be needed to process then act upon the perception toward the behavior [2].

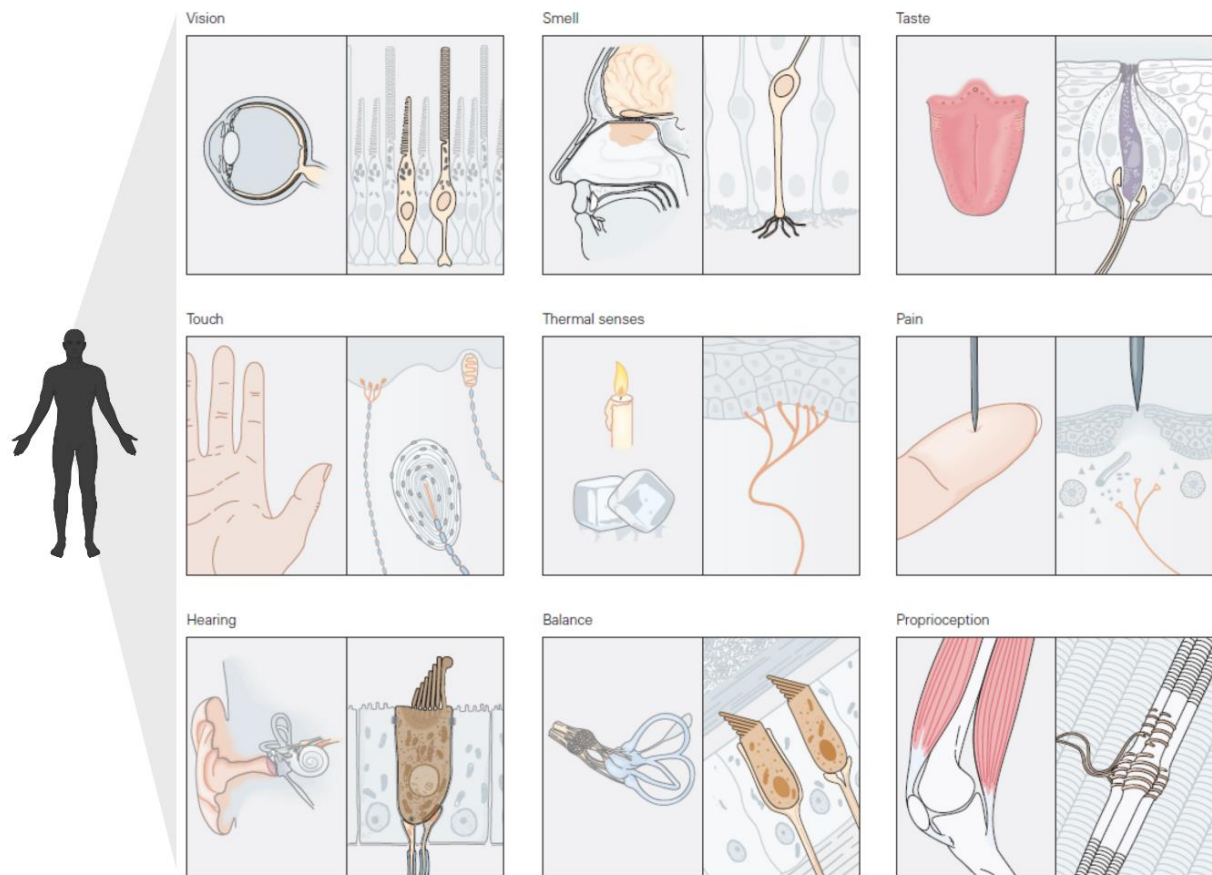


Figure 1: Vision of the organs and neurons associated with the main human senses dedicated to exteroception (vision, smell, taste, touch, and hearing) and to interoception (pain, visceral sensations, proprioception). Each sense is associated with a specific type of neuron expressing certain types of receptors. Vision is mediated by the photoreceptors within the eyes; smell, taste and pain by chemoreceptors located respectively in the nose, the mouth, and the whole body; touch by mechanoreceptors located on the skin; proprioception and visceral sensations by receptors distributed on the whole body. Adapted from “Principles of neural science 6th edition” [4].

Perception has multiple forms in the natural kingdom and is dependent on the organs. Human can feel touch, pain, pressure, cold, warm, the position of their body in space (proprioception), see, hear, smell and taste (figure 1) [5]. Even if there is often overlapping between our senses and the senses of animals, the perceptual space can be different: Mosquitos can smell CO₂ [6] and bats can hear ultrasounds [7]. Some animals have unique abilities like perceiving magnetic [8] or electrical fields [9]. Perception can even evolve in the lifetime of an organism, like the sea squirt that is equipped with a basic sensory system at the larvae stage for detecting a suitable environment to become sessile for its adult life [1]. Some researchers support the idea that sensory systems evolve at the pace of the change in the specie's environment and may drive the evolution of new species [10].

Chemosensory perception

Definition and diversity of chemosensory perception

Chemosensory perception is the ability to perceive a chemosensory (chemical) signal from the environment. Chemosensory perception can be described as a discrete sense, as opposed to the continuous physical parameters that hearing (air vibration), vision (light wavelength), or touch (pressure) detect [11, 12].

The diversity of chemicals that can be recognized by living organisms through chemosensory perception is astonishing. A recent paper focusing exclusively on volatile compounds estimated that more than 40 billion chemicals could be odorous [12]. Chemical space is multidimensional, and its boundaries are unknown, unlike the physical properties of signals for vision and hearing. Chemosensory signals can range from a single compound to complex mixtures of different concentrations. Chemosensory molecules can be classified according to their solubility in water and volatility in air. Pheromones are sub-class of chemical signals specifically designed to mediate intraspecific behaviors [13].

Chemosensory perception may be the most basic and widespread sensory modality in animals [14]. Chemosensory perception is found in various forms in most living species, even the simplest (figure 2). Indeed, some form of chemosensory perception has been described in unicellular and multicellular organisms, including bacteria [15], yeast [16], plants [17], worms [18], crustaceans [19], insects [20], fishes [21], birds [22] and mammals [23]. Conversely, every living organism emits some chemosensory chemicals due to active communication or to its metabolism.

Perhaps the most basic use of chemosensory perception is chemotaxis, exemplified by the simple but remarkable ability of *Escherichia Coli* (*E. Coli*) to orient itself to food and steer clear of toxic environments [23, 25]. But in addition to aiding navigation, chemosensation can shape more complex behaviors essential for survival and reproduction, such as detection and evaluation of food availability and quality, predator avoidance, mate identification and attraction, offspring care, housing, and complex intra- and interspecific communication [18].

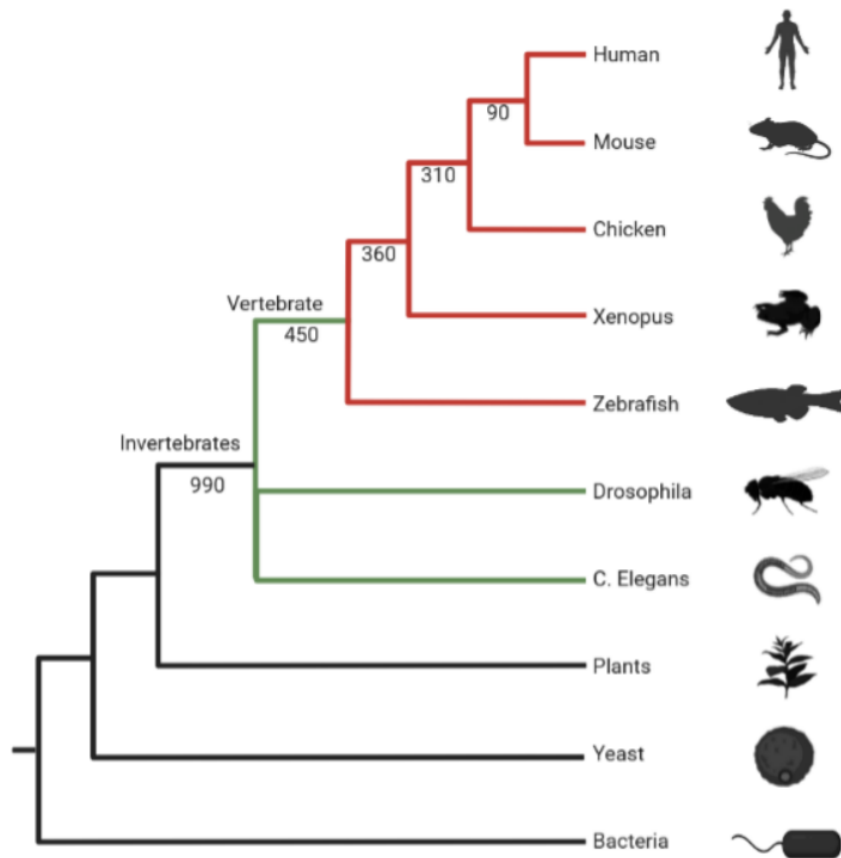


Figure 2: Phylogenetic tree with model organisms: bacteria (*Escherichia coli*), unicellular eukaryotes (*Saccharomyces cerevisiae*), nematodes (*Caenorhabditis elegans*), insects (*Drosophila melanogaster*), fish (Zebrafish or *Danio rerio*), amphibians (*Xenopus laevis*), birds (*Gallus gallus domesticus*), rodents (*Mus musculus*), primates (*Homo sapiens sapiens*). Number in million years. Adapted from Wheeler and Brändli (2009) [24].

Evolution of chemosensation

Since all living cells are "irritable to chemicals" [26], this ability may have led to the evolution of specific receptors to detect these chemicals, and from complexification to the creation of specific chemosensory organs and neural systems [27]. The evolution of complex chemosensory organs is a gradual and cumulative process, as environmental variations, and natural selection form new adaptations to maximize survival [28].

At the anatomical level, chemosensory perception in humans is divided into two main senses, taste, and olfaction, due to the presence of two specific organs (the tongue with taste buds and the nose with the olfactory epithelium). Chemesthesis is a third chemosensory perception capable of detecting chemical substances, including the trigeminal neurons. Even if taste and

olfaction have specific organs in human, there is a huge connectivity and integration of smell and taste signals in the brain to give flavor [29].

The boundaries between the chemical senses become blurred as one moves down the evolutionary ladder. Neurons in the gustatory cortex of rats can respond to both gustatory and/or olfactory signals, demonstrating convergence of flavor and palatability coding in this area [30]. The vomeronasal organ (VNO), an organ specifically designed to recognize pheromones, has been lost in humans but is present in most mammals. Pheromones can be detected by the olfactory epithelium and the VNO in mammals, but they use different neural pathways in the brain like taste, olfaction and chemesthesis in humans [13]. In birds, the proximity of chemosensory neurons to olfactory sensory neurons in the nasal cavity, and to the gustatory sensory neuron in the oral cavity, made the separation between these chemosensory senses difficult [31]. In snake, the tongue can express olfactory but not gustatory genes [32] and is used to provide volatile and non-volatile chemical cues to the VNO [33].

These examples show that taste and olfaction are not that well defined for terrestrials' animals even if olfaction is in general considered as a distance sense and gustation as a contact sense. But the limit blur even more when we consider aquatic life forms.

One view of chemosensory evolution from water to terrestrial life state that this transition forced chemosensory systems to adapt from detecting mostly hydrophilic compounds to their new hydrophobic and volatile ones [34]. Mollo *et al.* present an opposite view and suggest that aquatic life forms can detect both hydrophobic and hydrophilic chemical clues and this ability was conserved during the transition [35, 36]. The nudibranchs gastropods can sense hydrophobic odorant with their oral tentacles like an aquatic nose [35]. They can defend themselves by secreting terpenoids (molecules recognized as volatiles odorant in terrestrial milieu) on sacrificial parts of their body to survive predator attack [37]. In *Caenorhabditis elegans* (*C. elegans*), taste and olfaction present ambiguity as ammonium acetate is sensed both by olfactory and gustatory receptors [38]. As a result, Mollo *et al.* (2022) propose to unify "all chemosensory modalities into a single sense" to go out from the anthropomorphic view of chemical senses [39]. With these details in mind, this work will focus on olfactory perception in the broad sense.

Overview of olfactory perception

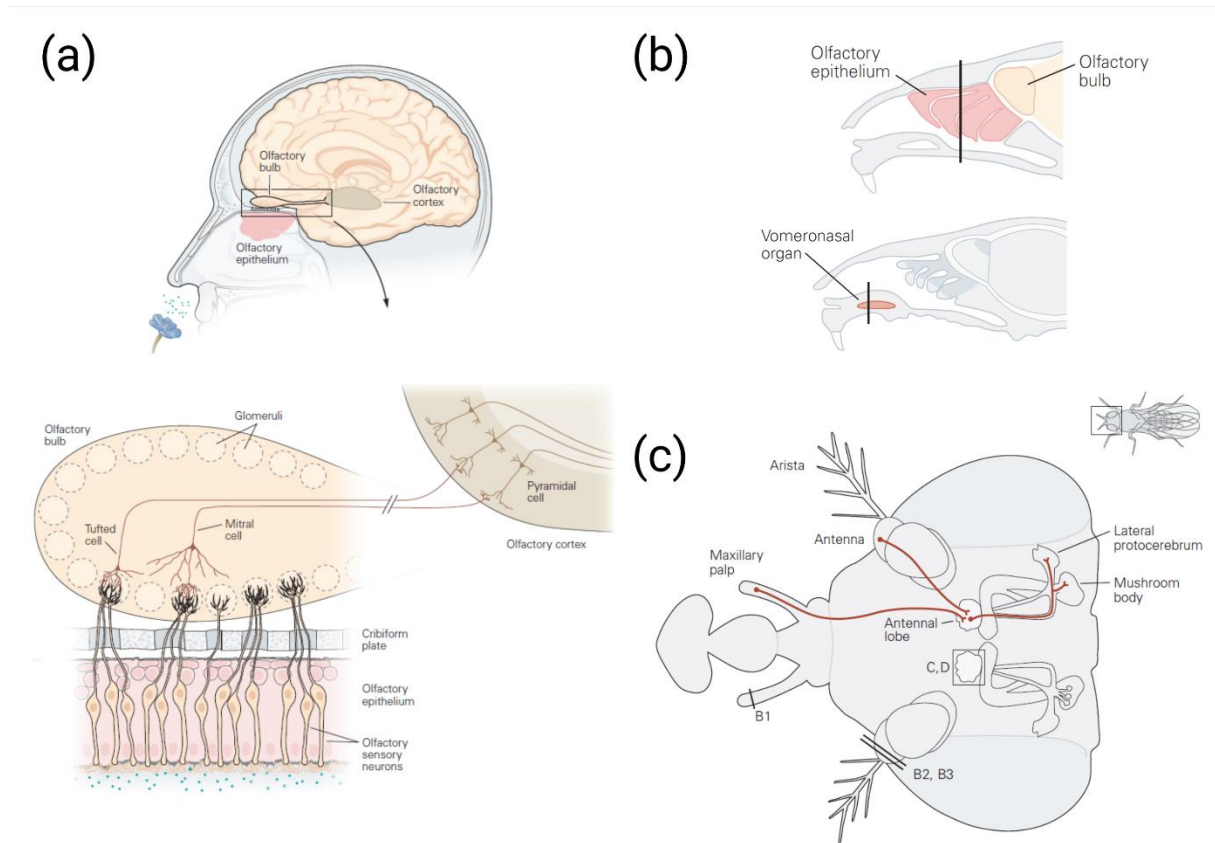


Figure 3: Anatomy of olfactory perception in mammals and insects. **(a)** Overview of the human olfactory system. Odorants enter the nasal cavity to reach the olfactory epithelium. Olfactory sensory neurons detect the odorant through their ORs. The axons of the neurons project to the glomeruli of the olfactory bulb through the sieve plate. The olfactory signal is then transmitted further into the olfactory cortex to ultimately create the odor percept. **(b)** Compared to humans, rodents have a specific organ for detecting pheromones called the vomeronasal organ (VNO). **(c)** Overview of the insect olfactory system. Odors are detected by olfactory sensory neurons located in both the antenna and maxillary palps via their ORs. The axons of these neurons project to the antennal lobes. The olfactory signal is then transmitted into the lateral protocerebrum and the mushroom body to create odor perception. Adapted from "Principles of Neural Science 6th Edition" [4].

The olfactory system presents remarkable similarities within the animal kingdom. Odorant molecules eventually arrive in the vicinity of highly diverse olfactory receptors (ORs) on the surface of olfactory sensory neurons in the animal's sensory organ. Certain perireceptor events can modify or impact the concentration, availability of chemicals. The chemical information is

then transformed into electrical information through olfactory signal transduction. For all animals except mollusks, the information is then transmitted directly in the central nervous system to the first synaptic relay: the olfactory lobe in crustaceans, the antennal lobe in insects, and the olfactory bulb in vertebrates. In both invertebrates and vertebrates, the first synaptic relay is organized in the glomerulus where all neurons expressing one type of receptor converge on one or a few glomeruli. The conservation of this strategy during evolution suggests the importance of this spatial organization for processing odorant information. Axons from the first relay project then extend into the olfactory cortex for mammals, and into the lateral protocerebrum and mushroom body for insects, to create odor perception and subsequent behavior (Figure 3) [27].

Odorants

Chemical olfactory cues are extremely diverse and can be classified according to their physicochemical properties and in particular to their ability to be volatile in air and soluble in water. Odorants can be volatile but insoluble in water, non-volatile but insoluble in water, volatile and soluble in water, non-volatile and insoluble in water [36]. Recent work by Mayhew *et al.* (2022) tentatively predicts the boundaries of the odorant space using machine learning. Their models trained on previously known odorants suggest that transport from the environment to the binding site is the most crucial feature a molecule must possess to be an odorant. Using a database of 166 billion chemicals, their models predict at least 40 billion of possible odorant compounds [12].

Odorants can also be classified according to the function they perform. Environmental odorants include odors of food, water, nesting sites, or danger such as fire. Allelochemicals are odors emitted by other species or organisms and signal the presence of prey, homing, symbiotic associations, territorial marking, predator avoidance, metamorphosis and growth, or pollination. Pheromones are a group of odors emitted by conspecifics and can signal sexual attraction, mark individual identity or membership in a social group [27].

Olfactory neurons

Olfactory neurons are extremely similar between different animal species capable of olfactory perception, even though many invertebrates do not have complex organs like vertebrates [40].

A bipolar olfactory neuron with cilia bathed in a protective fluid, projects its axon directly to the organism's central nervous system. Different types of cells support the neuron. The cilia increase the contact surface between the parts of the neuron possessing the ORs, and the extracellular medium (see figure 4). Because olfactory neurons are continually in direct contact with the outside world, there is ongoing neurogenesis to replace them throughout the life of the animal, which is a rare feature for neurons in an adult animal. In humans, olfactory neurons are replaced approximately every 5 to 7 weeks [41, 42].

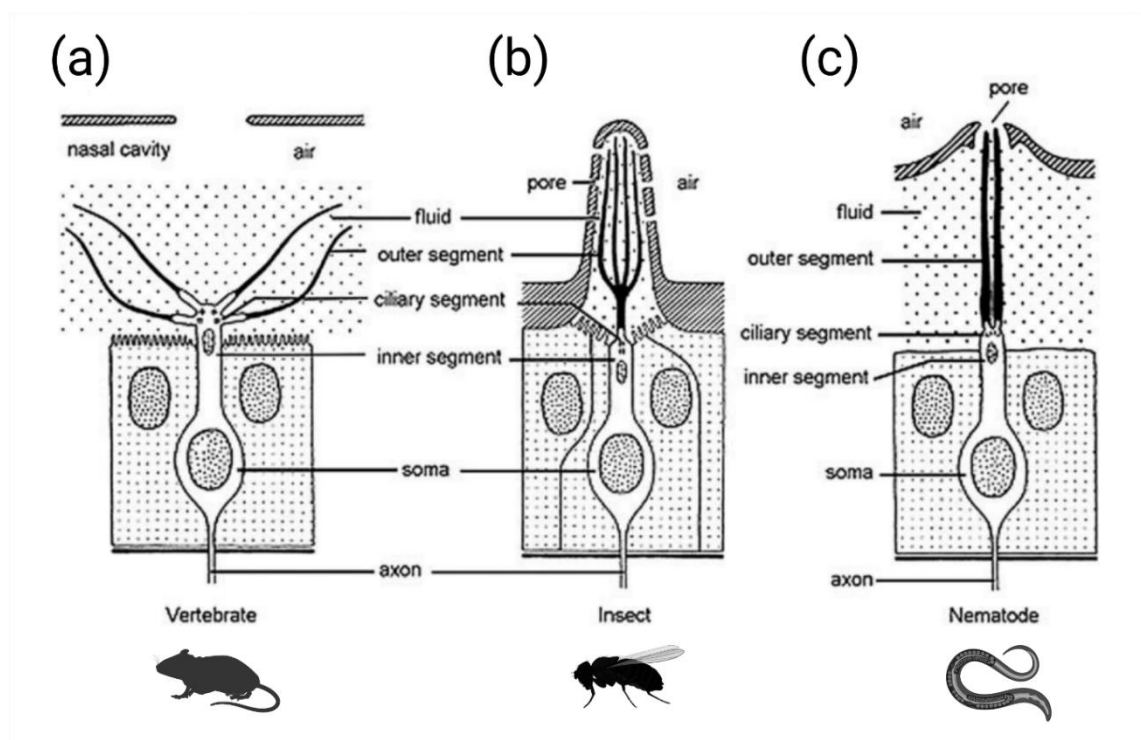


Figure 4: Comparison of olfactory sensory neurons in vertebrates (a), insects (b) and nematodes (c). All species have bipolar neurons with cilia containing ORs, bathed in a protective fluid. Adapted from Ache *et Young* (2005) [27].

Olfaction across evolution

Unicellular organisms

Prokaryotes

Bacteria use both an ionotropic and a metabotropic chemosensory system. They can sense amino acids both using precursors of ionotropic glutamate receptors [43] and a metabotropic two-component signal transduction system with chemosensory receptors (called MCPs for methyl-accepting chemotaxis proteins), a histidine kinase Chemotaxis protein A (CheA) and a response regulator (see figure 5). MCPs are transmembrane proteins that dimerize and assemble into trimers of dimers. Genes encoding the chemosensory receptors are by far the most diverse of the bacterial chemosensory system [44]. Where *E. coli* has 5 chemoreceptor subtypes triggering the same signaling cascade, some bacterial species have up to 80 chemoreceptors with up to 4 chemosensory pathways [45]. Among its 5 receptors, CheM (Chemotaxis protein M) and CheD (Chemotaxis protein D) allow to detect aspartate and serine respectively [15]. We can then consider bacterial gene CheM and CheD as encoding both “olfactory-like” or “gustatory-like” receptors. Indeed, in some teleost fishes, the olfactory system plays a dominant role in the detection of amino acids [46], whereas it’s the gustatory system in humans [47]. The study of the odor space of bacteria is particularly important in the context of antibiotic resistance. Microbial odorants can serve as odorant messengers and are involved in bacterial-eukaryotic interactions. A better understanding of these compounds could allow the design of antibacterial odorants [48, 49].

Eukaryotes.

Yeast, like bacteria, possess a dual metabotropic (G protein coupled receptor, GPCR) and ionotropic (transient receptor potential channel, TRP) chemosensory system. Most of the proteins involved in the metabotropic pathway of GPCR signaling were present in the last common ancestor of eukaryotes [51]. Even if lower eukaryotes have typically a lower number of GPCRs than mammals, unicellular eukaryotes have used this signaling pathway for chemoperception. Indeed, yeast possess 3 G protein coupled receptors to detect sugar and pheromones: Sterile 2 (Ste2), Sterile 3 (Ste3) and G-Protein coupled Receptor 1 (Gpr1) [52].

Yeast possess also a TRP ionic channel called TrpY1 who can sense odorants like indole and other aromatic compounds. This channel is considered as an ancient form of chemosensory receptors ion channels [53].

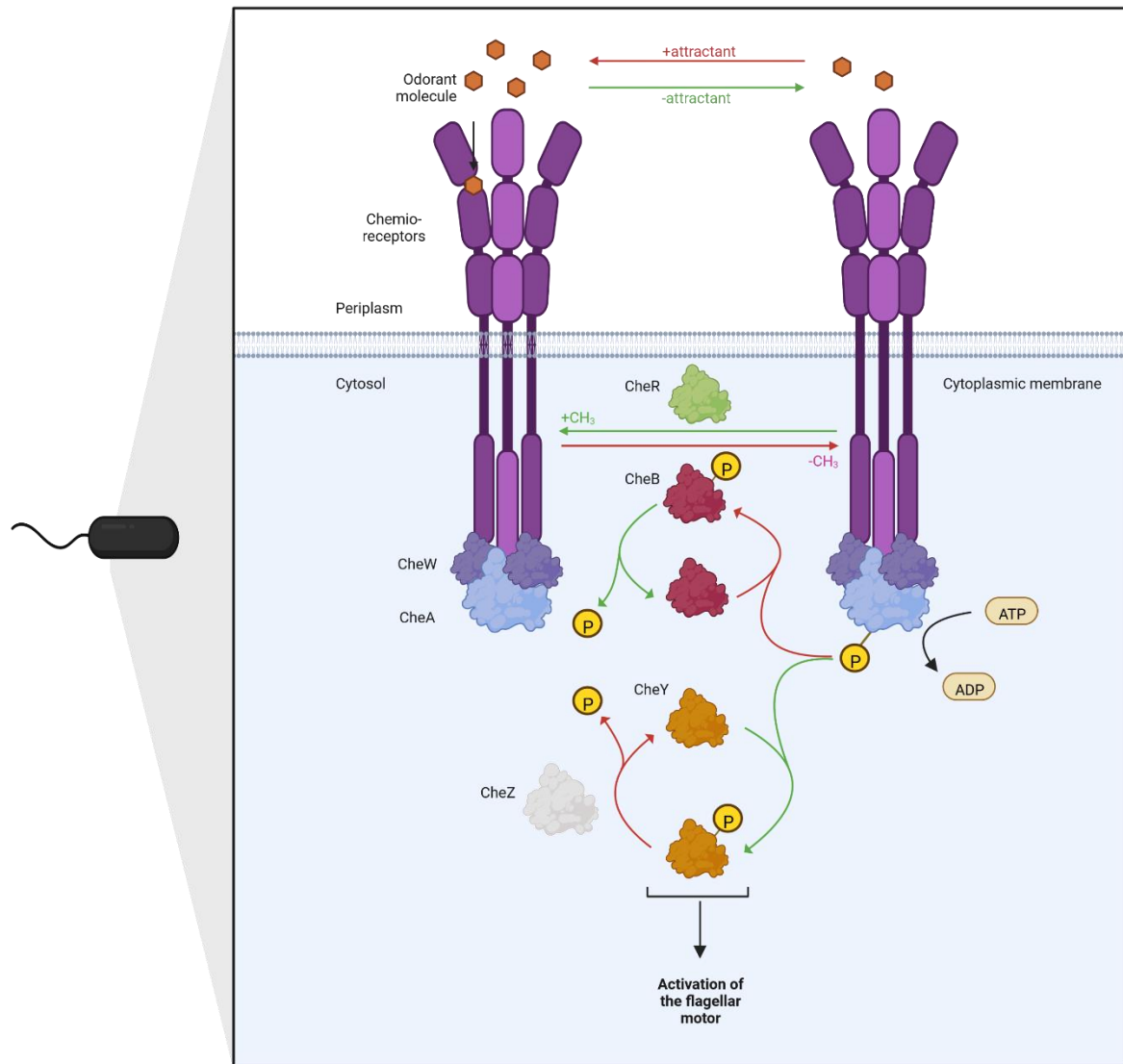


Figure 5: Chemotaxis of *E. Coli*. CheA, CheB, CheR, CheW, CheY and CheZ stand for Chemotaxis protein A/B/R/W/Y and Z respectively. Attractive or repulsive compounds are detected by chemosensory receptors composed of trimers or dimers. CheW forms a complex with CheA (protein kinase) and the cytosolic part of the receptor. Other modulating proteins are present, such as CheB and CheR (methyl donors or acceptors) and CheZ (phosphatase). The role of these proteins is to regulate the phosphorylation of CheY, which is responsible for triggering flagellar movement. Adapted from Houten (2015) [50].

Multicellular organisms

The transition from unicellular to multicellular organisms has allowed many evolutions. Cells were able to specialize, and sensory neurons emerged, along with sensory organs. At the receptor level, however, it is remarkable that evolution has preserved the signal transduction methods of unicellular organisms: metabotropic receptors or ionotropic receptors are still found on chemosensory neurons. However, the activation of chemosensory receptors following an olfactory signal ultimately induces a change in the membrane potential of the neuron. The chemical information is transformed into electrical information and modifies the activity of the neurons that receive this information [54].

Plants

Surprisingly, plants can emit, receive, and react to olfactory signals called semiochemicals. When some species are attacked by herbivorous insects, they may emit volatiles to attract predators of the herbivorous insects or warn their conspecifics to induce a defensive response [55]. The molecular basis of the perception of volatile odorant molecules in plants is still not well understood. For tobacco plants, it seems that plants have evolved an olfactory recognition system using nuclear proteins rather than membrane receptors like animals. Caryophyllene emitted by neighboring plants diffuses passively (or with unknown transporters) towards the nucleus. Caryophyllene then binds to transcriptionally co-repressive proteins, called TOPLESS-like proteins (TPLs), thus modulating gene expression (figure 6). Although this mechanism of olfactory perception in plants cannot be generalized at present, the presence of a large quantity of transcription factor genes, as well as their evolutionary increase compared to animals, points in this direction [56].

Nematodes

C. elegans, belonging to the nematode phylum, has about 1300 chemosensory receptors and 400 pseudogenes classified into 19 families. These genes belong to the G protein-coupled receptor family and represent 8.5% of all *C. elegans* genes. Yet only 6 receptor subtypes have been deorphanized [57]. *C. elegans* possess also ionotropic TRP channels [53] and different

transmembrane guanylyl cyclases like ODRant response abnormal receptor 1 (ODR-1) [58]. Each of the 32 sensory neurons expresses several chemosensory receptors and can discriminate multiple odorants. This ability seems to derive from arrestin-mediated desensitization of stimulated ORs [59]. However, despite the diversity of receptors, the spectrum of odorant recognition remains limited [18].

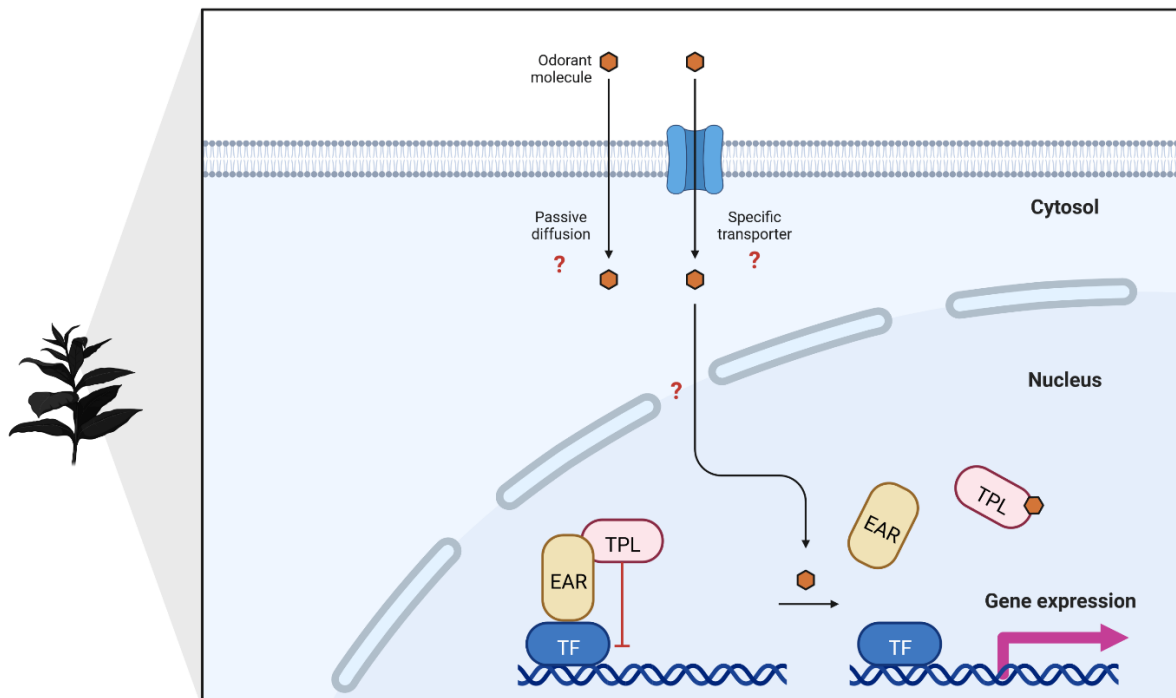


Figure 6: A tentative model of olfactory transduction in plants (tobacco). TPLs are transcriptional co-repressors and can bind caryophyllene. Caryophyllene diffuses from the extracellular to the cytosol and nucleus by passive diffusion or by specific transporters. Upon binding to TPL, NtOsomotion expression is induced. Adapted from Nagashima *et al.* (2019) [56].

Genesis of the combinatorial code of olfaction

Olfactory perception after nematodes shows great similarities between different animal species. The olfactory neurons begin to express a single OR subtype unlike *C. elegans* (fishes [60], insects [61], rodents [62], mammals [63]). In this organization, an OR can detect one or more odorants, and an odorant is detected by one or more receptors, which is called the combinatorial code.

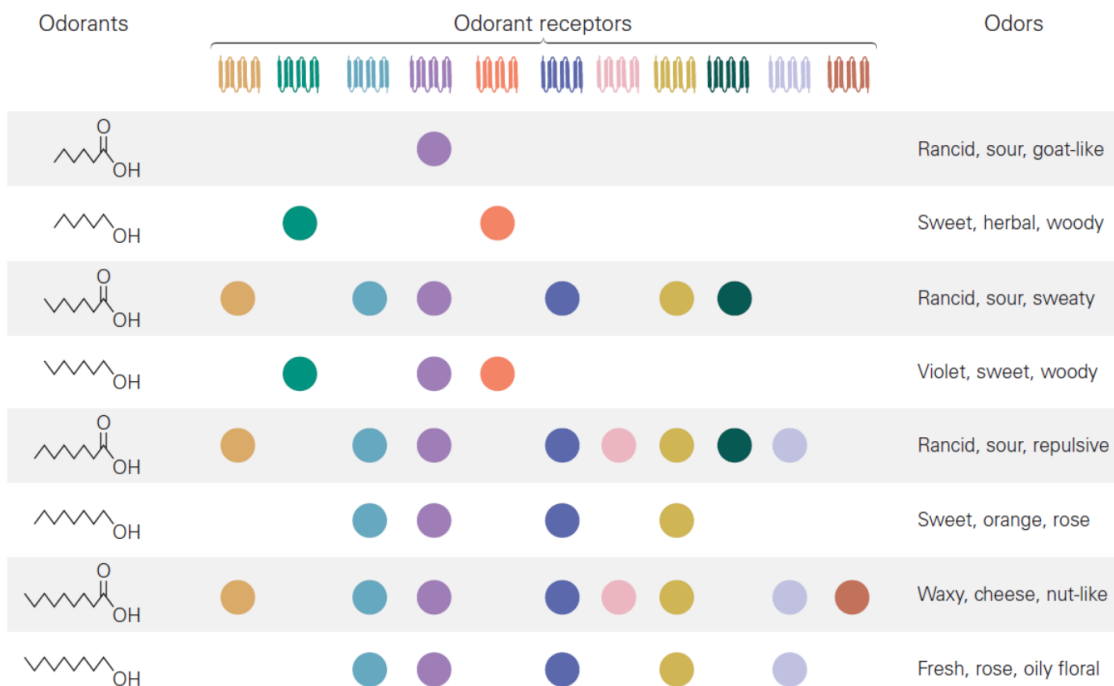


Figure 7: The combinatorial code of olfaction. Each odorant is recognized by a combination of receptors. A receptor can recognize one or several odorants (narrowly to broadly tune recognition spectrum). This organization explains why 2 chemically close molecules can produce very different olfactory perceptions. Adapted from “Principles of neural science 6th edition” [4].

Genesis of the perireceptors processes

In the protective fluid around the neurons, there are enzymes and buffers that help olfactory perception and are retained in most animals. These are the so-called perireceptors processes. Odorant-binding proteins (OBPs) are small soluble proteins present in terrestrial animals but absent in crustaceans and fish. They have been described as an adaptation to terrestrial life [36]. Insect OBPs are multiple, some are used to transport and solubilize odorant molecules while others are specialized to bind pheromones. The insect OBP repertoire varies greatly between species, suggesting an accelerated rate of gene turnover driven by a need to adapt to the environment [64]. Mammalian OBPs, belonging to the lipocalin family, are completely different in terms of sequence [65]. The second group of proteins found in the mucus of the olfactory epithelium are xenobiotic enzymes necessary for the deactivation of the odorant stimulus. This diverse family called odorant degrading enzymes (ODE), includes esterases, transferases and cytochrome P450 monooxygenases, the latter being abundantly expressed in the olfactory tissues of both insects and mammals [66, 67].

A divergence in the transduction mechanism

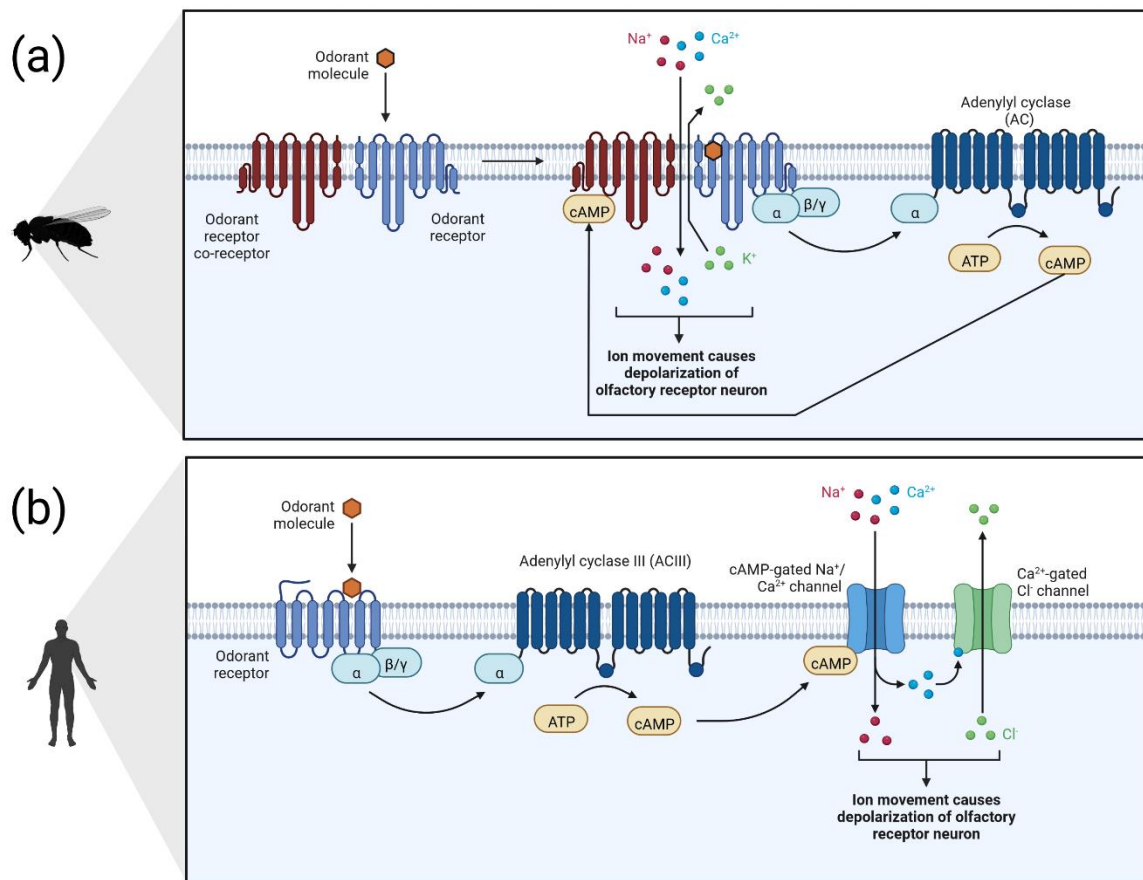


Figure 8: Mechanisms of olfactory transduction in insects and mammals. **(a)** Ionotropic transduction in insects. An odorant opens the heteromeric ion channel by binding to the OR part. The movement of ions causes a depolarization of the neuron. Recent work highlights the activation of the G protein by the OR, activating adenylyl cyclase and producing the second messenger cAMP. The cAMP then binds to the odorant co-receptor (Orco), increasing the activation of the channel further. **(b)** Metabotropic transduction in mammals. An odorant activates the OR, in turn activating the G protein. The G protein then activates adenylyl cyclase, initiating the production of the second messenger cAMP. The cAMP opens the $\text{Na}^+/\text{Ca}^{2+}$ channels managed by cAMP. The cytosolic increase in Ca^{2+} concentration in turn opens Cl^- channels, which further depolarizes the neuron.

However, the similarities of the olfactory system stop at the level of receptors when considering invertebrates and vertebrates: There is a strong difference in the mechanism of signal transduction between insects (see Figure 8a) on the one hand, and mollusks, fish, birds, and mammals on the other hand (see Figure 8b). While mammalian ORs are metabotropic GPCRs, olfaction in insects is primarily ionotropic. Why nature evolved two different transduction systems remains a matter of debate.

Vertebrate ORs are members of the class A (or rhodopsin-like) GPCR family. They have a common 7 transmembrane helices fold (Figure 9c), are about 300 amino acids long, and possess several motifs characteristic of GPCRs (GN in TM1, DRY in TM3, KA in TM6, and NPxxY in TM7), but also OR-specific motifs such as the PMYxFL motif in TM2, MAYDRYVAIC in TM3, SY residues in TM5, RxKAxxTCxSH and FY in TM6. The orthosteric cavity shows extreme variability, responsible for the wide spectrum of recognition power of the OR family ligands [68, 69].

Insects have a system of highly variable ORs subtypes that associate with a highly conserved olfactory receptor co-receptor (Orco) without which they cannot function (Figure 9a). Insect ORs and Orcos are 7-helix transmembrane proteins with an N-terminal cytosolic part and a C-terminal extracellular part, unlike GPCRs (Figure 9b). They assemble in a yet unknown stoichiometry to form a central pore, and their helices 4, 5, and 6 extend well beyond the cytoplasmic side to form the anchor domain [70, 71].

These two pathways have different temporal properties: Transduction through ion channels is rapid, in the micro to millisecond range, compared with metabotropic transduction, which takes 50 to 150 milliseconds to produce second messengers and activate secondary effectors. Signal deactivation is also faster, as dissociation of the ligand from the ionotropic receptor results in rapid channel closure. In contrast, metabotropic pathways require the metabolization of second messengers to terminate the signal. In addition, the energy cost of metabotropic transduction is higher than that of ionotropic transduction [73]. In contrast, metabotropic transduction allows both a strong amplification of the signal and a finer regulation of it. Indeed, a GPCR will be able to activate several G proteins, which will lead to the production of a large quantity of second messengers. Signal regulation can occur at several levels: phosphorylation of the receptor by kinases and its binding to beta-arrestin, but also regulation of each piece of the transduction ($G_{\alpha\text{olf}}$, adenylyl cyclase III, cAMP, CNG and Ca^{2+}). This is not the case for transmission in insects, although signal amplification through the G protein pathway has been demonstrated [74]. Thus, mammalian olfactory neurons are certainly more sensitive to variations in odorant concentration, whereas insect olfactory neurons are very efficient at rapidly detecting the presence or absence of an odorant. It seems that the choice of the olfactory perception system in insects could be explained by the mechanical advantages of the ionotropic pathway for their perceptual needs. This point of view is supported when we consider olfaction in tobacco plants whose perception speed is of the order of several hours [56].

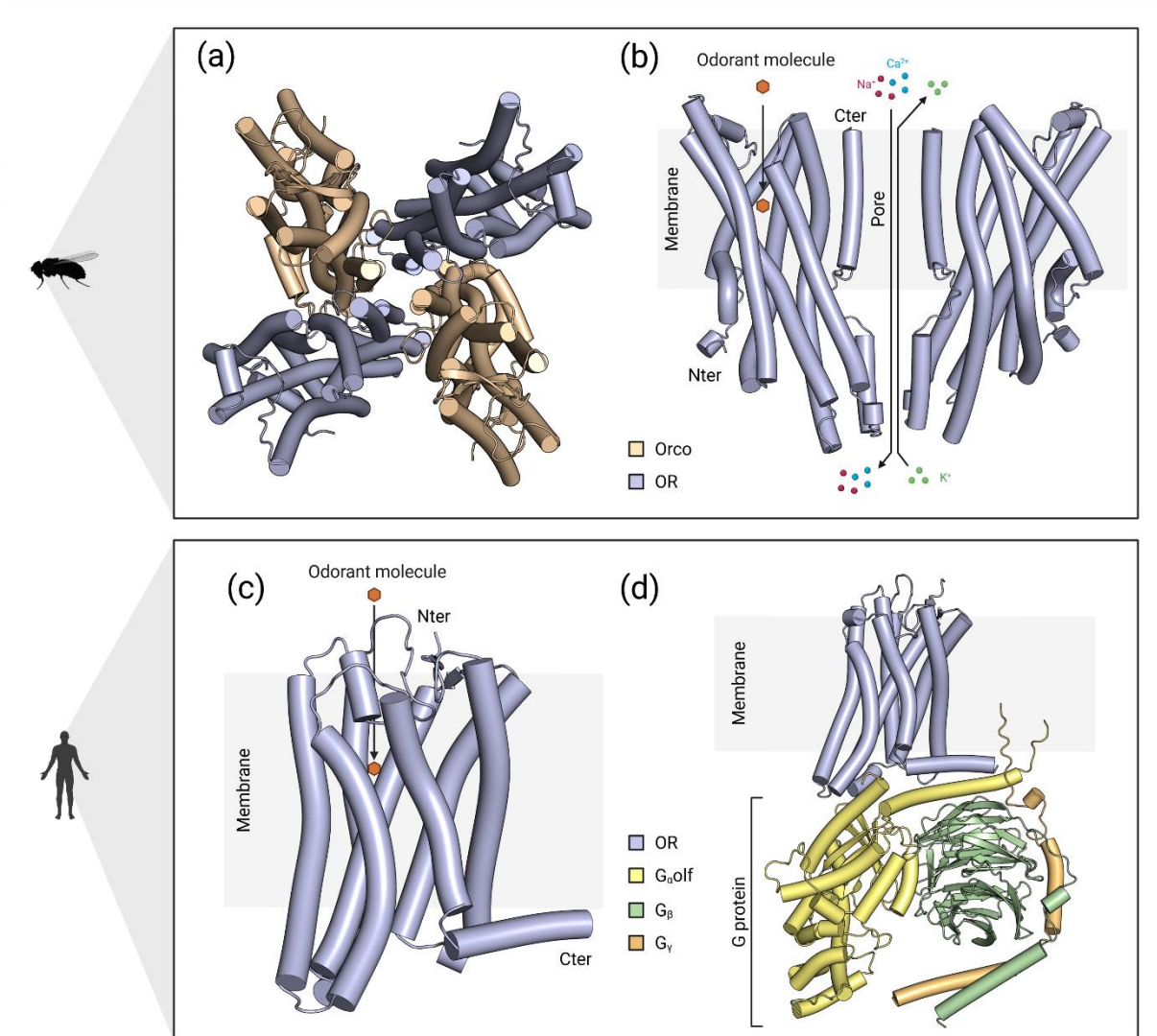


Figure 9: (a) Top view of a DmelOrco and DmelOR5 complex ion channel model (flesh-colored and blue, respectively) with a suggested stoichiometry of 2:2. (b) Side view of the channel, centered on the OR subunits in their membranes (gray). Insect OR and Orcos subunits share the same fold consisting of 7 transmembrane helices with the N terminus on the cytosolic side and the C terminus on the extracellular side. An odorant molecule binds to the OR subunits and, by an allosteric process, opens the channel pore, allowing the passage of Na⁺, Ca²⁺ and K⁺ ions. (c) Side view of a model of the human OR hOR1A1 (blue). ORs are GPCRs consisting of 7 transmembrane helices, with the N-terminus on the extracellular side and the C-terminus on the cytosolic side. A scent molecule binds in a cleft buried in the receptor and, through an allosteric process, allows the G protein complex to bind to the cytosolic side of the receptor. (d) Model of hOR1A1 (blue) complexed with its G protein composed of G_αolf, G_β and G_γ (yellow, green, and orange respectively). The models were generated with AlphaFold2 multimer version [72].

Evolution of gene repertoires

Olfaction is extremely plastic, with the gene repertoires of ORs undergoing intense rearrangements during evolution. There have been both intense gene duplications and intense gene deletions (pseudogenizations) in a process called life-and-death evolution. In general, ORs repertoires have expanded and diversified, from the ionotropic ORs of invertebrates to the GPCR ORs of amphibians, fish, birds, and mammals. However, there is considerable variation between species, not least because of the need for organisms to adapt to the chemical space of the biotope in which they are found. The large number of pseudogenes possessed by some species may reflect the extent to which each species has relied on olfaction during recent evolution time, because a reduced need means a relaxation of the selective pressure that normally inactivates gene mutations [27].

Ionotropic OR genes.

Crustaceans possess ionotropic glutamate receptors (IRs) which are found in olfactory perception in insects. In insects, IRs are composed of up to 3 subunits formed by a receptor recognizing the odor, and between 1 and 2 subunits of the co-receptors IR25a, IR8a, and IR76b. These receptors are found in the olfactory neurons of animals of the genus Protostomia, notably the American lobster. Crustaceans being the closest organisms to insects, it is likely that IRs are the ancestors of the ORs of insects [75].

Drosophila melanogaster (*D. melanogaster*) has an olfactory system composed of 60 ORs subtypes that associate with a highly conserved olfactory receptor co-receptor (Orco) in a stoichiometry not yet known. Ancestral to *D. melanogaster*, the order Archaeognatha is the most basal group of insects with ORs. Belonging to this order and unlike more evolved species, the jumping bristletail species *Machilis hrabei* (*M. hrabei*) has only 5 ORs and does not express any Orco. The structure of MhOR5, was solved in 2021 by Marmol *et al.* in its free or agonist-bound form [76]. The receptor is a functional homotetramer and has been proposed as the ancestor of the Orco/OR system. The need to recognize an increasingly broad spectrum of odors would have driven evolution to create the modular system that is the Orco/OR, to increase the number of ORs, and to favor a hybrid system of narrow and broad ORs [77, 78].

Metabotropic OR genes.

Metabotropic ORs are few in most basal fishes, as evidenced by sharks, which possess only one family of ORs [21]. Teleost fish also have few but diverse OR genes, covering a large part of the olfactory chemical space [68]. The big "winners" are rats with 6% of their genome dedicated to olfaction and about ~1200 genes, and elephants with ~2000 genes and 2200 pseudogenes. This impressive diversity of genes could be explained by the importance of olfaction for elephants, for food, social communication, and reproduction. The number of genes would help in the recognition of odorants with a similar structure, and the number of genes would determine the resolution of the olfactory world and not the sensitivity to a given smell [27, 79]. To enable this extreme increase in the number and diversity of ORs while maintaining functionality and expression at the membrane, Ikegami *et al.* (2020) speculate that olfactory sensory neurons would have had to develop multiple OR-specific chaperone proteins [80]. Following their reasoning, it is interesting to draw a parallel between the chaperone role of Orco in insects and the different mammalian chaperones, which are different responses to the same need for rapid adaptation of ORs repertoires.

Primates have fewer OR genes than rodents (humans have ~400 OR genes and 400 pseudogenes, like chimpanzees), and a higher percentage of pseudogenes. However, humans retain genes in nearly all mouse OR subfamily. This suggests that humans can smell a similar number of odors, but that subtle differences between closely related molecules may go unnoticed. Ache *et Young* (2005) [27] suggest that the loss of part of the repertoire coincides with the onset of trichromatic vision. Niimura *et al.* (2018) nuances this view stating that "the rate of gene loss can be retraced to the haplorrhine basal branch of primates, which coincided with development of acute vision", and conclude that multiple factors may have contributed to the olfactory degeneration of primates, like phylogeny, anatomy, and habitat evolutions [81]. Interestingly, there is also a notable or total decrease in the available OR for species returning to the sea in both reptiles like sea snakes [32] or mammals like dolphins [82], pointing to a loss of olfactory perception needs for these species.

Computational methods applied to olfaction

Challenges raised by olfaction

To better understand the mechanisms of olfactory perception, several challenges stand out. If we understood the combinatorial code of odors, we would be able to predict the odor of a molecule from its chemical structure, just as we can predict the color induced by a photon by knowing its wavelength in vision [63, 83].

This aim is more delicate than it seems. Indeed, some odors that share functional groups seem to smell the same, such as esters that smell fruity or floral and thiols that smell rotten. But this is more the exception than the rule. Odor of a molecule can vary according to its concentration. Some molecules that are chemically very similar have different odors; a phenomenon called an "activity cliff". This is the case of enantiomeric compounds such as (-)-carvone which has a caraway smell and (+)-carvone which has a fresh mint smell [84]. Conversely, completely different molecules can smell the same, such as muscone and androstenol [85]. To make matters even more complex, olfactory sensory neurons can be excitatory or inhibitory, and the smell of a molecule can vary with its concentration, adding additional degrees of freedom to the complexity of the combinatorial code [27].

Olfactory genes are extremely diverse and vary greatly between species, both in quantity and specificity. However, individual variation is also important. Individuals may have different repertoires of pseudogenes, different polymorphisms, or numbers of copies of genes. Indeed, a single genetic modification (i.e., single nucleotide polymorphism or SNP) can change the perception of an odorant, and a decrease in receptor function can be linked to a decrease in the intensity of perception. A better understanding of these intraspecific variations would improve our understanding of the relationship between the activity of an OR and the resulting olfactory perception [86]. With the 400 human ORs and their polymorphisms facing the 10000 to 40 billion possible odorants [12], deciphering the combinatorial code means solving a complex many-to-many relationship problem.

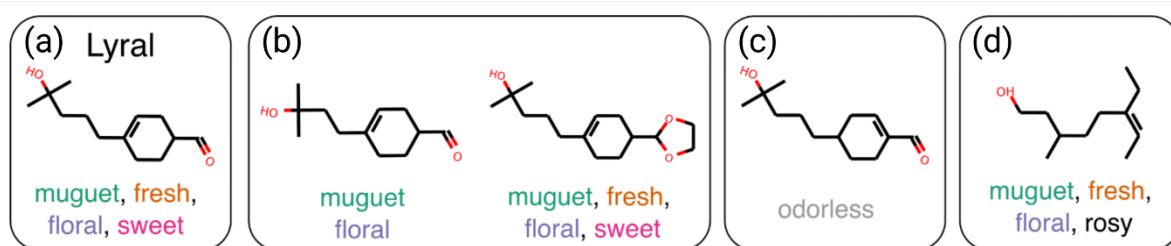


Figure 10: Molecules with similar structures do not necessarily have similar odors. (a) Example of Lyrals. (b) Molecules with similar structures may have similar descriptors. (c) But some molecules with minute variations have different semantic descriptors. (d) On the contrary, very different molecules can have several semantic descriptors in common. From Sanchez-Lengeling *et al.* (2019) [87].

Even if the number of odorant-receptor pairs is overwhelming, the more we know about them, the better our understanding of the combinatorial code becomes. Unfortunately, many of the known ORs are orphans (we do not know any ligand to the receptor in question). Multiple *in vitro* and *in vivo* deorphanization efforts have been done, with drawbacks and benefits for each approach. *In vivo* approaches have the merit of recreating the real conditions of an olfactory neuron, but they operate at low throughput. On the contrary, *in vitro* approaches allow high throughput screening, but face problems of expression, addressing the receptor to the membrane and be far from the real conditions of the olfactory perception [80]. The amount of data available currently remains critically low [88, 89].

Available structural information on ORs is also limited. In insects, the structure of the Orco homotetramer was resolved by cryo-electron microscopy in 2018 [71], followed by the structure of *MhOR5* in 2021 [76]. Given the low sequence identity of ORs to Orcos or to basal receptors such as *MhOR5*, an experimental structure of the Orco/OR complex would be extremely useful. On the vertebrate side, there is currently no structure of an OR. Furthermore, the mammalian ORs, although class A GPCRs, all have less than 20% sequence identity with the known experimental structures in this family [80].

This complexity has naturally driven scientists toward the promises of *in silico* approaches. Machine learning can use available information about a problem to find hidden correlations in the data set, leading to predictions and sometimes better understanding. Molecular modeling, through a wide range of structural techniques, can provide new insights into the relationships between sequences, structures and functions of the system considered.

Machine learning in olfaction

Machine learning (ML) has been used to help in various olfaction problems. The use of ML coupled with *in vitro* tests to find new ligand-receptor pairs has proven its effectiveness in the case of olfaction, both for invertebrate [90, 91], and vertebrate ORs [89, 92–95]. ML has also been used for various problems (non-exhaustive) such as predicting the odor perception threshold [96], the expression of ORs on the surface of olfactory neurons on the basis of receptor sequence [80] or predicting whether a chemical compound is an odorant [12].

But the holy grail of the use of machine learning in olfaction, would be to succeed in predicting and understanding with precision the smell of a molecule starting from its chemical structure (the structure-odor relationships or SOR). Several attempts have been made in this direction [97], but the applicability domain or performance of the models were limited. It appeared that it was easier to predict intensity or hedonicity than odor [98–100]. The quality and quantity of the data was questioned. To overcome these shortcomings, the DREAM project [101] assembled a dataset of 480 molecules with various structural and perceptual properties on a panel of 55 people. The following year, Keller *et al.* (2017) [102] proposed models that significantly improved performance for the prediction of intensity and hedonicity, as well as 8 semantic descriptors out of the 19 considered. Very recently, Lee *et al.* (2022) [103] may have pushed the limit near to the glass ceiling. Based on the principle that the chemical descriptors usually used are not suitable for the discontinuities between the chemical space and the perceptual space of odors (figure 10), they use a Message Passing Neural Network (MPNN) where each odorant is represented in graph. The predictive performance of the model is comparable to that of a panel of untrained individuals [103].

Molecular modeling in olfaction

Molecular modeling is the structural and functional study of matter through the creation of computer models. Molecular modeling encompasses a wide range of fast evolving techniques. We will focus here on some of the techniques used in the study of the molecular basis of olfactory perception that give insights about the sequence-structure-function relationships of ORs and their ligands. The initial step is to create a model of the system at atomic resolution. If experimental structures are not available, there are different techniques to generate a model of the protein structure. This first step enable (non-exhaustive): localization of the orthosteric

or allosteric sites of the receptor and characterize their chemical properties, finding new ligand-receptor pairs through docking and virtual screening or study the energy and dynamics of the receptor through molecular dynamic simulations (MD).

Creating a model

As previously mentioned, the difficulty of obtaining an experimental structure requires sometimes the creation of a model of the receptor. Multiple approaches are available and used depending on the availability of experimental reference structures.

Ab initio modeling

At the time of the discovery of ORs genes by Lina and Axel (1991) [104], there were no crystallographic structure of GPCRs available. In 2000, researchers created the first OR model using ab initio modeling techniques, where a model is created from scratch [105] coupled with coarse grain molecular dynamics (CG-MD) to refine the model, guided only by the rhodopsin 7.5 Å electron density map that was available at this time. ROSETTA method has been used to model *de novo* GPCRs with relatively good accuracy [106]. Coevolutionary sequence analysis is a third *de novo* approach that identifies evolutionary couplings between amino acids and thus predicts the physical contact between them. In addition to providing evolutionary information per se, these structural constraints can be used to generate a model [107]. This approach has been used in 2015 to propose a model of the insect ORs [108].

Threading methods

Threading methods take advantage of the available structural information by assuming that the number of different folds is small (estimated to be less than 10000 in nature [109, 110], and that the Protein Data Bank (PDB) already contains most of the possible folds. The query sequence is compared to the experimental structures looking for similar folds, even if there is no evolutionary relationship. The I-TASSER workflow is a composite approach that couples both threading methods with evolutionary homology search for better prediction [111, 112]. This method was recently used to model the mouse TAAR9 receptor [113] using the GPCR-I-TASSER composite web server [114].

Homology modelling

Homology/comparative modeling assumes that the amino acid sequence dictates the three-dimensional structure of proteins, and that folding is more conserved than sequence during evolution (a perfect example of this is the extreme diversity of OR genes with conserved folding, both in vertebrates and invertebrates). Therefore, similar sequences are likely to share similar structures [115].

The homology modeling process starts with the identification of an appropriate experimental structure (called template) to model the target protein. Sequence similarity, but also quality (resolution), or the state of the protein (ligand bound, active or inactive) are important parameters to consider when choosing the experimental structure. Multi-template modeling may in some case better the quality of the final structure [116]. The second step is to correct the alignment between the template sequence and the target sequence if necessary. This step is crucial because, for example, a single residue missing in an alpha helix region can rotate the residues of the rest of the helix. There are then different methods to generate the 3D model. Rigid-body assembly split the model into basic conserved regions (core regions, loops, side-chains) taken from the template that are fitted together [117]. Segment mapping method selects specific atomic positions in the template as the primary positions for the modeling process, chosen based on sequence identity, geometry, and energy on a known database. The spatial constraint method creates a set of structural constraints based on the template structure. These constraints are then applied to the targeted structure based on the alignment [118]. The artificial evolution method uses the rigid body assembly method and stepwise template evolutionary mutations until the template sequence is identical to the target sequence [115].

Loops are often not or partially defined in experimental structures due to their high structural and conformational variability [119]. This is a major problem, as they often perform a critical function in receptors, modulating for example specificity or promiscuity in aminergic GPCRs or vertebrate ORs [120, 121]. This problem carries over to homology modeling where modeling loops is not a trivial task. Because of their flexibility, but also because of their high variability in length, sequence and structure from one protein to another sharing the same fold, they deserve special care [122]. To improve the quality of the loops, two techniques are possible. The first one is based on a similarity search with a database of loops, to guide the modeling process. The second is an *ab initio* method that samples the conformations of the considered

loop [122, 123]. The side chains are then optimized by searching a rotamer library and selected using a score function. The model is finally optimized by energy minimization using force fields from molecular mechanics and evaluated with diverse tools.

Homology modeling has been extensively used in olfaction since the 2000, even if sequence identity between available templates and ORs are less than 20% [105]. In 2013, Charlier *et al.* (2013) propose a homology workflow specific to ORs using multi-templates [124]. Groups have attempted to attack the problem of low sequence identity through the use of experimentally guided alignment as described by Gelis *et al.* (2011) [125] or De March *et al.* (2015) [126] and many attempts have followed [127–131].

Model evaluation and the arrival of deep learning

Various methods are available to assess the accuracy of the model and determine its validity for the application. A basic stereochemistry check is required, such as the Ramachandran plot [115]. Most web servers and software offer scoring functions to evaluate the quality of the generated models [127]. But a true evaluation of modeling performance between approaches can ultimately only be done by comparing the model with its experimental structure. The CASP (critical assessment of the protein prediction) dataset was designed with this idea in mind [128]. Every two years, CASP identify an array of experimental structure just solved and not yet published. The respective sequences of these proteins are then given to the community which has three weeks to submit models for evaluation. The competition is divided in 2 parts: *ab initio* methods and comparative modeling.

The most effective method was generally comparative modeling, provided that the available experimental structures were close enough to the target. However, for the thirteenth CASP in 2018 [128], deep learning techniques were used to drastically increase the quality of the structures generated by *ab initio*, reaching the performance level of homology modeling [129]. For the next edition, a redesign of the AlphaFold deep learning model named AlphaFold2 (AF2) gave experimental accuracy on more than 2/3 of the targets, making it obsolete to classify proteins by difficulty with respect to their degree of similarity to other structures already available [130].

AF2 is a machine learning approach that integrates both structural and evolutionary data into a deep learning model. Its predictions are accurate both at the structural domain and at the amino acid side chain orientation level [131]. AF2's predicted local-distance difference test (pLDDT)

metric for assessing the quality of the model has already proven its usefulness in predicting the flexibility of different areas of the modeled protein, in accordance with molecular dynamics [132]. AF2 makes its prediction with only the sequence information of the protein of interest, plus a multi-sequence alignment (MSA) of homologous sequences coupled with structural information from available experimental models. These inputs are processed by several layers of a new neural network block called Evoformer. The information contained in the MSA becomes a matrix of M sequences x N residues, the structural information becomes a matrix of N residues x N residues representing the residue pairs of the target protein. Next comes the structure module which creates an explicit 3D model with the rotation and translation of each residue of the protein, which iteratively converges to the final result after a gradient descent relaxation phase using the AMBER force field [131].

Baek *et al.* (2021) quickly followed by developing Rosetta-Fold, a deep neural network with similar performance to AF2 and capable of predicting protein-protein complexes [133]. Since then, Evans *et al.* (2021) has also released a multimeric version of their model with state-of-the-art protein-protein docking accuracy [134]. The Universal Protein Resource (UniProt) database had 190 million protein sequences in 2019 [135], while the PDB has only 180,000 structures for about 55,000 distinct proteins [136]. Noting the near-experimental accuracy of AF2, Varadi *et al.* (2021) sought to bridge the gap between structural and sequence information available by creating a database of AF2-predicted structures. Today, more than 360000 high quality models are available for multiple proteomes of model organisms and the aim is to cover most UniProt sequences [137]. AF2 ORs models are already used by different group [138].

Docking and screening *in silico*

Experimental structure or modeling of a receptor paves the way for structural studies of ligand/receptor interactions and structure-based drug discovery (SBDD). Depending on the quality of the model, docking techniques allow to obtain a prediction of the binding mode and the approximate evaluation of the binding energy of a ligand in a target cavity. On a larger scale, high throughput *in silico* screening (virtual screening) can be used to rapidly select potential ligands on a target from databases of several million compounds. However, it is important to keep in mind that docking score functions cannot estimate the affinity of a ligand with high accuracy. A virtual screening in which 5-10% of the molecules are detected as active in *in vitro* assays should be considered successful [139].

Docking is a two-step process: 1. the sampling step generates a large number of conformations and orientations of the ligand within the protein binding site. 2. these positions are evaluated by a score function to identify the best ligand pose. Multiple approximations are performed to speed up the calculations in virtual screening. The ligand is considered flexible, but the receptor is generally fixed (although flexible docking is possible). Multiple scoring functions have been developed, but most only consider receptor-ligands interactions. The contributions of desolvation and entropy to the binding energy are important but generally neglected [140, 141]. There are several ways to improve the quality of docking, such as considering part of the receptor as flexible, or screening on a set of its conformations (named ensemble docking). The consensus between several score functions improves the results, as well as considering the water molecules in the active site [139]. Despite all these limitations, docking and structure-based virtual screening are important techniques in the arsenal of molecular modeling, not least because of their ability to propose new scaffolding molecules, which is difficult to do with machine learning techniques that are limited by to the applicability domain of the models [142]. In the field of olfaction, several studies have successfully used this approach to understand odorant-OR interactions [105, 113, 125, 143, 144] or discover new odorant-OR pairs [145], despite the inherent difficulty of using non-experimental models [146].

Molecular dynamic simulations

To understand running, it is better to see a movie than a picture of a runner. Since proteins are essentially molecular machines, their functions are intrinsically linked to their dynamic processes. Molecular dynamics (MD) simulations, like a movie, can bring experimentally resolved or model-generated protein structures to life with the promise of understanding the relationship between their structure, dynamics and functions [147].

Molecular dynamics starts with the preparation of a system. To model the physiological environment of an OR, one can model the protein embedded in a lipid bilayer and solubilized in a box of water molecules [148]. Once we know the position of each atom in the finite system, we can then calculate the force that it exerts on all the others, and vice versa. The so-called "classical" MD uses Newtonian mechanics to calculate and update, step by step, the position and the velocities applied on each atom of the system. The concatenation of each step forms a trajectory describing the three-dimensional evolution of the system during the simulation time [149].

The duration of time step should be as small as the fastest motion of the system, but as long as possible to decrease the computational cost and maximize the observation time. In the case of a classical all-atom MD, this time is of the order of femtoseconds, limited by the vibrational movements of the hydrogen bonds. A trajectory of several microseconds thus counts several billion frames, which is computationally expensive. This is both a strength, as few experimental techniques can observe motions with such fine temporal resolution, and a weakness, as many crucial protein motions take place on, microseconds to millisecond and higher time scales: For example, side chain flipping takes 10-100 ns, ligand binding to a GPCR takes a few ns to a few μ s, GPCR activation takes a few μ s to a few ms, GPCR/GPCR dimerization takes a few ms, and G protein or ligand release can take seconds to minutes [150].

The calculation of the energy and velocities of atoms is based on so-called mechanical force field models. Each atom is approximated by a sphere, each bond and angle by a spring. This last approximation implies that there can be no broken bonds. Depending on the type of atom and bond, the force field defines the constant of the forces applied to the atoms. These parameters are generated from experimental spectroscopic data and quantum mechanical calculations. Typically considered are non-bonded interactions with electrostatic interactions and van der Waals forces (as Coulomb and Lennard Jones potentials), and bonded interactions with bond lengths and torsion angles (as harmonic potentials) and dihedral torsions (as periodic potentials). The sum of all these terms represents the potential energy of the system at a given time [149].

Given enough time, a MD can sample all possible conformations (or microstates) of a system across its degrees of freedom. The relative probability of finding the system in a specific conformation is linked to its free energy. We can then conceptualize a free energy landscape of N dimensions equal to the degree of freedom of the system, where the low energy zones are highly populated (stable microstates), and conversely the high energy zones are lowly populated (unstable microstates). The MD simulation therefore samples the shapes of the free energy landscape across simulation time (figure 11a, [151]).

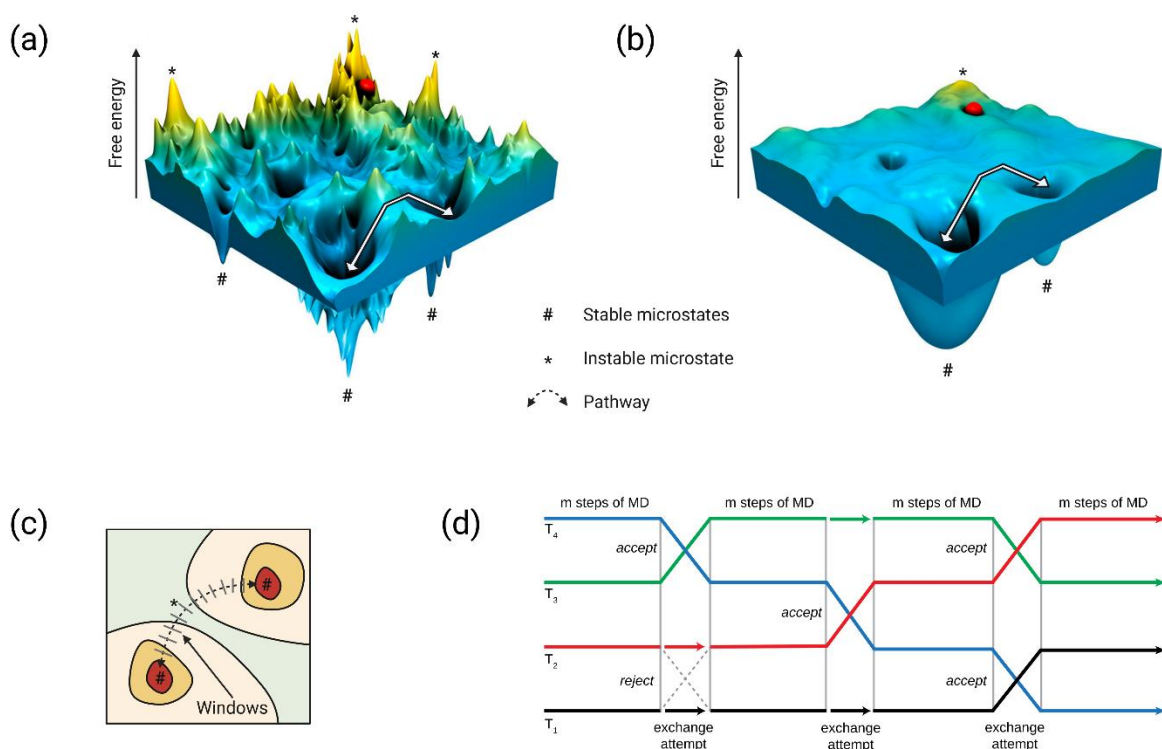


Figure 11: (a) Representation of the free energy landscape and its coarse-grained smoothed version (b). Regions of low or high free energy are colored blue or yellow, respectively. The smoothed CG potential allows efficient sampling without falling into local minima. Adapted from Kmiecik *et al.* (2018) [152]. (c) Umbrella sampling scheme. An artificial bias is applied across a reaction coordinate to sample the energy barrier between two stable microstates. The trajectory is partitioned into overlapping sampling windows. Adapted from Kästner (2010) [153]. (d) Replica exchange molecular dynamics (REMD) scheme. Multiple parallel MD trajectories are launched at different temperatures. At each given time interval, configurations are tentatively exchanged based on a metropolis criterion. Adapted from Mori *et al.* (2016) [154].

The fine temporality of the MD simulations allows us to describe the paths that separate one stable conformation from another and contains significant information about the molecular processes (Figure 11a). In reality, the observation of unstable/high-energy intermediate states between these stable conformations is rare and the ideal case where the ergodic hypothesis is validated, i.e., when the simulation has explored the entire free energy landscape/phase space, is never satisfied due to lack of simulation time. This has an impact on the quality of the conclusions that can be drawn from the MD simulations. The need to observe changes on time scales inaccessible to current computing power has led researchers to develop a series of methods to accelerate the sampling of the free energy landscape [155]. These techniques can

be divided into two classes: the first is based on adding a bias potential along predefined collective variables (CVs), the second does not require predefined CVs. A CV is a coarse-grained description of a system, used instead of atomic coordinates to describe a particular process of interest [156].

If the description of the microstates separating the initial and final states is not important, alchemical free energy perturbation techniques (FEP) can be used to quantify the difference in free energy between 2 states. These techniques are mainly used to obtain the relative binding energy between 2 simple perturbations. The CVs are for example the physicochemical properties of the first ligand which will be progressively transformed to become the second ligand [157]. Another example is umbrella sampling which biases the system by applying an artificial force along one or more CVs to force the system from one thermodynamic state to another (figure 11c, [153]).

In the case where there is no a priori knowledge of the system and the movements to consider, there are accelerated sampling techniques that do not require the definition of CVs. An artificial increase of the system temperature can be used as a means to smooth the free energy landscape and thus decrease the size of the energy barriers. The system is then cooled and can reach a new local minimum. The replica exchange molecular dynamics (REMD) and its variations is an appropriate technique for exploratory studies, despite its high computational cost (figure 11d, [154]). Another way to increase the sampling speed is to decrease the number of degrees of freedom of the system. We can define a force field where each particle corresponds to 2 or 4 atoms as in the Martini3 case [158]. This results in an increase of the sampling speed of several orders of magnitude and a smoothing of the free energy landscape, at the cost of a decrease of the resolution of the system (figure 3b).

The information that MD can provide is varied. One can study the conformational flexibility and stability of proteins. One of the first (and perhaps most intuitive) historical uses of MD was indeed to explore the flexibility of the first experimental structure of a protein [159]. MD can also be used to improve the quality of a model [144, 160], although this approach has shown its limitations [161].

One of the main advantages of MD is the total control the user has over the parameters of his system. One can thus disturb the system in a certain way and observe its reaction to better understand it. A ligand can be placed or removed from a protein structure to observe induced rearrangements and interactions [69]. An amino acid can be mutated ([162], chapter 2) or protonated/deprotonated [163] to predict its impact on protein function. Although residence

time is known to be more predictive of drug efficacy than affinity [164], ligand unbinding occurs on a time scale (ms to s) inaccessible to conventional MD simulations. A possible way to observe unbinding is to force the ligand out of the binding site through an artificial force [165] and have been applied in olfaction [166]. It is also possible to study ion channel permeation by creating an artificial transmembrane potential by artificially creating a system with asymmetric ion distributions [167]. Finally, dynamic processes can be observed without adding bias to the system. It is for example possible, with sufficient computing power, to observe the activation process of a receptor ([168], chapter 2), or the diffusion of a ligand through its protein to its binding site [169], chapter 1 and 3b).

Applied to the study of olfaction, MD is a valuable tool that can guide *in vitro* experiments such as directed mutagenesis and functional assays. This multidisciplinary approach allows to predict receptor stability [80], to identify residues important in the allosteric mechanisms of ORs ([170], chapter 1, 2, 3), to predict the impact of a polymorphism on receptor activation ([170], chapter 2), to characterize an orthosteric site ([144, 171], chapter 1, 3a) or allosteric ([69], chapter 1, 2, 3b), to guide the discovery of novel odorant-receptor pairs [144] and to describe the diffusion channels of ligands to their binding sites ([138, 172], chapter 1 and 3b).

Contents

This thesis is focused on obtaining a better understanding of the molecular basis of olfactory perception at the level of the first component of the olfactory system: the receptor. The general objective is to clarify the sequence-structure-function relationships of insect and mammalian ORs. Several practical objectives have been set: 1. to predict the structure of a receptor from its sequence. Given its structure, 2. to predict the binding sites of a receptor. 3. to predict the impact of a ligand and/or a mutation on the function of a receptor. 4. to evaluate the functional significance of specific domains on the function of a receptor.

My contribution to olfaction research is articulated in 3 chapters in which I have placed the scientific papers to which I contributed. The 4 papers are based on the generation of a model of the targeted receptor, the identification of key residues in the considered function thanks to structural information obtained by molecular dynamics, and then *in vitro* functional tests to confirm or refute the *in silico* observations. In the first chapter, we attempt to identify and characterize the diffusion channel of an agonist to its binding site for the olfactory coreceptor

(Orco) in *D. melanogaster*. The second chapter discusses the functional impact that a mutation in the human OR TAAR5 gene can have. The third and last chapter focuses on the importance of two non-ordered structures on the function and recognition spectrum of mammalian ORs: the extracellular loop ECL2 and ECL3, in mouse and human ORs respectively.

In addition to these papers, I was able to contribute to several parallel projects. The methodology developed in the OR reclassification project presented in the conclusion was applied to the TAS2R16 family of bitter taste receptors and led to the discovery of a relationship between cavity hydrophobicity and the recognition spectrum of this receptor family. I also provided structural information and physicochemical descriptors of all human and murine OR cavities to guide a machine learning model to discover new odorant-receptor pairs. Finally, I participated in the data curation of a project on COVID19 based on an online self-administered questionnaire identifying self-identified symptoms of ageusia and anosmia as early indicators of hospital overload.

This thesis illustrates the synergistic effect of combined *in silico* and *in vitro* methods to further understand the molecular mechanisms of ORs and more globally of metabotropic and ionotropic membrane receptors.

References

1. Mashour GA, Alkire MT (2013) Evolution of consciousness: Phylogeny, ontogeny, and emergence from general anesthesia. *Proceedings of the National Academy of Sciences* 110:10357–10364. <https://doi.org/10.1073/pnas.1301188110>
2. Niven JE, Laughlin SB (2008) Energy limitation as a selective pressure on the evolution of sensory systems. *J Exp Biol* 211:1792–1804. <https://doi.org/10.1242/jeb.017574>
3. Swain A, Hoffman T, Leyba K, Fagan WF (2021) Exploring the Evolution of Perception: An Agent-Based Approach. *Frontiers in Ecology and Evolution* 9:
4. Kandel ER, Koester JD, Mack SH, Siegelbaum SA (2021) *Principles of Neural Science, Sixth Edition, 6th edition*. McGraw Hill / Medical
5. Gadhvi M, Waseem M (2022) *Physiology, Sensory System*. In: StatPearls. StatPearls Publishing, Treasure Island (FL)
6. Kumar A, Tauxe GM, Perry S, et al (2020) Contributions of the conserved insect carbon dioxide receptor subunits to odor detection. *Cell Rep* 31:107510. <https://doi.org/10.1016/j.celrep.2020.03.074>

7. Darras K, Yusti E, Knorr A, et al (2022) Sampling flying bats with thermal and near-infrared imaging and ultrasound recording: hardware and workflow for bat point counts. *F1000Res* 10:189. <https://doi.org/10.12688/f1000research.51195.2>
8. Cochran WW, Mouritsen H, Wikelski M (2004) Migrating songbirds recalibrate their magnetic compass daily from twilight cues. *Science* 304:405–408. <https://doi.org/10.1126/science.1095844>
9. Chittka L, Brockmann A (2005) Perception Space—The Final Frontier. *PLoS Biol* 3:e137. <https://doi.org/10.1371/journal.pbio.0030137>
10. Dangles O, Irschick D, Chittka L, Casas J (2009) Variability in sensory ecology: expanding the bridge between physiology and evolutionary biology. *Q Rev Biol* 84:51–74. <https://doi.org/10.1086/596463>
11. Hettinger TP (2011) Olfaction is a chemical sense, not a spectral sense. *Proc Natl Acad Sci U S A* 108:E349. <https://doi.org/10.1073/pnas.1103992108>
12. Mayhew EJ, Arayata CJ, Gerkin RC, et al (2022) Transport features predict if a molecule is odorous. *Proc Natl Acad Sci U S A* 119:e2116576119. <https://doi.org/10.1073/pnas.2116576119>
13. Ma M (2012) Odor and pheromone sensing via chemoreceptors. *Adv Exp Med Biol* 739:93–106. https://doi.org/10.1007/978-1-4614-1704-0_6
14. Hildebrand JG, Shepherd GM (1997) Mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annu Rev Neurosci* 20:595–631. <https://doi.org/10.1146/annurev.neuro.20.1.595>
15. Sourjik V, Berg HC (2004) Functional interactions between receptors in bacterial chemotaxis. *Nature* 428:437–441. <https://doi.org/10.1038/nature02406>
16. Palková Z, Janderová B, Gabriel J, et al (1997) Ammonia mediates communication between yeast colonies. *Nature* 390:532–536. <https://doi.org/10.1038/37398>
17. Hu L (2022) Integration of multiple volatile cues into plant defense responses. *New Phytol* 233:618–623. <https://doi.org/10.1111/nph.17724>
18. Prasad BC, Reed RR (1999) Chemosensation: molecular mechanisms in worms and mammals. *Trends in Genetics* 15:150–153. [https://doi.org/10.1016/S0168-9525\(99\)01695-9](https://doi.org/10.1016/S0168-9525(99)01695-9)
19. Harzsch S, Krieger J (2018) Crustacean olfactory systems: A comparative review and a crustacean perspective on olfaction in insects. *Prog Neurobiol* 161:23–60. <https://doi.org/10.1016/j.pneurobio.2017.11.005>
20. Sokolinskaya EL, Kolesov DV, Lukyanov KA, Bogdanov AM (2020) Molecular Principles of Insect Chemoreception. *Acta Naturae* 12:81–91. <https://doi.org/10.32607/actanaturae.11038>

21. Sharma K, Syed AS, Ferrando S, et al (2019) The Chemosensory Receptor Repertoire of a True Shark Is Dominated by a Single Olfactory Receptor Family. *Genome Biology and Evolution* 11:398–405. <https://doi.org/10.1093/gbe/evz002>
22. Abankwah V, Deeming DC, Pike TW (2020) Avian Olfaction: A Review of the Recent Literature. *Comparative Cognition & Behavior Reviews* 15:149–161
23. Gomez-Marin A, Duistermars B, Frye M, Louis M (2010) Mechanisms of odor-tracking: multiple sensors for enhanced perception and behavior. *Frontiers in Cellular Neuroscience* 4:
24. Wheeler GN, Brändli AW (2009) Simple vertebrate models for chemical genetics and drug discovery screens: Lessons from zebrafish and *Xenopus*. *Developmental Dynamics* 238:1287–1308. <https://doi.org/10.1002/dvdy.21967>
25. Signal transduction in bacterial chemotaxis - PubMed. <https://pubmed.ncbi.nlm.nih.gov/16369945/>. Accessed 29 Sep 2022
26. Shapiro JA (2021) All living cells are cognitive. *Biochem Biophys Res Commun* 564:134–149. <https://doi.org/10.1016/j.bbrc.2020.08.120>
27. Ache BW, Young JM (2005) Olfaction: Diverse Species, Conserved Principles. *Neuron* 48:417–430. <https://doi.org/10.1016/j.neuron.2005.10.022>
28. Yohe LR, Brand P (2018) Evolutionary ecology of chemosensation and its role in sensory drive. *Curr Zool* 64:525–533. <https://doi.org/10.1093/cz/zoy048>
29. Suen JLK, Yeung AWK, Wu EX, et al (2021) Effective Connectivity in the Human Brain for Sour Taste, Retronasal Smell, and Combined Flavour. *Foods* 10:2034. <https://doi.org/10.3390/foods10092034>
30. Processing of Intraoral Olfactory and Gustatory Signals in the Gustatory Cortex of Awake Rats | Journal of Neuroscience. <https://www.jneurosci.org/content/37/2/244?etoc=>. Accessed 29 Sep 2022
31. Clark L, Hagelin J, Werner S (2015) Chapter 7 - The Chemical Senses in Birds. In: Scanes CG (ed) *Sturkie's Avian Physiology (Sixth Edition)*. Academic Press, San Diego, pp 89–111
32. Kishida T, Go Y, Tatsumoto S, et al (2019) Loss of olfaction in sea snakes provides new perspectives on the aquatic adaptation of amniotes. *Proceedings of the Royal Society B: Biological Sciences* 286:20191828. <https://doi.org/10.1098/rspb.2019.1828>
33. Function of Oscillatory Tongue-Flicks in Snakes: Insights from Kinematics of Tongue-Flicking in the Banded Water Snake (*Nerodia fasciata*) | Chemical Senses | Oxford Academic. <https://academic.oup.com/chemse/article/37/9/883/327681>. Accessed 29 Sep 2022
34. Freitag J, Ludwig G, Andreini I, et al (1998) Olfactory receptors in aquatic and terrestrial vertebrates. *J Comp Physiol A* 183:635–650. <https://doi.org/10.1007/s003590050287>

35. Mollo E, Fontana A, Roussis V, et al (2014) Sensing marine biomolecules: smell, taste, and the evolutionary transition from aquatic to terrestrial life. *Front Chem* 2:92. <https://doi.org/10.3389/fchem.2014.00092>
36. Mollo E, Garson MJ, Polese G, et al (2017) Taste and smell in aquatic and terrestrial environments. *Nat Prod Rep* 34:496–513. <https://doi.org/10.1039/c7np00008a>
37. Carbone M, Gavagnin M, Haber M, et al (2013) Packaging and Delivery of Chemical Weapons: A Defensive Trojan Horse Stratagem in Chromodorid Nudibranchs. *PLoS One* 8:e62075. <https://doi.org/10.1371/journal.pone.0062075>
38. Frøkjær-Jensen C, Ailion M, Lockery SR (2008) Ammonium-Acetate Is Sensed by Gustatory and Olfactory Neurons in *Caenorhabditis elegans*. *PLoS One* 3:e2467. <https://doi.org/10.1371/journal.pone.0002467>
39. Taste and Smell: A Unifying Chemosensory Theory | *The Quarterly Review of Biology*: Vol 97, No 2. <https://www.journals.uchicago.edu/doi/10.1086/720097>. Accessed 29 Sep 2022
40. Short History of Nearly Every Sense—The Evolutionary History of Vertebrate Sensory Cell Types | *Integrative and Comparative Biology* | Oxford Academic. <https://academic.oup.com/icb/article/58/2/301/4993961>. Accessed 29 Sep 2022
41. Olfactory neuron turnover in adult *Drosophila* | *bioRxiv*. <https://www.biorxiv.org/content/10.1101/2020.11.08.371096v1>. Accessed 29 Sep 2022
42. Fletcher RB, Das D, Gadye L, et al (2017) Deconstructing Olfactory Stem Cell Trajectories at Single Cell Resolution. *Cell Stem Cell* 20:817–830.e8. <https://doi.org/10.1016/j.stem.2017.04.003>
43. Chiu J, DeSalle R, Lam HM, et al (1999) Molecular evolution of glutamate receptors: a primitive signaling mechanism that existed before plants and animals diverged. *Molecular Biology and Evolution* 16:826–838. <https://doi.org/10.1093/oxfordjournals.molbev.a026167>
44. Gumerov VM, Andrianova EP, Zhulin IB (2021) Diversity of bacterial chemosensory systems. *Curr Opin Microbiol* 61:42–50. <https://doi.org/10.1016/j.mib.2021.01.016>
45. Kaneko T, Minamisawa K, Isawa T, et al (2010) Complete Genomic Structure of the Cultivated Rice Endophyte *Azospirillum* sp. B510. *DNA Res* 17:37–50. <https://doi.org/10.1093/dnares/dsp026>
46. Hara TJ (2006) Feeding behaviour in some teleosts is triggered by single amino acids primarily through olfaction. *Journal of Fish Biology* 68:810–825. <https://doi.org/10.1111/j.0022-1112.2006.00967.x>
47. Kawai M, Sekine-Hayakawa Y, Okiyama A, Ninomiya Y (2012) Gustatory sensation of (L)- and (D)-amino acids in humans. *Amino Acids* 43:2349–2358. <https://doi.org/10.1007/s00726-012-1315-x>
48. Schmidt R, Cordovez V, de Boer W, et al (2015) Volatile affairs in microbial interactions. *ISME J* 9:2329–2335. <https://doi.org/10.1038/ismej.2015.42>

49. Avalos M, van Wezel GP, Raaijmakers JM, Garbeva P (2018) Healthy scents: microbial volatiles as new frontier in antibiotic research? *Curr Opin Microbiol* 45:84–91. <https://doi.org/10.1016/j.mib.2018.02.011>
50. Houten JV (2015) Microbial Chemical Sensing. In: *Handbook of Olfaction and Gustation*. John Wiley & Sons, Ltd, pp 511–530
51. The Evolution of the GPCR Signaling System in Eukaryotes: Modularity, Conservation, and the Transition to Metazoan Multicellularity - PMC. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3971589/>. Accessed 29 Sep 2022
52. Sieber B, Coronas-Serna JM, Martin SG (2022) A focus on yeast mating: From pheromone signaling to cell-cell fusion. *Seminars in Cell & Developmental Biology*. <https://doi.org/10.1016/j.semdb.2022.02.003>
53. Nilius B, Owsianik G (2011) The transient receptor potential family of ion channels. *Genome Biol* 12:218. <https://doi.org/10.1186/gb-2011-12-3-218>
54. Wicher D (2012) Functional and evolutionary aspects of chemoreceptors. *Frontiers in Cellular Neuroscience* 6:
55. Plant Volatiles as a Defense against Insect Herbivores | *Plant Physiology* | Oxford Academic. <https://academic.oup.com/plphys/article/121/2/325/6096239>. Accessed 29 Sep 2022
56. Nagashima A, Higaki T, Koeduka T, et al (2019) Transcriptional regulators involved in responses to volatile organic compounds in plants. *J Biol Chem* 294:2256–2266. <https://doi.org/10.1074/jbc.RA118.005843>
57. Ferkey DM, Sengupta P, L'Etoile ND (2021) Chemosensory signal transduction in *Caenorhabditis elegans*. *Genetics* 217:iyab004. <https://doi.org/10.1093/genetics/iyab004>
58. L'Etoile ND, Bargmann CI (2000) Olfaction and odor discrimination are mediated by the *C. elegans* guanylyl cyclase ODR-1. *Neuron* 25:575–586. [https://doi.org/10.1016/s0896-6273\(00\)81061-2](https://doi.org/10.1016/s0896-6273(00)81061-2)
59. Merritt DM, MacKay-Clackett I, Almeida SMT, et al (2022) Arrestin-mediated desensitization enables intraneuronal olfactory discrimination in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* 119:e2116957119. <https://doi.org/10.1073/pnas.2116957119>
60. Sato Y, Miyasaka N, Yoshihara Y (2007) Hierarchical Regulation of Odorant Receptor Gene Choice and Subsequent Axonal Projection of Olfactory Sensory Neurons in Zebrafish. *J Neurosci* 27:1606–1615. <https://doi.org/10.1523/JNEUROSCI.4218-06.2007>
61. Goldman AL, Van der Goes van Naters W, Lessing D, et al (2005) Coexpression of Two Functional Odor Receptors in One Neuron. *Neuron* 45:661–666. <https://doi.org/10.1016/j.neuron.2005.01.025>
62. Serizawa S, Ishii T, Nakatani H, et al (2000) Mutually exclusive expression of odorant receptor transgenes. *Nat Neurosci* 3:687–693. <https://doi.org/10.1038/76641>

63. Malnic B, Hirono J, Sato T, Buck LB (1999) Combinatorial Receptor Codes for Odors. *Cell* 96:713–723. [https://doi.org/10.1016/S0092-8674\(00\)80581-4](https://doi.org/10.1016/S0092-8674(00)80581-4)
64. Rondón JJ, Moreyra NN, Pisarenco VA, et al (2022) Evolution of the odorant-binding protein gene family in *Drosophila*. *Frontiers in Ecology and Evolution* 10:
65. Pelosi P, Knoll W (2022) Odorant-binding proteins of mammals. *Biol Rev Camb Philos Soc* 97:20–44. <https://doi.org/10.1111/brv.12787>
66. Baldwin SR, Mohapatra P, Nagalla M, et al (2021) Identification and characterization of CYPs induced in the *Drosophila* antenna by exposure to a plant odorant. *Sci Rep* 11:20530. <https://doi.org/10.1038/s41598-021-99910-9>
67. Thiebaud N, Veloso Da Silva S, Jakob I, et al (2013) Odorant Metabolism Catalyzed by Olfactory Mucosal Enzymes Influences Peripheral Olfactory Responses in Rats. *PLoS One* 8:e59547. <https://doi.org/10.1371/journal.pone.0059547>
68. Niimura Y (2012) Olfactory Receptor Multigene Family in Vertebrates: From the Viewpoint of Evolutionary Genomics. *Curr Genomics* 13:103–114. <https://doi.org/10.2174/138920212799860706>
69. Bushdid C, de March CA, Topin J, et al (2019) Mammalian class I odorant receptors exhibit a conserved vestibular-binding pocket. *Cell Mol Life Sci* 76:995–1004. <https://doi.org/10.1007/s00018-018-2996-4>
70. Gomez-Diaz C, Martin F, Garcia-Fernandez JM, Alcorta E (2018) The Two Main Olfactory Receptor Families in *Drosophila*, ORs and IRs: A Comparative Approach. *Frontiers in Cellular Neuroscience* 12:
71. Butterwick JA, Del Mármol J, Kim KH, et al (2018) Cryo-EM structure of the insect olfactory receptor Orco. *Nature* 560:447–452. <https://doi.org/10.1038/s41586-018-0420-8>
72. Bryant P, Pozzati G, Elofsson A (2022) Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* 13:1265. <https://doi.org/10.1038/s41467-022-28865-w>
73. Silbering AF, Benton R (2010) Ionotropic and metabotropic mechanisms in chemoreception: “chance or design”? *EMBO reports* 11:173–179. <https://doi.org/10.1038/embor.2010.8>
74. Fleischer J, Pregitzer P, Breer H, Krieger J (2018) Access to the odor world: olfactory receptors and their role for signal transduction in insects. *Cell Mol Life Sci* 75:485–508. <https://doi.org/10.1007/s00018-017-2627-5>
75. Regier JC, Shultz JW, Zwick A, et al (2010) Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463:1079–1083. <https://doi.org/10.1038/nature08742>
76. del Mármol J, Yedlin MA, Ruta V (2021) The structural basis of odorant recognition in insect olfactory receptors. *Nature* 597:126–131. <https://doi.org/10.1038/s41586-021-03794-8>

77. Andersson MN, Löfstedt C, Newcomb RD (2015) Insect olfaction and the evolution of receptor tuning. *Frontiers in Ecology and Evolution* 3:
78. Brand P, Robertson HM, Lin W, et al (2018) The origin of the odorant receptor gene family in insects. *eLife* 7:e38340. <https://doi.org/10.7554/eLife.38340>
79. Niimura Y, Matsui A, Touhara K (2014) 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. <https://doi.org/10.1101/gr.169532.113>
80. Ikegami K, de March CA, Nagai MH, et al (2020) Structural instability and divergence from conserved residues underlie intracellular retention of mammalian odorant receptors. *Proceedings of the National Academy of Sciences* 117:2957–2967. <https://doi.org/10.1073/pnas.1915520117>
81. Niimura Y, Matsui A, Touhara K (2018) Acceleration of Olfactory Receptor Gene Loss in Primate Evolution: Possible Link to Anatomical Change in Sensory Systems and Dietary Transition. *Molecular Biology and Evolution* 35:1437–1450. <https://doi.org/10.1093/molbev/msy042>
82. Liu A, He F, Shen L, et al (2019) Convergent degeneration of olfactory receptor gene repertoires in marine mammals. *BMC Genomics* 20:977. <https://doi.org/10.1186/s12864-019-6290-0>
83. Genva M, Kenne Kemene T, Deleu M, et al (2019) Is It Possible to Predict the Odor of a Molecule on the Basis of its Structure? *International Journal of Molecular Sciences* 20:3018. <https://doi.org/10.3390/ijms20123018>
84. Leitereg TJ, Guadagni DG, Harris Jean, et al (1971) Chemical and sensory data supporting the difference between the odors of the enantiomeric carvones. *J Agric Food Chem* 19:785–787. <https://doi.org/10.1021/jf60176a035>
85. Jacob S, Garcia S, Hayreh D, McClintock MK (2002) Psychological effects of musky compounds: comparison of androstadienone with androstenol and muscone. *Horm Behav* 42:274–283. <https://doi.org/10.1006/hbeh.2002.1826>
86. Trimmer C, Keller A, Murphy NR, et al (2019) Genetic variation across the human olfactory receptor repertoire alters odor perception. *Proceedings of the National Academy of Sciences* 116:9475–9480. <https://doi.org/10.1073/pnas.1804106115>
87. Sanchez-Lengeling B, Wei JN, Lee BK, et al The Chemistry of Smell: Learning Generalizable Perceptual Representations of Small Molecules. 6
88. Peterlin Z, Firestein S, Rogers ME (2014) The state of the art of odorant receptor deorphanization: A report from the orphanage. *J Gen Physiol* 143:527–542. <https://doi.org/10.1085/jgp.201311151>
89. Cong X, Ren W, Pacalon J, et al (2022) Large-Scale G Protein-Coupled Olfactory Receptor–Ligand Pairing. *ACS Cent Sci* 8:379–387. <https://doi.org/10.1021/acscentsci.1c01495>

90. Kepchia D, Xu P, Terry R, et al (2019) Use of machine learning to identify novel, behaviorally active antagonists of the insect odorant receptor co-receptor (Orco) subunit. *Sci Rep* 9:4055. <https://doi.org/10.1038/s41598-019-40640-4>
91. Caballero-Vidal G, Bouysset C, Gévar J, et al (2021) Reverse chemical ecology in a moth: machine learning on odorant receptors identifies new behaviorally active agonists. *Cell Mol Life Sci* 78:6593–6603. <https://doi.org/10.1007/s00018-021-03919-2>
92. Bushdid C, de March CA, Matsunami H, Golebiowski J (2018) Numerical Models and In Vitro Assays to Study Odorant Receptors. In: Simoes de Souza FM, Antunes G (eds) *Olfactory Receptors: Methods and Protocols*. Springer, New York, NY, pp 77–93
93. Bushdid C, de March CA, Fiorucci S, et al (2018) Agonists of G-Protein-Coupled Odorant Receptors Are Predicted from Chemical Features. *J Phys Chem Lett* 9:2235–2240. <https://doi.org/10.1021/acs.jpcclett.8b00633>
94. Kowalewski J, Ray A (2020) Predicting Human Olfactory Perception from Activities of Odorant Receptors. *iScience* 23:101361. <https://doi.org/10.1016/j.isci.2020.101361>
95. Jabeen A, de March CA, Matsunami H, Ranganathan S (2021) Machine Learning Assisted Approach for Finding Novel High Activity Agonists of Human Ectopic Olfactory Receptors. *Int J Mol Sci* 22:11546. <https://doi.org/10.3390/ijms222111546>
96. Toropov AA, Toropova AP, Cappellini L, et al (2016) Odor threshold prediction by means of the Monte Carlo method. *Ecotoxicology and Environmental Safety* 133:390–394. <https://doi.org/10.1016/j.ecoenv.2016.07.039>
97. Kovatcheva A, Golbraikh A, Oloff S, et al (2004) Combinatorial QSAR of ambergris fragrance compounds. *J Chem Inf Comput Sci* 44:582–595. <https://doi.org/10.1021/ci034203t>
98. Edwards PA, Jurs PC (1989) Correlation of odor intensities with structural properties of odorants. *Chemical Senses* 14:281–291. <https://doi.org/10.1093/chemse/14.2.281>
99. Sell CS (2006) On the unpredictability of odor. *Angew Chem Int Ed Engl* 45:6254–6261. <https://doi.org/10.1002/anie.200600782>
100. Khan RM, Luk C-H, Flinker A, et al (2007) Predicting Odor Pleasantness from Odorant Structure: Pleasantness as a Reflection of the Physical World. *J Neurosci* 27:10015–10023. <https://doi.org/10.1523/JNEUROSCI.1158-07.2007>
101. Keller A, Vosshall LB (2016) Olfactory perception of chemically diverse molecules. *BMC Neurosci* 17:55. <https://doi.org/10.1186/s12868-016-0287-2>
102. Keller A, Gerkin RC, Guan Y, et al (2017) Predicting Human Olfactory Perception from Chemical Features of Odor Molecules. *Science* 355:820–826. <https://doi.org/10.1126/science.aal2014>
103. Lee BK, Mayhew EJ, Sanchez-Lengeling B, et al (2022) A Principal Odor Map Unifies Diverse Tasks in Human Olfactory Perception. 2022.09.01.504602

104. Buck L, Axel R (1991) A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* 65:175–187. [https://doi.org/10.1016/0092-8674\(91\)90418-X](https://doi.org/10.1016/0092-8674(91)90418-X)
105. Floriano WB, Vaidehi N, Goddard WA, et al (2000) Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor. *Proc Natl Acad Sci U S A* 97:10712–10716
106. Yarov-Yarovoy V, Schonbrun J, Baker D (2006) Multipass Membrane Protein Structure Prediction Using Rosetta. *Proteins* 62:1010–1025. <https://doi.org/10.1002/prot.20817>
107. Hopf TA, Green AG, Schubert B, et al (2019) The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics* 35:1582–1584. <https://doi.org/10.1093/bioinformatics/bty862>
108. Hopf TA, Morinaga S, Ihara S, et al (2015) Amino acid coevolution reveals three-dimensional structure and functional domains of insect odorant receptors. *Nat Commun* 6:6077. <https://doi.org/10.1038/ncomms7077>
109. Govindarajan S, Recabarren R, Goldstein RA (1999) Estimating the total number of protein folds. *Proteins: Structure, Function, and Bioinformatics* 35:408–414. [https://doi.org/10.1002/\(SICI\)1097-0134\(19990601\)35:4<408::AID-PROT4>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-0134(19990601)35:4<408::AID-PROT4>3.0.CO;2-A)
110. Koonin EV, Wolf YI, Karev GP (2002) The structure of the protein universe and genome evolution. *Nature* 420:218–223. <https://doi.org/10.1038/nature01256>
111. Wu S, Skolnick J, Zhang Y (2007) Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol* 5:17. <https://doi.org/10.1186/1741-7007-5-17>
112. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725–738. <https://doi.org/10.1038/nprot.2010.5>
113. Jia L, Li S, Dai W, et al (2021) Convergent olfactory trace amine-associated receptors detect biogenic polyamines with distinct motifs via a conserved binding site. *J Biol Chem* 297:101268. <https://doi.org/10.1016/j.jbc.2021.101268>
114. Zhang J, Yang J, Jang R, Zhang Y (2015) GPCR-I-TASSER: A hybrid approach to G protein-coupled receptor structure modeling and the application to the human genome. *Structure* 23:1538–1549. <https://doi.org/10.1016/j.str.2015.06.007>
115. Muhammed MT, Aki-Yalcin E (2019) Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical Biology & Drug Design* 93:12–20. <https://doi.org/10.1111/cbdd.13388>
116. Larsson P, Wallner B, Lindahl E, Elofsson A (2008) Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci* 17:990–1002. <https://doi.org/10.1110/ps.073344908>
117. Waterhouse A, Bertoni M, Bienert S, et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46:W296–W303. <https://doi.org/10.1093/nar/gky427>

118. Eswar N, Webb B, Marti-Renom MA, et al (2006) Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics* 15:5.6.1-5.6.30. <https://doi.org/10.1002/0471250953.bi0506s15>
119. van Beusekom B, Joosten K, Hekkelman ML, et al (2018) Homology-based loop modeling yields more complete crystallographic protein structures. *IUCrJ* 5:585–594. <https://doi.org/10.1107/S2052252518010552>
120. Launay G, Sanz G, Pajot-Augy E, Gibrat J-F (2012) Modeling of mammalian olfactory receptors and docking of odorants. *Biophys Rev* 4:255–269. <https://doi.org/10.1007/s12551-012-0080-0>
121. Yu Y, Ma Z, Pacalon J, et al (2022) Extracellular loop 2 of G protein–coupled olfactory receptors is critical for odorant recognition. *Journal of Biological Chemistry* 298:. <https://doi.org/10.1016/j.jbc.2022.102331>
122. Hameduh T, Haddad Y, Adam V, Heger Z (2020) Homology modeling in the time of collective and artificial intelligence. *Comput Struct Biotechnol J* 18:3494–3506. <https://doi.org/10.1016/j.csbj.2020.11.007>
123. Barozet A, Chacón P, Cortés J (2021) Current approaches to flexible loop modeling. *Current Research in Structural Biology* 3:187–191. <https://doi.org/10.1016/j.crstbi.2021.07.002>
124. Charlier L, Topin J, de March CA, et al (2013) Molecular modelling of odorant/olfactory receptor complexes. *Methods Mol Biol* 1003:53–65. https://doi.org/10.1007/978-1-62703-377-0_4
125. Gelis L, Wolf S, Hatt H, et al (2012) Prediction of a ligand-binding niche within a human olfactory receptor by combining site-directed mutagenesis with dynamic homology modeling. *Angew Chem Int Ed Engl* 51:1274–1278. <https://doi.org/10.1002/anie.201103980>
126. de March CA, Kim S-K, Antonczak S, et al (2015) G protein-coupled odorant receptors: From sequence to structure. *Protein Science : A Publication of the Protein Society* 24:1543–1548. <https://doi.org/10.1002/pro.2717>
127. Shen M-Y, Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15:2507–2524. <https://doi.org/10.1110/ps.062416606>
128. Kryshtafovych A, Schwede T, Topf M, et al (2019) Critical Assessment of Methods of Protein Structure Prediction (CASP) – Round XIII. *Proteins* 87:1011–1020. <https://doi.org/10.1002/prot.25823>
129. Senior AW, Evans R, Jumper J, et al (2020) Improved protein structure prediction using potentials from deep learning. *Nature* 577:706–710. <https://doi.org/10.1038/s41586-019-1923-7>
130. Kryshtafovych A, Schwede T, Topf M, et al (2021) Critical assessment of methods of protein structure prediction (CASP)-Round XIV. *Proteins* 89:1607–1617. <https://doi.org/10.1002/prot.26237>

131. Jumper J, Evans R, Pritzel A, et al (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
132. Guo H-B, Perminov A, Bekele S, et al (2022) AlphaFold2 models indicate that protein sequence determines both structure and dynamics. *Sci Rep* 12:10696. <https://doi.org/10.1038/s41598-022-14382-9>
133. Baek M, DiMaio F, Anishchenko I, et al (2021) Accurate prediction of protein structures and interactions using a 3-track neural network. *Science* 373:871–876. <https://doi.org/10.1126/science.abj8754>
134. Evans R, O’Neill M, Pritzel A, et al (2022) Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034. <https://doi.org/10.1101/2021.10.04.463034>
135. (2020) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49:D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
136. PDBe-KB consortium (2020) PDBe-KB: a community-driven resource for structural and functional annotations. *Nucleic Acids Research* 48:D344–D353. <https://doi.org/10.1093/nar/gkz853>
137. Varadi M, Anyango S, Deshpande M, et al (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50:D439–D444. <https://doi.org/10.1093/nar/gkab1061>
138. Renthal R, Chen LY (2022) Tunnel connects lipid bilayer to occluded odorant-binding site of insect olfactory receptor. *Biophysical Chemistry* 289:106862. <https://doi.org/10.1016/j.bpc.2022.106862>
139. Ballante F, Kooistra AJ, Kampen S, et al (2021) Structure-Based Virtual Screening for Ligands of G Protein–Coupled Receptors: What Can Molecular Docking Do for You? *Pharmacol Rev* 73:1698–1736. <https://doi.org/10.1124/pharmrev.120.000246>
140. Winkler DA (2020) Ligand Entropy Is Hard but Should Not Be Ignored. *Journal of Chemical Information and Modeling*. <https://doi.org/10.1021/acs.jcim.0c01146>
141. Li Y, Gao Y, Holloway MK, Wang R (2020) Prediction of the Favorable Hydration Sites in a Protein Binding Pocket and Its Application to Scoring Function Formulation. *J Chem Inf Model* 60:4359–4375. <https://doi.org/10.1021/acs.jcim.9b00619>
142. Rognan D (2017) The impact of in silico screening in the discovery of novel and safer drug candidates. *Pharmacol Ther* 175:47–66. <https://doi.org/10.1016/j.pharmthera.2017.02.034>
143. Geithe C, Protze J, Kreuchwig F, et al (2017) Structural determinants of a conserved enantiomer-selective carvone binding pocket in the human odorant receptor OR1A1. *Cell Mol Life Sci* 74:4209–4229. <https://doi.org/10.1007/s00018-017-2576-z>
144. Yuan S, Dahoun T, Brugarolas M, et al (2019) Computational modeling of the olfactory receptor Olfr73 suggests a molecular basis for low potency of olfactory receptor-activating compounds. *Commun Biol* 2:1–10. <https://doi.org/10.1038/s42003-019-0384-8>

145. Bavan S, Sherman B, Luetje CW, Abaffy T (2014) Discovery of Novel Ligands for Mouse Olfactory Receptor MOR42-3 Using an In Silico Screening Approach and In Vitro Validation. *PLOS ONE* 9:e92064. <https://doi.org/10.1371/journal.pone.0092064>
146. Jaiteh M, Rodríguez-Espigares I, Selent J, Carlsson J (2020) Performance of virtual screening against GPCR homology models: Impact of template selection and treatment of binding site plasticity. *PLoS Comput Biol* 16:e1007680. <https://doi.org/10.1371/journal.pcbi.1007680>
147. Miller MD, Phillips GN (2021) Moving beyond static snapshots: Protein dynamics and the Protein Data Bank. *Journal of Biological Chemistry* 296:. <https://doi.org/10.1016/j.jbc.2021.100749>
148. McRobb FM, Negri A, Beuming T, Sherman W (2016) Molecular dynamics techniques for modeling G protein-coupled receptors. *Current Opinion in Pharmacology* 30:69–75. <https://doi.org/10.1016/j.coph.2016.07.001>
149. Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. *Neuron* 99:1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011>
150. Wang Y, Bugge K, Kragelund BB, Lindorff-Larsen K (2018) Role of protein dynamics in transmembrane receptor signalling. *Current Opinion in Structural Biology* 48:74–82. <https://doi.org/10.1016/j.sbi.2017.10.017>
151. Wales DJ (2018) Exploring Energy Landscapes. *Annu Rev Phys Chem* 69:401–425. <https://doi.org/10.1146/annurev-physchem-050317-021219>
152. Kmiecik S, Gront D, Kolinski M, et al (2016) Coarse-Grained Protein Models and Their Applications. *Chem Rev* 116:7898–7936. <https://doi.org/10.1021/acs.chemrev.6b00163>
153. Kästner J (2011) Umbrella sampling. *WIREs Computational Molecular Science* 1:932–942. <https://doi.org/10.1002/wcms.66>
154. Mori T, Miyashita N, Im W, et al (2016) Molecular dynamics simulations of biological membranes and membrane proteins using enhanced conformational sampling algorithms. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1858:1635–1651. <https://doi.org/10.1016/j.bbamem.2015.12.032>
155. Harpole TJ, Delemotte L (2018) Conformational landscapes of membrane proteins delineated by enhanced sampling molecular dynamics simulations. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1860:909–926. <https://doi.org/10.1016/j.bbamem.2017.10.033>
156. Liao Q (2020) Chapter Four - Enhanced sampling and free energy calculations for protein simulations. In: Strodel B, Barz B (eds) *Progress in Molecular Biology and Translational Science*. Academic Press, pp 177–213
157. Matricon P, Ranganathan A, Warnick E, et al (2017) Fragment optimization for GPCRs by molecular dynamics free energy calculations: Probing druggable subpockets of the A2A adenosine receptor binding site. *Sci Rep* 7:6398. <https://doi.org/10.1038/s41598-017-04905-0>

158. Souza PCT, Alessandri R, Barnoud J, et al (2021) Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat Methods* 18:382–388. <https://doi.org/10.1038/s41592-021-01098-3>
159. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. *Nature* 267:585–590. <https://doi.org/10.1038/267585a0>
160. Dutagaci B, Heo L, Feig M (2018) Structure refinement of membrane proteins via molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* 86:738–750. <https://doi.org/10.1002/prot.25508>
161. Kapla J, Rodríguez-Espigares I, Ballante F, et al (2021) Can molecular dynamics simulations improve the structural accuracy and virtual screening performance of GPCR models? *PLOS Computational Biology* 17:e1008936. <https://doi.org/10.1371/journal.pcbi.1008936>
162. de March CA, Topin J, Bruguera E, et al (2018) Odorant Receptor 7D4 Activation Dynamics. *Angewandte Chemie International Edition* 57:4554–4558. <https://doi.org/10.1002/anie.201713065>
163. Paladino A, Costantini S, Colonna G, Facchiano AM (2008) Molecular modelling of miraculin: Structural analyses and functional hypotheses. *Biochem Biophys Res Commun* 367:26–32. <https://doi.org/10.1016/j.bbrc.2007.12.102>
164. Sykes DA, Stoddart LA, Kilpatrick LE, Hill SJ (2019) Binding kinetics of ligands acting at GPCRs. *Mol Cell Endocrinol* 485:9–19. <https://doi.org/10.1016/j.mce.2019.01.018>
165. You W, Tang Z, Chang CA (2019) Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? *J Chem Theory Comput* 15:2433–2443. <https://doi.org/10.1021/acs.jctc.8b01142>
166. Hajjar E, Perahia D, Débat H, et al (2006) Odorant Binding and Conformational Dynamics in the Odorant-binding Protein*. *Journal of Biological Chemistry* 281:29929–29937. <https://doi.org/10.1074/jbc.M604869200>
167. Delemotte L, Dehez F, Treptow W, Tarek M (2008) Modeling Membranes under a Transmembrane Potential. *J Phys Chem B* 112:5547–5550. <https://doi.org/10.1021/jp710846y>
168. Dror RO, Arlow DH, Maragakis P, et al (2011) Activation mechanism of the β 2-adrenergic receptor. *Proc Natl Acad Sci U S A* 108:18684–18689. <https://doi.org/10.1073/pnas.1110499108>
169. Dror RO, Pan AC, Arlow DH, et al (2011) Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci U S A* 108:13118–13123. <https://doi.org/10.1073/pnas.1104614108>
170. de March CA, Yu Y, Ni MJ, et al (2015) Conserved Residues Control Activation of Mammalian G Protein-Coupled Odorant Receptors. *J Am Chem Soc* 137:8611–8616. <https://doi.org/10.1021/jacs.5b04659>

171. Sekharan S, Ertem MZ, Zhuang H, et al (2014) QM/MM Model of the Mouse Olfactory Receptor MOR244-3 Validated by Site-Directed Mutagenesis Experiments. *Biophys J* 107:L5–L8. <https://doi.org/10.1016/j.bpj.2014.07.031>
172. Xu Z, Guo L, Qian X, et al (2022) Two entry tunnels in mouse TAAR9 suggest the possibility of multi-entry tunnels in olfactory receptors. *Sci Rep* 12:2691. <https://doi.org/10.1038/s41598-022-06591-z>

Chapter I
**Insect Odorant Receptor Co-
Receptor (Orco)**

Prior to 1900, pest control was based primarily on inorganic compounds or botanical substances such as nicotine. The isolation and discovery of the structure of these pharmaceutical compounds led to great advances in pest control. About 70 years ago, the discovery of potent synthetic organic insecticides eliminated most research needs in this area. Public concern about the toxicity of pesticides and their impact on the environment began to grow around 1950. Pesticides were identified as a possible cause of cancer and were identified as having an impact on biodiversity (death of birds, fish, bees). Most of the pesticides used are neurotoxins acting on acetylcholinesterases (AChE), gamma-aminobutyric acid (GABA) chloride channels or voltage-gated sodium channels. Through Darwinian selection, pests have also adapted and developed tolerance to insecticides that act on the nervous system, which revives the need for new approaches [1].

Repellents are playing an increasingly important role in pest control. With molecular knowledge, it is possible to design new odorants with attractive or repellent properties, with the significant advantage of being environmentally friendly. A recent work has identified new odorant compounds that modify the behavior of larvae of the noctuid moth *Spodoptera littoralis* (*S. littoralis*), by combining machine learning and electrophysiological validation [2].

In this chapter, the main objective was to develop the structural understanding of the insect olfactory system at the receptor level, taking advantage of the recent elucidation of the experimental structure of the Orco homotetramer [3]. The strategy of targeting the Orco portion of the insect olfactory channel to achieve a repellent effect may already be used by plants, as suggested by the preliminary work of Chen *et al.* (2014) [4]. Orco antagonist molecules also appear to allosterically block the activation of ORs by odorants. Because Orco associates with all OR subtypes and is also well conserved among insect species, the design of new synthetic volatile compounds that block it could enable the creation of less dangerous and more specific broad spectrum insect repellents. We identified and described the diffusion pathway of the synthetic agonist VUAA1 into its binding site through unbiased molecular dynamic simulations. The ligand was placed unrestrictedly in the extracellular portion of the simulation box, free to sample into the cavity. From these simulations, a selection of positions was tested in vitro to probe the importance of these residues.

Contribution: Jérôme Golebiowski, Jérémie Topin and our collaborator Christophe Moreau designed the study. I set up, launched, and analyzed the molecular dynamics simulations with

the help of Jérémie Topin. Guillaume Audic performed the *in vitro* assays. Guillaume Audic and I contributed equally as first authors.

Reference: *Pacalon J, Audic G, Magnat J, Philip M, Golebiowski J, Moreau C, Topin J. Ligand of the conserved insect odorant receptor co-receptor reaches its binding site through a dynamic translocation pathway. 2022, under review.*

1. Casida JE, Quistad GB (1998) Golden age of insecticide research: past, present, or future? *Annu Rev Entomol* 43:1–16. <https://doi.org/10.1146/annurev.ento.43.1.1>
2. Caballero-Vidal G, Bouysset C, Gévar J, et al (2021) Reverse chemical ecology in a moth: machine learning on odorant receptors identifies new behaviorally active agonists. *Cell Mol Life Sci* 78:6593–6603. <https://doi.org/10.1007/s00018-021-03919-2>
3. Butterwick JA, Del Mármol J, Kim KH, et al (2018) Cryo-EM structure of the insect olfactory receptor Orco. *Nature* 560:447–452. <https://doi.org/10.1038/s41586-018-0420-8>
4. Chen S, Luetje CW (2014) Trace amines inhibit insect odorant receptor function through antagonism of the co-receptor subunit. *F1000Res* 3:84. <https://doi.org/10.12688/f1000research.3825.1>

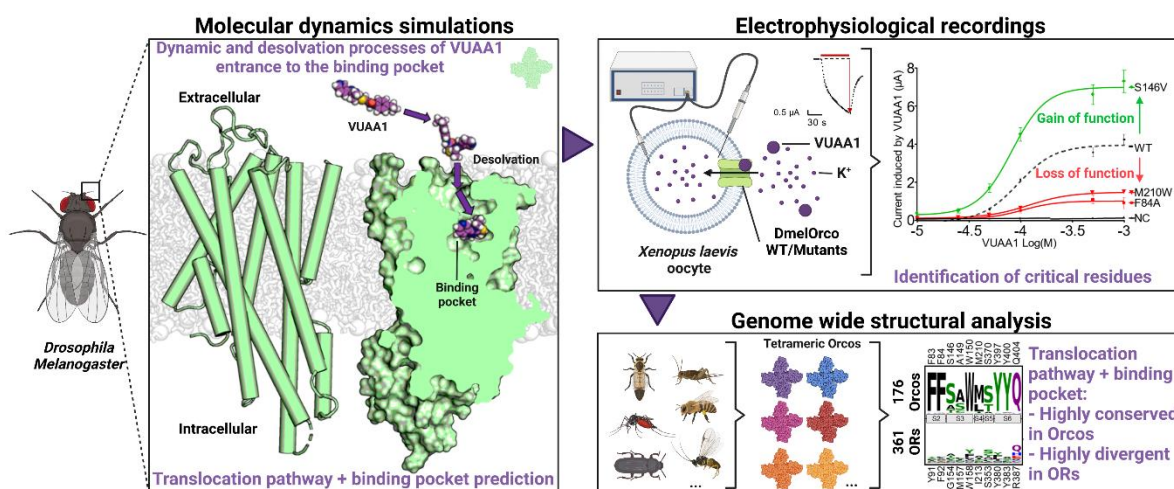
Publication 1

Ligand of the conserved insect odorant receptor co-receptor reaches its binding site through a dynamic translocation pathway.

Pacalon J, Audic G, Magnat J, Philip M, Golebiowski J, Moreau C, Topin J. 2022, under review.

Abstract

Insects are of major importance for our society, being both beneficial for agriculture and detrimental for human health as pathogen vectors. Olfaction is an essential sense for insects, notably for food and host seeking. In numerous insects, the olfactory receptor family forms a unique class of heteromeric cation channels with two subunits that evolved in opposite directions. The signal-generating subunit (Orco) is extremely conserved across species, while the odorant-binding subunit (OR) diverged to recognize specific ligands present in the environment of the insect. Despite this divergent evolution, ORs have remarkably preserved their ability to interact with Orcos, even from different species. Due to its high degree of conservation, Orco is an attractive target for the development of repellents with very broad-spectrum effects. Recent advances have revealed the homomeric structures of both an "ancient" OR and an Orco without the OR subunits. Unexpectedly, these structures in apo or ligand-bound states did not reveal the pathway taken by the ligands between the extracellular space and the deep internal cavities that have been identified as ligand binding sites in MhOR5. Using a combination of dynamic simulations and structure-function approaches, this article highlights: i) the original molecular entry mechanism of a ligand (VUAA1) into an Orco, which involves a process of dehydration of the compounds; and ii) the ligand binding site of VUAA1 in the Orco. These mechanisms are potentially common to a very large variety of insect species including winged insects.



Significance Statement

For insects, olfaction is a critical sense that can be manipulated to control their behavior either for attractive or repulsive applications. Recent progress in resolving the odorant receptors structures at an atomic scale offers unprecedented opportunities for deciphering their molecular mechanisms and particularly the binding mode of their ligands. By combining homology modeling, molecular dynamics simulations and electrophysiological recordings, we identified residues involved in the dynamic entry pathway and the binding of VUAA1 to *Drosophila melanogaster*'s Orco. These results enhance our understanding on how insect olfactory receptors decipher their volatile chemical environment, and they open new avenues for the rational design of pest control tools.

Introduction

Among all living multicellular organisms, insects represent more than half of all identified species on Earth, thus forming the most diverse group of animals (1). Insects show a remarkable capacity to adapt to a wide range of ecological niches. The rapid evolution of insect olfactory receptors is thought to contribute to this adaptation (2), endowing each insect species with the ability to selectively detect volatile chemicals associated with its specialized habitat and lifestyle. Therefore, olfaction is a vital sense necessary for them to find food, a mate, an oviposition site and a host. Moreover, the insect olfactory receptors are the main targets for the rational design of repulsive or attractive compounds for protection against vector-borne species or pest control.(3, 4)

Ground breaking studies have provided a structural description of the proteins involved in odor recognition by insects(5, 6). In addition to the gustatory receptors, the repertoire of odorant receptors is mainly composed of two distinct families: i) the olfactory receptors (ORs) that form a complex with the highly conserved odorant receptor co-receptor (Orco) (7); and ii) the ionotropic receptors (IRs) that are structurally similar to the ionotropic glutamate receptor (8). The OR/Orco receptors are mainly expressed in olfactory sensory neurons (OSNs) found in insects' antennae. An individual OSN typically expresses only a single type of OR (9), which defines the neuron's response spectrum (10). The OR/Orco complexes are proposed to form a unique class of heteromeric cation channels composed of the two related 7-transmembrane

subunits. It has been shown that Orcos could form homotetrameric channels (Figure 1A), which have a different recognition spectrum than ORs (11, 12).

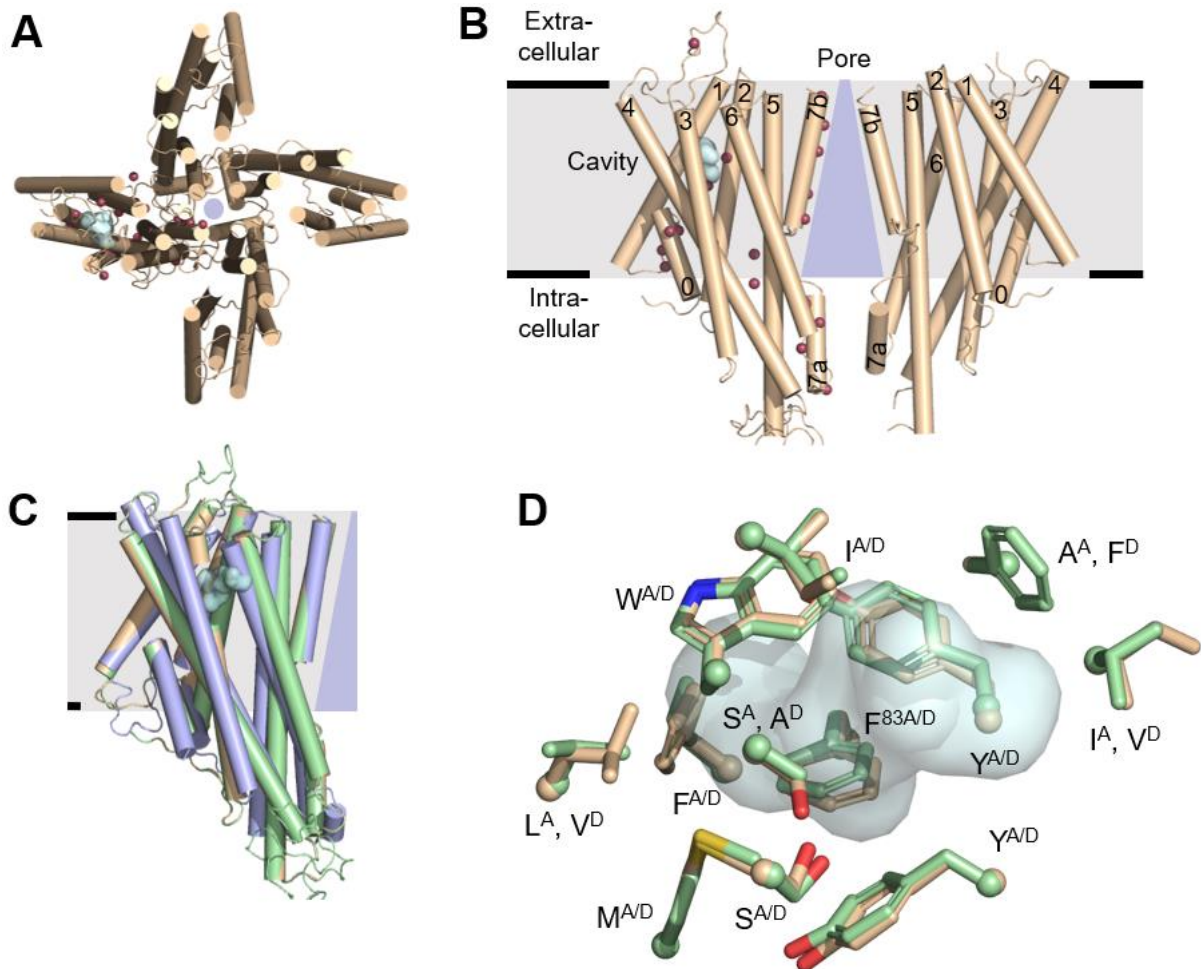


Figure 1: **A.** Extracellular view of the cryo-electronic microscopy structure of the homotetramer of *Apocrypta bakeri* Orco (AbakOrco) (pdb: 6C70). Ligand-binding pocket of related Orco receptors. **B.** Side view of two Orco subunits with a schema of the channel pore (blue trapezoid). Residues shown in red spheres are equivalent to residues critical for VUAA1 response found in Orcos from point mutations that alter channel function in *Drosophila melanogaster*, *Agrotis segetum*, *Mayetiola destructor*, *Bombyx mori*, or *Apocrypta bakeri*. **C.** AbakOrco (pdb: 6C70, beige color) membrane view, superposed on *Machilis hrabei* odorant receptor 5 (MhOR5, pdb: 7LID, blue) and *Drosophila melanogaster* Orco (DmelOrco, green) homology model. Cavity analysis reveals the conserved position of a pocket (cyan) in these 3 structures. **D.** Close view of the cavities (cyan) of DmelOrco and AbakOrco with their amino acids represented as sticks (respectively green and beige). DmelOrco and AbakOrco cavities share 73% of sequence identity (82% of similarity). The superscript letters A and D refer to AbakOrco and DmelOrco, respectively.

Orcos seems to appear late in the evolution of insects and constitute a remarkable example of an adaptive system, with a unique highly conserved signaling subunit (Orco) that can associate with a large repertoire of odorant receptor subunits that diverge to recognize specific ligands.(13, 14) The evolution of ORs that led to the appearance of Orcos induced a total loss of odorant binding for this subunit, while engendering the ability to bind few synthetic ligands, like VUAA1 (15-21). On the other hand, the "ancestral" OR5 receptor from *Machilis hrabei* (MhOR5) is activated by a large set of odorants, but not by VUAA1. (6) DmelOrco and MhOR5 share 18.3% sequence identity and adopt the same tertiary fold (Figure 1C). However, the origin of the differences in the recognition spectra of the two receptors is still not fully understood.

To decipher the molecular mechanisms governing the response of Orcos to ligands, different structure-function studies were previously employed based on site-directed mutagenesis combined with two-electrode voltage-clamp (TEVC) measurements. Figure 1B summarizes the position of different residues that showed a functional impact when mutated (5, 22-25). These studies have highlighted the central role of residues from helix S7 in forming a hydrophobic gate that contributes to ion selectivity. Moreover, the structures of MhOR5 in complex with two agonists, eugenol and DEET, revealed the ligand binding cavity of this receptor (Figure 1C, D) (6).

Despite these highly informative structural studies, several questions remain, in particular the entry pathway and the binding site of ligands in Orcos. Their identification is essential for understanding the high specificity of action of Orco ligands and for the rational design of new molecules for attractive or repulsive applications. To identify the binding pocket and the translocation pathway of VUAA1 from the extracellular space to the Orco binding site, we combined molecular modeling approaches with site-directed mutagenesis and functional characterization by the TEVC technique.

Results

Determination of the optimal Orco.

Olfactory receptors are notorious for weakly expressing in heterologous systems, which impedes their functional characterization. Before initiating MD simulations, we searched for the optimal Orco that generates the highest response to VUAA1 when expressed in *Xenopus*

oocytes. Orcos from *Apocrypta bakeri*, *Drosophila melanogaster*, *Aedes albopictus* and *Culex quinquefasciatus* were functionally characterized by the TEVC technique. The results (Figure 2) clearly demonstrate that DmelOrco generated the highest current amplitude in the presence of VUAA1 and it was chosen as the model for both computations and experiments.

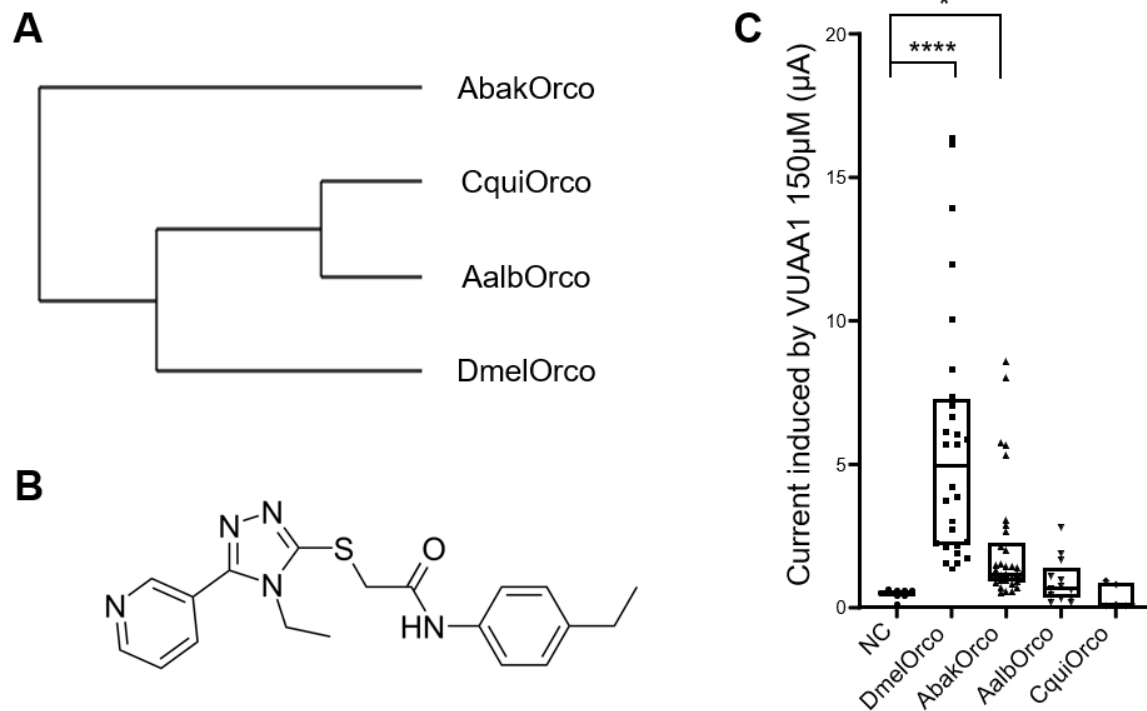


Figure 2: **A.** Phylogenetic tree of Orcos from different species. NC: negative control (water injected oocytes), Dmel: *Drosophila melanogaster*, Abak: *Apocrypta bakeri*; Aalb: *Aedes albopictus*. **B.** structure of VUAA1. **C.** Response to VUAA1 of Orcos from different species. Amplitude of currents induced by VUAA1 at 500μM applied on xenopus oocytes heterologously expressing Orcos. Results are median +/- SEM. P values are 0.92 (Aalb), 0.0438 (Abak, *) and <0.0001 (Dmel, ****) with NC as reference in the Kruskal-Wallis test.

Orcos show a conserved cavity.

A 3D model of DmelOrco was built by homology modeling using the experimental structure of AbakOrco homotetramer (pdb ID: 6C70) as a template (5). The two protein sequences are highly similar (76 % of sequence identity) prefiguring a high confidence in the accuracy of the model of DmelOrco(26). The full protocol is detailed in the Materials and Methods section. We compared it to a model obtained by AlphaFold2 (extracted from the Alpha Fold Protein database) (27). Both structures show a high similarity of transmembrane segments (RMSD =

0.7 Å). The largest deviation between the structures is observed at the intracellular loop (see Figure S1). This loop is not resolved on the cryoEM structure of DmelOrco, which reflects a high flexibility.

The structure of AbakOrco (6) and the homology model of DmelOrco (Figure 1C&D), revealed a cavity between helices S1, S2, S3, S4 and S6 (Figure 1) that could play the role of the ligand binding site for VUAA1 and its analogues. Interestingly, this cavity has a position similar to the ligand binding site found in the structure of MhOR5(6)(Figure 1C). The amino acids lining the two cavities are highly conserved with 73% identity (Figure 1D). Notably, the cradle of this pocket would be formed by the residue F83^{Dmel}, which is critical for activation by VUAA1(25). In both structures and models, the access of VUAA1 to its putative binding site seems hindered by constrictions of the transmembrane helices, suggesting a progression of the molecule through a hidden and dynamic pathway.

MD simulations reveal the mechanism of VUAA1 entry and confirm the binding site.

We studied the entry of VUAA1 molecules to their putative binding site in DmelOrco by means of molecular dynamics (MD) simulations. To reach this deeply embedded site, residing in the core of the transmembrane helices, the molecule must transit through a path that is assumed to be dynamic since it is closed in the structures of AbakOrco and MhOR5. To identify this path, multiple MD simulations were performed with several ligands to enhance the sampling of rare events such as ligand migration (28, 29) and protein conformational changes (30, 31). We constructed a system containing 4 DmelOrco monomer with five VUAA1 molecules randomly placed in the extracellular part of the simulation box. Then, 22 replicas were subjected to MD simulations, leading to a total of 88 simulations on DmelOrco monomers. The total simulation time is approximately 31 μs (see Materials and Methods). A constraint was applied between each VUAA1 and the top of the channel pore to increase sampling speed without biasing the binding process. Thus, the ligands were free to sample the extracellular region of the simulation box and to diffuse into the receptor core. The migration of VUAA1 through the protein core was evaluated by the evolution of the distance between the VUAA1 center of mass and the center of mass of the binding cavity (defined as the center of mass of the eugenol molecule in MhOR5, pdb: 7LID).

The results of our simulations revealed a predominant pathway of VUAA1 entry into the binding site. From the 88 trajectories, 19 showed an entry of VUAA1 within the receptor bundle (i.e. Figure 3A, area c). Out of these 19 trajectories, 7 full entries into the binding pocket (i.e. Figure 3A, area d) were observed. The other trajectories resulted in a partial binding event, where VUAA1 remains within a vestibular site, half-way to the pocket cradle (Figure S2).

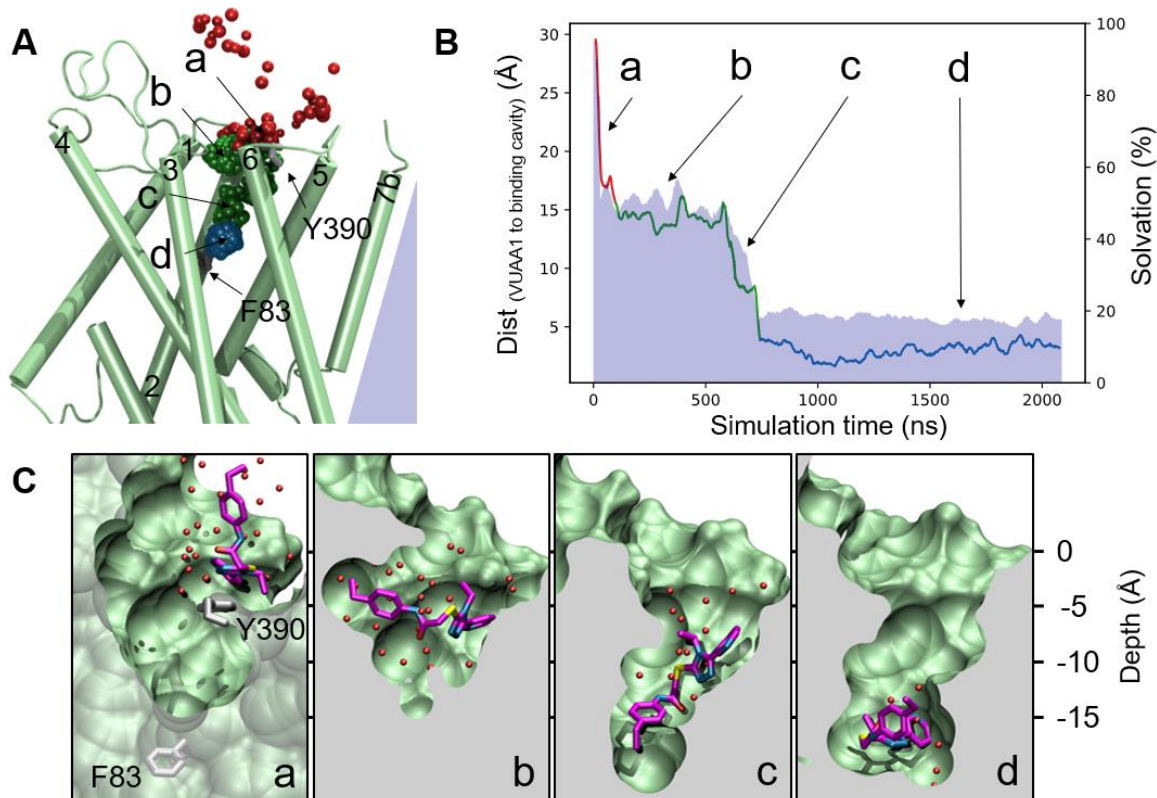


Figure 3: **A.** Prototypical trajectory of one prototypical VUAA1 binding event. The Orco monomer is shown in green. F83 and Y390 labels give their localization. The VUAA1 center of mass is represented by beads colored from red to blue according to the simulation time. **B.** Evolution of the distance between the VUAA1 centers of mass and the center of mass of the binding cavity (defined as the center of mass of the eugenol molecule in MhOR5, pdb: 7LID) The red curve represents the positions outside of the receptor. The green part of the curve represents the migration event and the blue one the sample of the binding cavity. The blue area shows the percentage of ligand solvation during the binding process (normalized to the solvation of the ligand outside the protein) **C.** close view of the path of migration of VUAA1 inside the Orco monomer corresponding to the positions a, b, c and d. Carbon atoms of VUAA1 are colored purple, carbon atoms of F83S2 and Y390S6 are grey and water molecules found less than 3 Å away from VUAA1 are represented by red spheres.

In all the seven observed binding events, VUAA1 consistently enters the receptor through the same gate and showed contact with residues belonging to helices S2 to S6. Interestingly, most of these residues are highly conserved among various Orcos (see file supdata_conservation.xlsx) in line with the similar action of VUAA1 observed in the majority of insect Orcos (25). The ligand does not interact with helix S7, which forms the tetrameric pore, suggesting that VUAA1 acts indirectly on the gate through conformational changes in Orco.

The migration of VUAA1 appears to be governed by stepwise hydrophobic and hydrophilic interactions throughout the ingress of the ligand towards the cradle of the binding site. This process can be divided into four distinct steps, summarized in Figure 3C (and Figure S2 for more details). In each step, the desolvation of VUAA1 increases (Figure 3B), playing an essential role in the progression of the molecule toward the binding site. During its progression toward the binding site from the position b to d (Figure 3C), VUAA1 is mostly orthogonal to the membrane plane (position c). In addition to the desolvation process, the flexibility of the molecule appears to greatly facilitate the migration of VUAA1. Thus, during its progression toward the internal cavity, VUAA1 adopts several conformations to adapt to the local constraints, which allow the entrance into the protein either by its pyridine or its phenylethyl moiety.

In the simulations, VUAA1 is stabilized by a subset of residues and must overcome an energetic barrier to reach the next metastable intermediate state. Several residues were identified as interacting with VUAA1 during its penetration into DmelOrco. A comprehensive list of these residues is provided as supplementary information (see file supdata_frequencies.xlsx). The initial binding event occurs through a contact between VUAA1 and Y390^{S6} at the extracellular end of S6 (Figure 3C position a). Starting from this position, VUAA1 makes regular contacts with the residue side chains (Figure 3C position b) and undergoes a large desolvation process upon its entry into the receptor bundle (Figure 3C position c). The ligand then establishes additional contacts with I79^{S2}, T80^{S2}, W150^{S3}, I181^{EL2}, V206^{S4}, K373^{S5} and Y397^{S6}, where it pauses for several nanoseconds (Figure 3C position c). The ligand finally enters the cavity (Figure 3C position d) that was previously identified in the structures AbakOrco and MhOR5, and in the model of DmelOrco (Figure 1D). The final position of VUAA1 in the cavity is parallel to the membrane, and it interacts with F83^{S2}, F84^{S2}, S146^{S3}, M210^{S4} and Y400^{S6}, similar to the position of the eugenol molecule in the MhOR5 structure (Figure S3).

These results further guided site-directed mutagenesis experiments combined with functional assays to assess the critical role of residues identified as interacting with VUAA1 during the simulations.

Site directed mutagenesis and electrophysiological characterizations confirm the entry mechanism of VUAA1.

To experimentally assess the functional role of residues that significantly interacted with VUAA1 in the simulations, different mutants were designed. The influence of the volume or the physicochemical properties of their side chains was evaluated according to the response of Orco to stimulation by VUAA1. To facilitate or block the translocation process of VUAA1, the residues were mutated to smaller (alanine) or larger (tryptophan) residues, respectively. For disrupting hydrophobic interactions or hydrogen bonds between side chains and VUAA1, mutations were made in a small hydrophilic residue with a hydroxyl group (serine) or in "non-hydroxylated tyrosine" (phenylalanine), respectively. To invert the charge at position 373^{S5}, the lysine was mutated in a negatively charged glutamate. The response to VUAA1 of each mutant was assessed by electrophysiological recordings with the TEVC method.

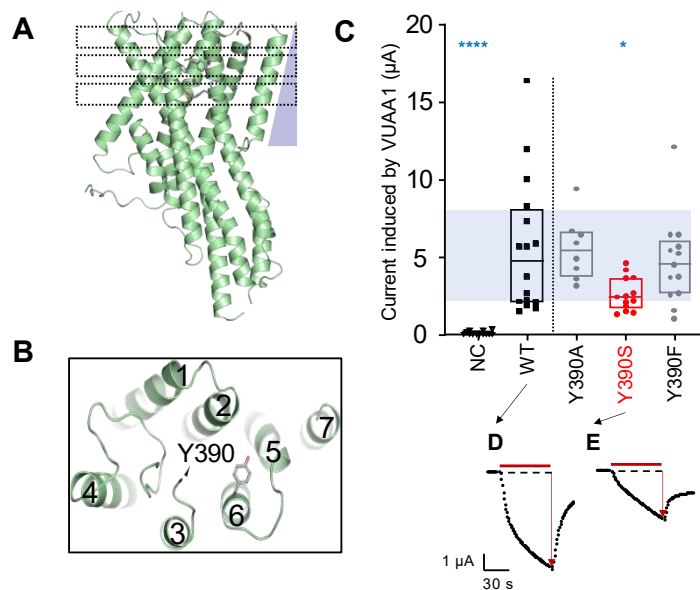


Figure 4: Results from electrophysiological measurements (entry). A. Side view of an Orco protomer. 15 Å cross-sections of the monomer as indicated on the ribbon representation of DmelOrco.

The position of the pore is indicated by a blue triangle. **B.** selected region with greater details. **C.** Boxplot showing the current induced by 500 μ M of the ligand VUAA1 and measured by TEVC recordings on DmelOrco WT and mutants expressed by the *Xenopus* oocyte. n= 8 minimum recordings from different oocytes; NC: Negative Control (non-injected oocytes). Data are analysed with one-way ANOVA with α -error =0.05 followed Dunn's post-hoc test. Results from mutant showing a statistical decrease from WT are coloured in red. **D-E.** Representative current measured on WT and mutant with statistical difference from DmelOrco WT.

The simulations revealed that Y390^{S6} is the first residue that has a significant interaction with VUAA1, interacting at a frequency of 0.47 averaged over all entry trajectories. Y390^{S6} was mutated into alanine (Y390A) and phenylalanine (Y390F) and both mutations did not show significant change in the response to VUAA1 (Figure 4). Thus, the reduction of the side chain into alanine or the removal of the hydroxyl group of Y390 did not favor or abolish the action of VUAA1. Consequently, neither aromaticity nor a hydroxyl group on the aromatic ring are necessary for the interaction with VUAA1 in position 390. On the contrary, its mutation into serine led to a decrease in the activation by VUAA1 (2.44 μ A vs 4.71 μ A for the WT). A Western-blot has been performed to verify that the expression level of the Y390S mutant was similar to the WT (Sup. Fig. S3), and the semi-quantitative analysis indicated no significant differences between both constructs. This result confirmed the role of Y390 in VUAA1 activation. In particular, the differences observed between the mutants emphasize the importance of the hydrophilic character of position 390^{S6}. Introduction of a serine in place of tyrosine generates a hydrophilic environment(32) that would hamper the first step of desolvation process that is crucial for the entry of VUAA1, as observed in the simulations (Figure 3C).

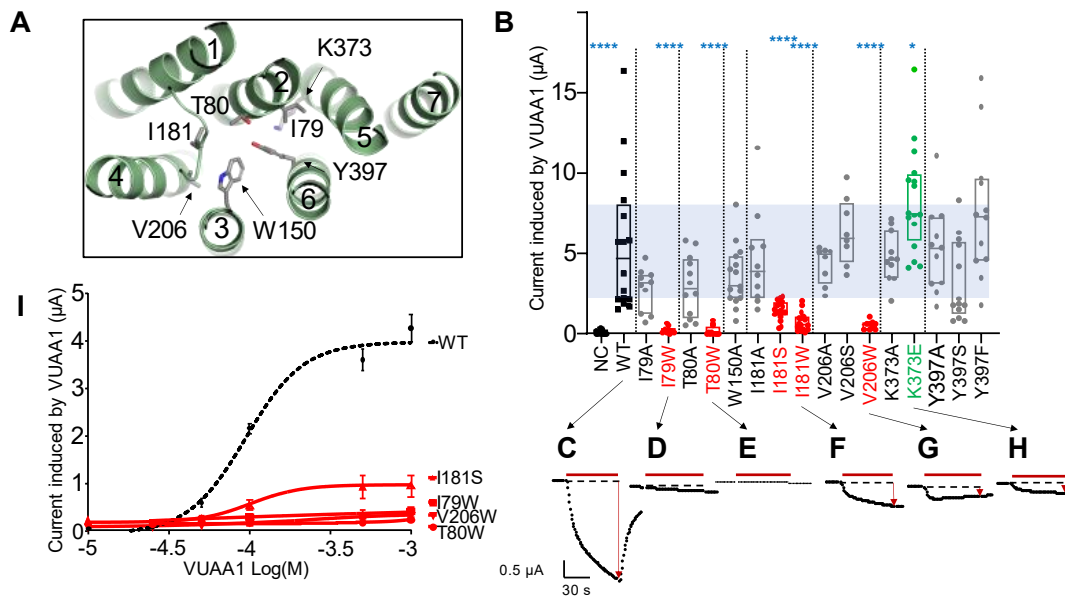


Figure 5: Results from electrophysiological measurements (vestibule). **A.** selected region with greater details. **B.** Boxplot showing the current induced by 500 μ M of the ligand VUAA1 and measured by TEVC recordings on DmelOrco WT and mutations expressed by the *Xenopus* oocyte. n= 8 minimum recordings from different oocytes; NC : Negative Control. Data are analysed with one-way ANOVA with α -error =0.05 followed by Dunn's post-hoc test. Results from mutations showing statistical decrease or increase from WT are coloured in red or green respectively. **C-H.** Representative current measured on mutants with statistical differences from DmelOrco WT. **I.** Dose-response curves for the mutants considered.

When going deeper in the protein, VUAA1 has shown high frequencies of interaction with a planar section of seven residues (I79^{S2}, T80^{S2}, W150^{S3}, I181^{EL2}, V206^{S4}, K373^{S5} and Y397^{S6} interacting with VUAA1 at frequencies of 0.44, 0.56, 0.68, 0.17, 0.11, 0.70 and 0.70, respectively, averaged over all entry trajectories) (Figure 5A). Mutations into alanine of all six residues did not significantly affect the amplitude of activation induced by VUAA1 (Figure 5B), indicating that the side chains of these residues are not critical or involved in a limiting step for the interaction with VUAA1. In contrast, mutations of the non-aromatic residues in the bulkier tryptophane significantly reduced or abolished the activation by VUAA1 (Figure 5B-H, red dots). Western-blot results (Sup. Fig. S3) showed a decrease of expression of T80W and V206W but not total. These results suggest that these mutations not only affected the expression level of the mutants but also the response to VUAA1. In contrast, mutations I181S and I181W showed an increase of expression in Western-blot results, but still a clear loss of VUAA1

activation confirming that the ability of VUAA1 to access this region is critical for the channel response.

As these residues are pointing into the simulated ligand pathway, these functional results reinforce those obtained by MD simulations that showed the implication of these residues in the entry of VUAA1. Interestingly, inserting the hydrophilic and shorter serine residue in place of the hydrophobic I181^{EL2} (I181S), significantly reduced the amplitude of activation (1.45 vs 4.71 μ A for the WT) (Figure 5B) as previously observed with Y390S mutant. This deleterious effect of the mutation into serine is site specific since the similar mutation of Y397^{S6} (Y397S) showed no significant effect on VUAA1 response (Figure 5B). Mutation of the only charged residue identified in the simulations (K373^{S5}) generated unexpected responses. Thus, mutation of K373^{S5} into Alanine (K373A) that profoundly modifies the physico-chemical properties by reducing the size of the side chain and by removing the positive charge, did not change the response to VUAA1 (Figure 5B). Inversion of the charge by mutation K373^{S5} into glutamate (K373E) did not abolish the response but increased it (8.29 μ A), potentially by decreasing the polarity of the binding cavity (Figure S5). Western-blot results confirmed the K373E was not overexpressed. All mutations made at position Y397^{S6} did not significantly change the amplitude of activation induced by VUAA1 (Figure 4). In the simulations, VUAA1 is in transit in this section of seven residues, and move on to a deeper cavity, which would constitute the binding site.

Site directed mutagenesis and electrophysiological characterization confirm the binding site of VUAA1.

In the deeper section, five residues were identified in the simulations to frequently interact with VUAA1 and formed a cavity suspected to be the binding site (supdata_frequencies.xlsx). The five positions F83^{S2}, F84^{S2}, S146^{S3}, M210^{S4} and Y400^{S6} (respectively interacting with VUAA1 at a frequency of 0.32, 0.02, 0.23, 0.19, 0.24, averaged on all entry trajectories) were mutated to defined more precisely the cradle of the VUAA1 binding cavity (Figure 6). Using the site-directed mutagenesis approach, all the six residues were mutated in alanine and tryptophane (Figure 6B) to reduce or increase the steric hindrance of the side chains, respectively. The

Western-blot results (Sup. Fig. 3) demonstrated that all mutants of the Fig. 6 were expressed at similar levels.

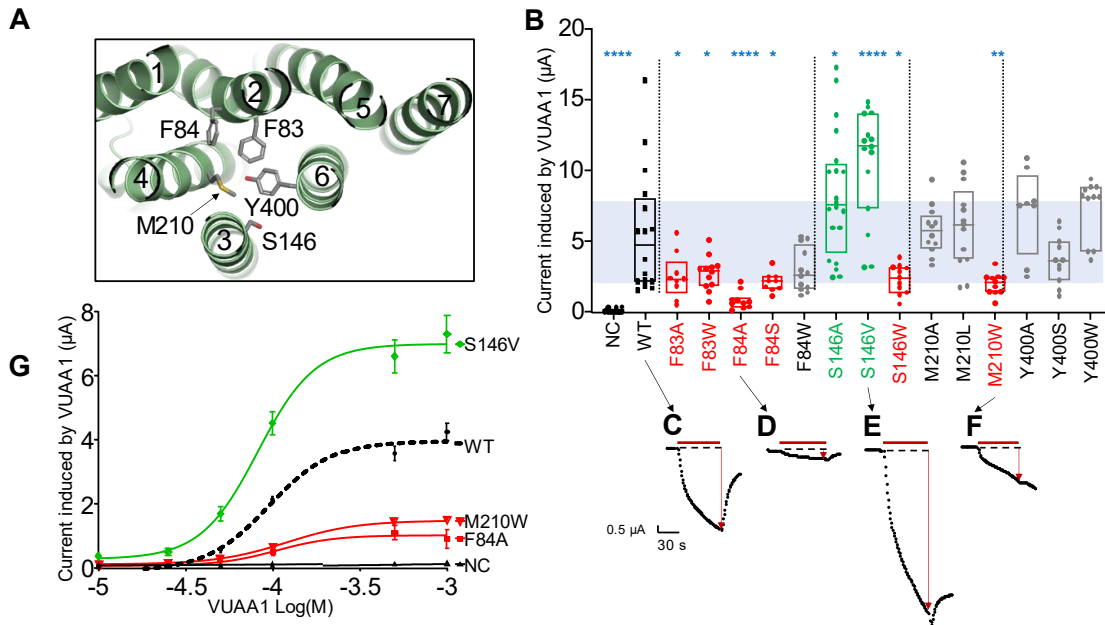


Figure 6: Results from electrophysiological measurements (binding site) **A.** selected region with greater details. **B.** Boxplot showing the current induced by 500µM of the ligand VUAA1 and measured by TEVC recordings on DmelOrco WT and mutants expressed by *Xenopus* oocyte. n= 8 minimum recordings from different oocytes; NC: Negative Control. Data are analysed with one-way ANOVA with α -error =0.05 followed by Dunn’s post-hoc test. Results from mutant showing a statistical decrease or increase from WT are coloured in red or green respectively. **C-F.** Representative current measured on mutants with statistical differences from DmelOrco WT. **G.** Dose-response curves for the mutants considered.

In contrast to previous results, mutation in alanine of two phenylalanines (F83A and F84A) decreased the response to VUAA1 (Figure 6B) with a greater extent for F84A (medians: 2.33, 0.67, 4.71 µA for F83A, F84A and WT, respectively). Mutation in tryptophane induced the same phenotype in position 83 (F83W) (Figure 6B-D), while the mutation in serine had the same impact in position 84 (F84S) (Figure 6B). Finally, the mutation F84W did not induce a significant change compared to the WT (Figure 6B). These results indicate that these two adjacent phenylalanine residues play a critical role in the activation by VUAA1, but with different characteristics. Position 83 must be a phenylalanine and cannot be replaced by a homologous residue like tryptophane, while position 84 is more tolerant to replacement by

tryptophane but much less to alanine and serine. The peripheral position of F84^{S2} in the cavity could explain this selective tolerance to large hydrophobic residues, while the central position of F83^{S2} in the cavity suggests a larger and more specific interaction with the ligand. These results are in agreement with those of Corcoran et al. (25), showing that F83^{S2} is one of the essential residues for the action of VUAA1.

On the opposite side of the cavity, S146^{S3} is also pointing toward the binding cavity. Mutation of this hydrophilic residue induced a unique phenotype of increased response to VUAA1 when mutated in alanine (medians: 7.55 μ A vs 4.71 μ A for WT). This effect is strengthened by the introduction of the bulkier and more hydrophobic residue, valine (32) (median: 11.68 μ A) (Figure 6B&E). This mutation S146V showed the highest response to VUAA1 and could be used in further studies to increase the amplitude of the response.

Mutations of M210^{S4} in shorter alanine (M210A) or leucine (M210L) residues did not change the response to VUAA1 (5.78 and 6.04 μ A, respectively vs 5.56 μ A for WT), while the mutation in the bulkier tryptophane induced a significant decrease in the amplitude of activation (2.00 μ A) (Figure 6B&F). Consequently, the methionine 210 that is in close proximity to F83^{S2} and F84^{S2} does not specifically interact with VUAA1, but this position does not tolerate steric hindrance.

Mutation of Y400^{S6} in either alanine, serine or tryptophane did not significantly change the response to VUAA1. Despite the high conservation of Y400, this result is consistent with the position of the residue, located deeply in the core of the protein, so its mutation is unlikely to change the properties of the binding cavity.

Concentration-effect curves performed on mutants with the most significant results (Figure 6G and Table S1) showed a change in E_{max} that was either negative (for F84A and M210W) or positive (for S146V), without affecting the EC₅₀. These results suggest a dominant effect of the mutations on the efficacy of VUAA1.

Western blot results show that mutants with a significant gain or loss of function are always expressed. Moreover, the level of expression has no correlation with the mutant response to VUAA1 (Figure S4).

Discussion

The simulations of VUAA1 binding onto DmelOrco were carried out with little knowledge about the amino acids involved in the mechanism of binding. They revealed both the entry pathway and the binding site in a model of Orco that is highly conserved across species. The binding mode of VUAA1 depends on interactions mediated by different hydrophobic, aromatic, and hydrophilic residues within the pocket. Finally, the experimental results confirmed the predictions made by the simulations and allowed the identification of residues critical for the entry of VUAA1 and for its binding in a deeply embedded cavity that is also found in the structures of AbakOrco and MhOR5.

The translocation of VUAA1 through the protein is highly conserved among Orcos.

The analysis of the sequence conservation reveals that the pathway followed by VUAA1 to reach the binding site of DmelOrco is highly conserved (Figure 7). As Orcos are known to recognize a remarkably restricted number of ligands, the high conservation of the translocation pathway can be interpreted as a molecular sieve, which filters the entrance of ligands to the binding site. These residues show a high conservation in Orco and are likely to be crucial for initiating the opening of the channel upon ligand binding. In contrast, ORs that recognize a large diversity of ligands (33-35) show a low conservation at similar positions. The chemical variation observed in residues that line the translocation pathway in ORs allows a large diversity of odorants to diffuse inside the protein and reach their binding sites.

Desolvation of VUAA1 is fundamental for its entry.

The recent advances in structural biology have led to greater insight into the role of desolvation in the thermodynamics and kinetics of binding (36-38). The importance of hydrophobic interactions as a ligand-desolvation penalty or a driving force for the induced fit of receptors is a long-term challenge in drug design (39, 40). In particular, it has been shown that water plays

a crucial role in the binding kinetics (41). The binding process of VUAA1 to Orco is accompanied by a desolvation at each metastable state. The most important decrease in the number of water molecules in the first solvation shell is observed when VUAA1 enters the protein. Accordingly, the mutation of the hydrophobic Y397^{S6} to a hydrophilic serine decreased the response of DmelOrco to VUAA1, most probably by preserving water molecules around VUAA1. Our results also suggest that I181^{EL2} could be involved in the desolvation process required for entry into the transmembrane core of Orco, which would explain why no continuous translocation pathway is observed in the structures of the apo state of AbakOrco and MhOR5. We further assessed the hydrophobic match between VUAA1 and Orco over the translocation pathway. The results show an increase in this complementarity during the ligand translocation process (Table S2). These observations suggest that progressive desolvation of VUAA1 during its entry into Orcos is a fundamental process not only for reaching the binding site but also for the high selectivity of Orcos for VUAA1.

Comparative analysis of the eugenol-bound MhOR5 structure (pdb: 7LID) with our VUAA1-bound DmelOrco model revealed a shared binding site position with a high conservation (16 amino acid pocket: 50 % identity, 62.5 % similarity; 24 amino acid pocket: 33 % identity, 62.5 % similarity) (Figure 7). However, the ORs show a remarkable diversity in the binding site composition. This particularity is also found in mammalian ORs, allowing for a broad detection of chemicals (42-44).

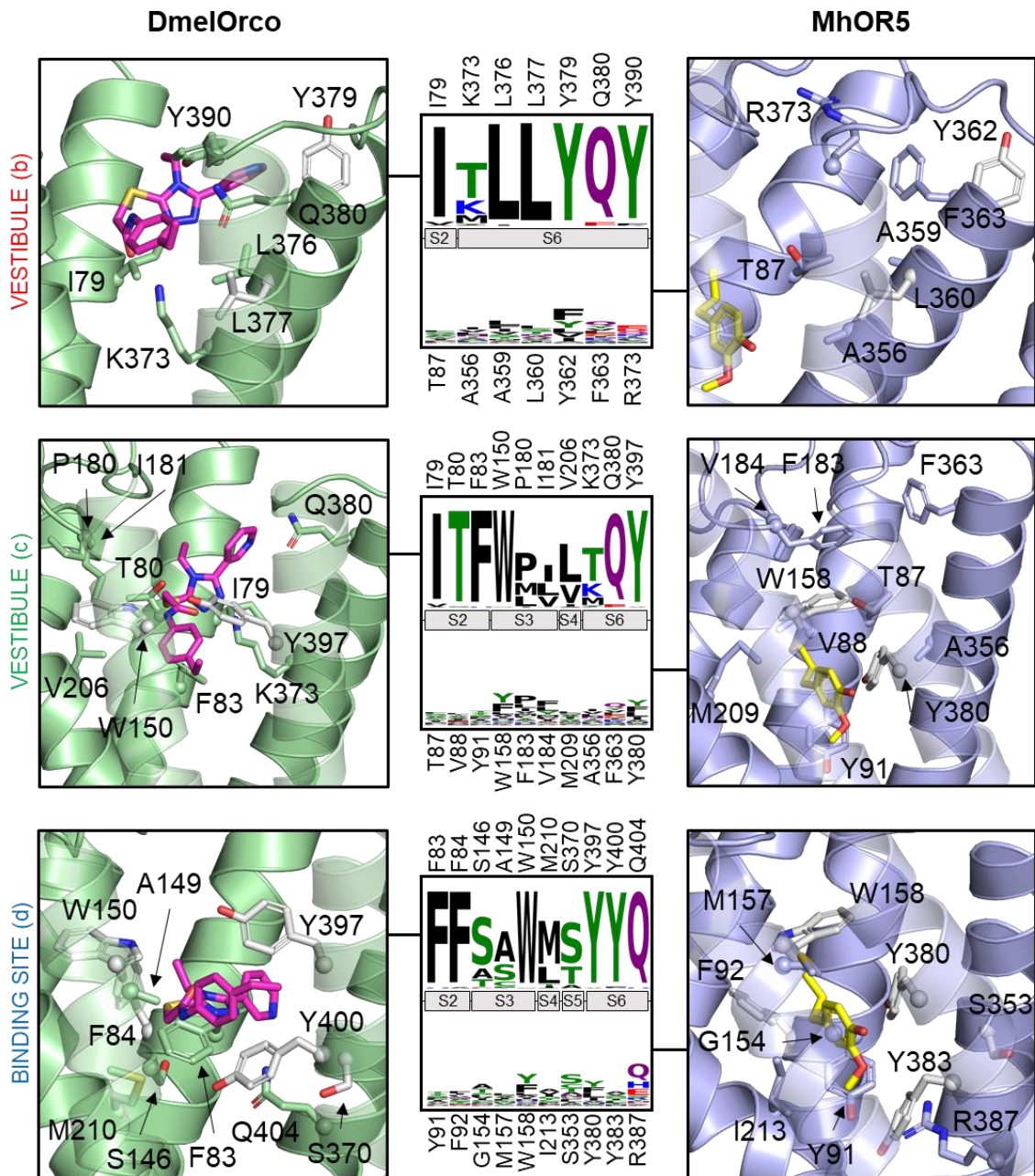


Figure 7: Sequence logos and molecular details of amino acids involved in the translocation pathway (vestibule b and c) and binding site in Orco and OR (d). The amino acids were selected according to their frequencies of interaction with VUAA1. Residue conservation among 176 Orcos from 174 species and 361 ORs from 4 species are coloured according to their side-chain chemistry. Carbon atoms from amino acids conserved between Orcos and ORs are coloured in white. Carbon atoms from VUAA1 and Eugenol are shown in purple and yellow, respectively.

The polarity of the ligand binding cavity influences the efficacy of VUAA1.

The polarity of the binding cavity appears to have a pronounced influence on the channel response to VUAA1: a decrease induces a gain of function while an increase leads to a loss of function (Figure S5). We further investigate this observation by evaluating the polarity of 176 Orcos from 174 species. This analysis reveals that the binding cavity of the VUAA1-insensitive *MdesOrco* is more polar than the responsive Orcos. When Corcoran et al. (25) replaced the hydrophilic H81^{S2} from *MdesOrco* by a more hydrophobic phenylalanine (H81F), it induced a response to VUAA1. In contrast, mutations that increased the polarity of the binding cavity abolished the response to VUAA1 in *AsegOrco*. The polarity of the cavity seems a good indicator to predict the response to VUAA1 of a given Orco or mutant (Figure S5).

The volume of the cavity also influenced the response of Orco to VUAA1 (Figure S6). A substantial reduction of the volume (such as the introduction of a tryptophane residue, in position F83^{S2}, S146^{S3}, V206^{S4} or M210^{S4}) induced a significant decrease in the response to VUAA1. In contrast, mutations that increased the volume of the cavity did not rationally induce a change in the response to VUAA1. An exception was the mutation F84A, which abolished the response to VUAA1, potentially due to an indirect interaction with the ligand. These results suggest that the protein could fluctuate to accommodate bulky ligands such as VUAA1, as has already been shown for olfactory receptors (6, 42, 43).

The architecture of the ligand binding site is conserved between Orcos and ORs

Once in the binding cavity, VUAA1 is stabilized by a combination of hydrophobic h-bond, Van der Waals and pi-stacking interactions and does not move back into the bulk within the simulation time. Our results highlight the importance of two residues from segment 2 (F83 and F84) to form the binding site. Mutations made at similar positions in *MhOR1* (Y106^{S2}A, I107^{S2}A) and *MhOR5* (Y91^{S2}A and F92^{S2}A) result in non-responsive channels (Table S3).

In this final pose, VUAA1 remained in the same orientation, with the ethyl phenyl moiety located between the helices S3 and S4 and the pyridine next to S2 and S5 (Figure 7 and Figure S3). This conclusion is strengthened by the increased sensitivity of S146^{S3} mutants (alanine and

valine). Interestingly, decreasing the ethyl moiety to a methyl almost abolished the response of Orco (17). In contrast, the replacement of the ethyl group by an isopropenyl one improved the potency of the VUAA1-derivative. All together, these observations show that increasing hydrophobicity by mutations or ligand modifications increases the response of Orco to its ligands.

To explore the potential binding modes of VUAA1, the ligand was structurally modified to VUAA2, VUAA3 and VUAA4, which display significant greater potency, and in VUAA0.5, which is less potent than VUAA1. The calculated hydrophobic and electrostatic matches between the ligands and the receptor correlated with the functional results. We further evaluated the orientation of VUAA1 by manually flipping it into the cavity and build the VUAA analogues (Figure S7). Scores of both electrostatic and hydrophobic matches were inferior to those of the initial binding mode (Table S3), suggesting that the initial orientation is preferred. In conclusion, this study revealed the translocation pathway and binding site of VUAA1 into DmelOrco using a combination of dynamic simulations and functional characterization. The results highlight the role of desolvation for the progression of the ligand, the role of the polarity of the binding cavity in the efficacy of VUAA1 and the lower limit of size of the cavity for VUAA1 binding. This study shows that the binding pocket location is conserved between ORs and Orcos. The striking difference between the two families is the high level of sequence conservation of the migration pathway and binding pocket observed in Orco compared to the high diversity in ORs. The conservation and the variability are then shared in between the two subunits forming the heterodimer. This combination of the highly conserved Orcos subunit with the more versatile ORs provides the insect with extremely high chemical discrimination power. Orcos have been shown to play a fundamental role in insect behavior such as foraging and oviposition and are thus a potential target for the development of behaviorally disruptive chemicals.(45, 46) Our results provide a fine description of the binding process, opening the way to a rational design of orthosteric and allosteric modulators.

Materials and Methods

In silico modelling

Alignment between Orcos and ORs with MhOR5

Alignment between MhOR5 and Orcos was based on the alignment files for 176 Orcos and 361 ORs from Butterwick et al.(5). MhOR5 was realigned with the Orcos using ClustalO(47) with default settings, then optimized by hand to conserve the existing alignment. The same process was applied for the ORs.

Orco modelling

The 176 Orcos tetramer models plus *DmelOrco* WT and mutants were generated by SWISS-model pipeline(48) using PDB 6C70 as a template with default settings. *DmelOrco* alpha fold model (version 07.01.2021) was retrieved from AlphaFold Protein structure database(49). RMSD between the SWISS-model and AlphaFold model was calculated using cptraj(50) after alignment of the structures on *i*) all the sequence, *ii*) all the sequence except IL2, and *iii*) only helices.

Cavity analysis of *DmelOrco*, *AbakOrco* and MhOR5

Detection of the pockets of the 176 Orcos plus *DmelOrco* mutants (SWISS-model), *AbakOrco* (pdb: 6C70) and MhOR5 (in APO form, pdb: 7LIC) cavities was carried out using fpocket3(51) with default settings. For each receptor, visual inspection was used to identify the pocket of interest.

Molecular dynamics setup

As IL2 is not resolved in the *AbakOrco* (pdb: 6C70) template structure, IL2 was discarded from the structure of each *DmelOrco* monomer. Propka(52) was used to predict protonation states of the protein at a target pH 6.5. The *DmelOrco* tetramer orientation in its membrane was

determined using OPM server(53). Five VUAA1 molecules were added in different orientations on the extracellular side. The system was embedded into a POPC-only model membrane using PACKMOL-memgen (54). The simulation box was completed using TIP3P water molecules and neutralized using K^+ and Cl^- ions with a final concentration of 0.15 M. The total system is made up of 286736 atoms, in a $3.4 \cdot 10^6 \text{ \AA}^3$ periodic box. Molecular dynamics simulations were performed with the sander and pmemd.cuda modules of AMBER18, with the ff14SB force field for the proteins and the lipid14 forcefield for the membrane(55). VUAA1 parameters were generated by calculating partial atomic charges with the HF/6-31G* basis set using Gaussian 09(56). The obtained electrostatic potential was fitted by the RESP program(57). The other parameters were taken from the General Amber Force Field 2 (gaff2). Bonds involving hydrogen atoms were constrained using the SHAKE algorithm and long-range electrostatic interactions were handled using Particle Mesh Ewald. The cut-off for non-bonded interaction was set to 10 Å. Each system was first minimized with the AMBER sander module, with 5000 steps of steepest descent algorithm then 5000 steps of conjugate gradient with a $50 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^2$ harmonic potential restraint on the protein part of the system. A second minimization of the same length without restraint was applied. The systems were then thermalized from 100 to 310 K for 10000 steps (restraining the protein and ligands with a $200 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^2$ harmonic potential). Each system underwent 50000 steps of equilibration in the NPT ensemble and 1 bar (restraining the protein and ligands with a $15 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^2$ harmonic potential) before the production phase. During equilibration and production phase, temperature was kept constant in the system at 310 K using a Langevin thermostat with a collision frequency of 5 ps^{-1} . To increase sampling, all 5 VUAA1 molecules were constrained in a sphere of 45-55 Å radius, centered on the center of mass of the Lys486 of the four Orco monomers (with a potential of $10 \text{ kcal}\cdot\text{mol}^{-1}$). To avoid VUAA1 aggregation, each VUAA1's sulfur atom was constrained to be a minimum of 20 Å from each other with a soft potential penalty of $5 \text{ kcal}\cdot\text{mol}^{-1}$. The VUAA1 system in water only was built solvating the molecule in a 20 Å TIP3P periodic box using the gaff2 and tip3p forcefield parameters. The system was minimized with the AMBER sander module, with 500 steps of steepest descent algorithm then 500 steps of conjugate gradient, then heated incrementally from 100 to 310K for 10000 steps. The first 10 nanoseconds of the production phase were considered as equilibration and not taken into account for analysis. The system stability was evaluated from the root mean square deviation (RMSD) evolution computed on the backbone of the full system. During the 22 replicas, the receptors underwent small fluctuations ($\text{RMSD} < 3\text{\AA}$) showing that they remained correctly folded during

microsecond simulations (Figure S8). Hydration of VUAA1 was calculated using the pytraj watershell function.

Minimum distance between VUAA1 and eugenol for all trajectories

The minimum distance between VUAA1 and eugenol was calculated for all trajectories by structurally aligning MhOR5 (pdb: 7LID) on each DmelOrco monomer using the cealign pymol command(58), then calculating the center of mass distance between eugenol and VUAA1 on each trajectory using the mindist pytraj module(50).

Selection of representative frames for b, c, and d

Representative frames of the diffusion were obtained by dividing the prototypical trajectory into 4 parts according to the curve shown in Figure 3B. For each part, a frequency analysis between VUAA1 and the receptor using the get_contacts module (<https://getcontacts.github.io/>) identified the critical residues. These residues, plus VUAA1, were selected and used to cluster each part by kmeans clustering, using cpptraj(50) with a fixed number of 4 clusters. The representative frame of the largest cluster was then extracted as the representative frame of that part of the trajectory.

Electrostatic and hydrophobic complementarity

For each representative frame (b, c and d), the protein was extracted with VUAA1 which was then replaced with VUAA0.5, VUAA2, VUAA3 and VUAA4 (17). For the representative frame of the binding site (d), VUAA1 was also manually flipped over using the pair fitting tool in PyMol, and then replaced again with VUAA0.5, VUAA2, VUAA3 and VUAA4. Each system was then minimized using the AMBER sander module, with 5000 steps of steepest descent algorithm then 5000 steps of conjugate gradient, while restraining the backbone of the protein with a 50 kcal·mol⁻¹ potential. Hydrophobic complementarity scores for each system were calculated using the PLATINUM web server(59) with default settings. Electrostatic

complementarity scores for each system were calculated using the Flare electrostatic complementarity tool (60).

In vitro experiments

Chemicals

VUAA1(N-(4-ethylphenyl)-2-((4-ethyl-5-(3-pyridinyl)-4H-1, 2, 4-triazol-3-yl)thio)acetamide) (CAS 525582-84-7) was purchased from Sigma-Aldrich. The stock solution was 110 mM in DMSO and subsequently diluted into appropriate buffer solution.

Molecular biology

All Orco gene sequences were optimized (61) for protein expression in *Xenopus laevis* oocytes with the GenSmart™ Codon optimization Tool and subcloned into a pGEMHE-derived vector. The wildtype gene of *Drosophila melanogaster* Orco (DmelOrco) was synthesized by Genscript and subcloned with XmaI/XhoI cloning sites. Site-directed mutagenesis of DmelOrco was done by PCR with the Q5® site directed mutagenesis kit (NEB) using primers optimized with the NEBase Changer online tool and following the supplier's protocol. After transformation of commercial competent bacteria (XL10 Gold) by standard heat-shock protocol and overnight culture in ampicillin-containing LB plates, positive clones were identified by electrophoretic restriction profile and external sequencing (Genewiz). DNAs of a positive clones were amplified with Qiagen MidiPrep Kit and the ORF fully sequenced. For in vitro transcription, DNAs were linearized with restriction enzyme NotI that cuts a unique site in the 3' region of the polyA tail. The linearized DNAs were purified by the standard phenol:chloroform extraction method and transcribed into mRNA using the T7 ultra mMessage mMachine kit (Thermo Fisher Scientific). mRNAs were purified with the NucleoSpin RNA plus XS kit (Machery-Nagel). DNA and RNA were analyzed by agarose-gel electrophoresis and quantified by spectrophotometry.

Electrophysiological recordings

Xenopus laevis oocytes were prepared as previously described (62). Briefly, oocytes were defolliculated after surgical retrieval by type 1A collagenase over 2-3h under smooth horizontal agitation. They were manually selected and incubated at 19°C in modified-Barth's solution (1 mM KCl, 0.82 mM MgSO₄, 88 mM NaCl, 2.4 mM NaHCO₃, 0.41 mM CaCl₂, Ca(NO₃)₂ 0.3 mM, 16 mM HEPES, pH 7.4) supplemented with 100 U·mL⁻¹ of penicillin and streptomycin and 0.1 mg·mL⁻¹ of gentamycin. Each oocyte was micro-injected with the Nanoject instrument (Drummond) with 50 nL of 20 ng of mRNA coding for the Orco of interest. Injected oocytes were incubated individually in 96-well plates for 4 days at 19°C in the same buffer.

Whole cell currents were recorded with the two-electrode voltage-clamp (TEVC) technique with the HiClamp robot (MultiChannel System). Microelectrodes were filled with 3M KCl. The high K⁺ buffer used for recordings was composed of 91 mM NaCl, 1 mM MgCl₂, 1.8 mM CaCl₂, 5 mM HEPES, pH 7.4. Membrane voltage was clamped to -50 mV and VUAA1 was applied for 60 s. Data were extracted with M. Vivaudou's programs (63) and statistically analyzed with Prism 8 (Graphpad).

Animal handling and experiments fully conformed to European regulations and were approved by the French Ministry of Higher Education and Research (APAFIS#30915-2021040615209331 v1 to CM). The animal facility was authorized by the Prefect of Isere (Authorization #E 38 185 10 001).

Western Blots

All expression experiments were assessed on 4-20% mini-Protean TGX SDS-PAGE gels (BioRad). All oocytes loaded on gel were from the same batch and injected as described above, with 4 days of incubation.

Oocytes were homogenized through several passes in a syringe with two sizes of needles (18g then 27g) into a solubilization buffer (PBS 1X, protease inhibitor cocktail tablets) and stored at -80°C. Western blots were performed by transferring proteins onto a nitrocellulose membrane using the trans-blot turbo system (BioRad). Membranes were blocked with PBS 1x-1% non-fat milk overnight at 4°C and incubated in primary antibody anti-Orco (1:500) and the secondary antibody Goat anti-rabbit IgG HRP conjugate (1:5000 Sigma-Aldrich) for 1 hour each. The immunoblot was revealed with ECL substrate kit (Abcam) and recorded on ChemiDoc

(BioRad) at different times for identifying the optimal exposition time without pixel saturation. Gels were stained with standard Coomassie blue staining protocol and the pictures were taken with the Chemidoc apparatus. Relative intensities of bands in blots and volume of lanes in gels were determined with the Image Lab software (BioRad).

The polyclonal primary antibody anti-Orco was purchased from Genscript and designed against the peptide sequence SSIPVEIPRLPIKSFYPW in the second extracellular loop (ECL2). Anti-Orco was produced in rabbit and purified by antigen affinity.

Supplementary data

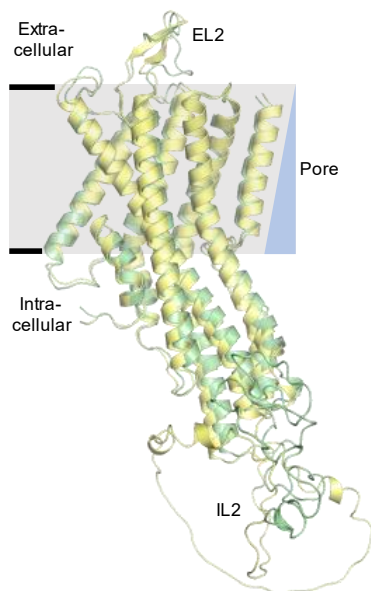


Fig. S1: Superposition of DmelOrco models. Homology model of DmelOrco built using the experimental structure of AbakOrco homotetramer (pdb ID: 6C70) as a template (in green) superposed on the AlphaFold2 model retrieved from the Alpha Fold Protein database (in yellow). The two models have an RMSD of 0.7 Å calculated on their transmembrane segments. To orient the protomer, part of the pore is shown on right as blue right-angled triangle. EL2: extracellular loop 2. IL2: intracellular loop 2.

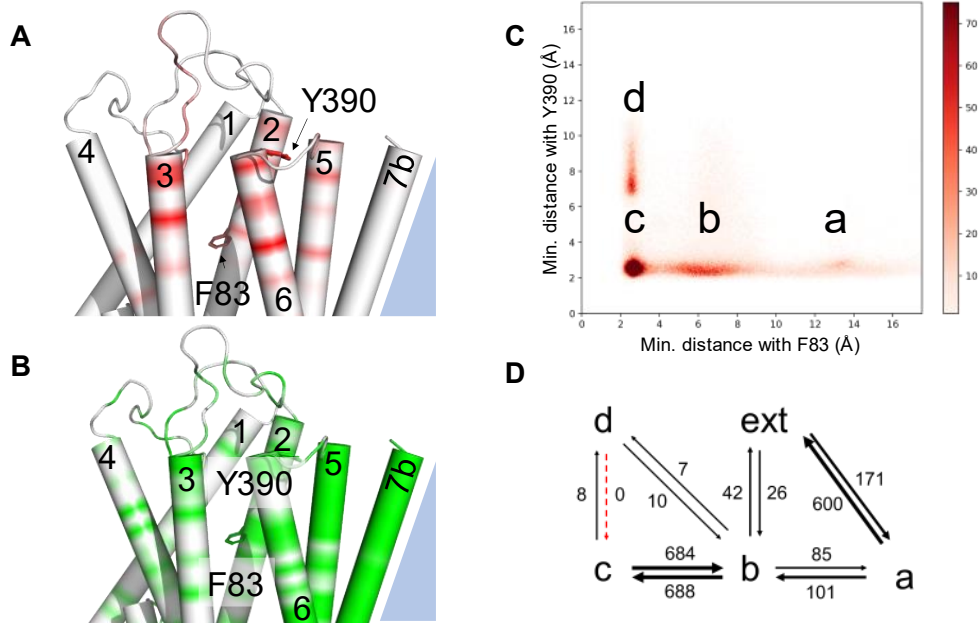


Fig. S2. **A.** Contact frequencies of VUAA1 mapped onto the structure of DmelOrco. **B.** Sequence conservation mapped onto the structure of DmelOrco. **C.** Contour map of VUAA1 migration as the minimum distance from F83S2 (distance from the entry) and minimum distance from Y390S6 (distance from the cradle of the cavity). The four basins account for the entry gate of the receptor a), two intermediate vestibules b) and c) and the sampling of the binding cavity d). **D.** Proposed mechanism of VUAA1 binding. The numbers indicate the transition from one state to another.

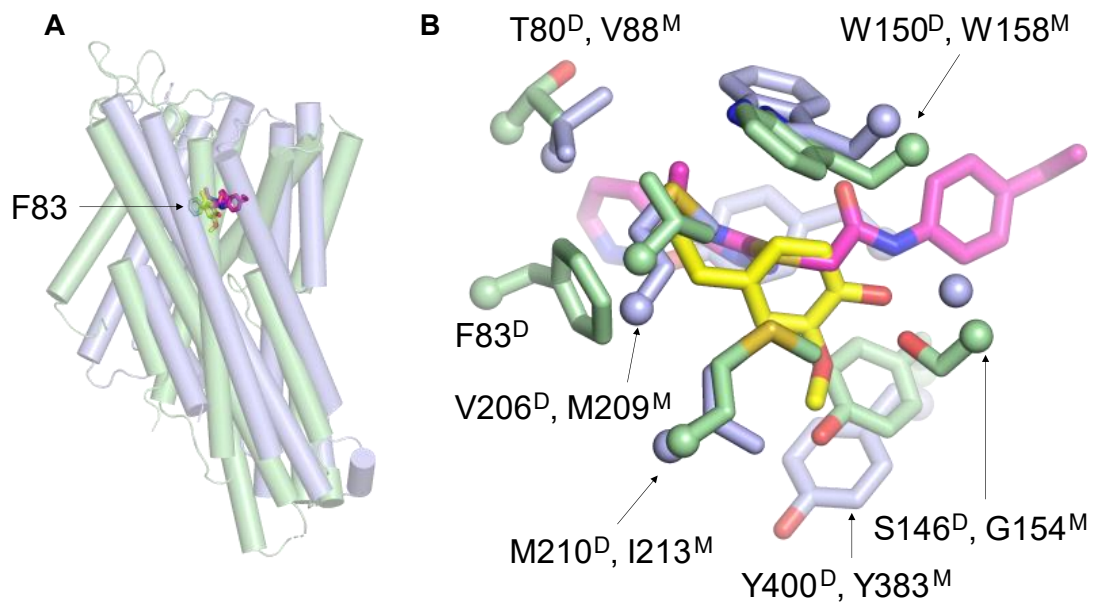


Fig. S3. A. Superposition of DmelOrco (green tubes) and MhOR5 (blue tubes), with their respective ligands VUAA1 (magenta sticks) and eugenol (yellow sticks). **B.** Close view on the binding site of DmelOrco (green licorice) and MhOR5 (blue licorice). The superscripts D or M on the amino acids names represent DmelOrco and MhOR5, respectively.

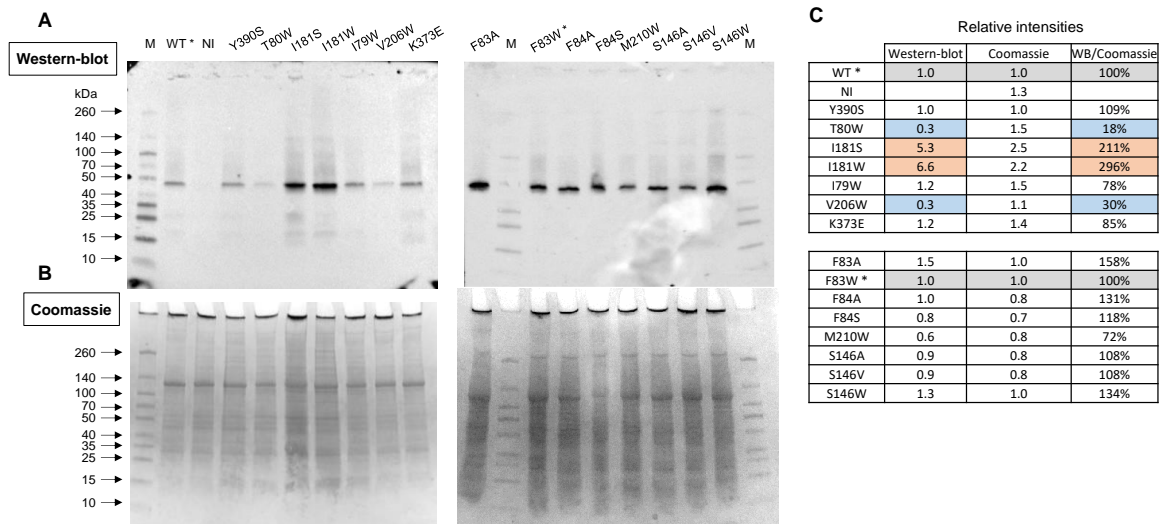


Fig. S4. A. Western blot with a polyclonal primary antibody directed against DmelOrco. The constructs indicated above the lanes were expressed in *Xenopus* oocytes and the samples are crude membrane extracts. M: Thermo Scientific Spectra Multicolor Broad Range Protein Ladder; WT*: wild-type used as reference for determining the relative intensities of bands of the first blot and gel, while the mutant F83W* was used for the second blot and gel. B. Coomassie blue –stained SDS PAGE gels (4-20%). C. Values of relative intensities of the bands corresponding to Orco (Western-blot column) in panel A and to the lanes (Coomassie) in panel B. Corrected intensities of bands relative to the intensities of lanes are indicated in the column WB/Coomassie.

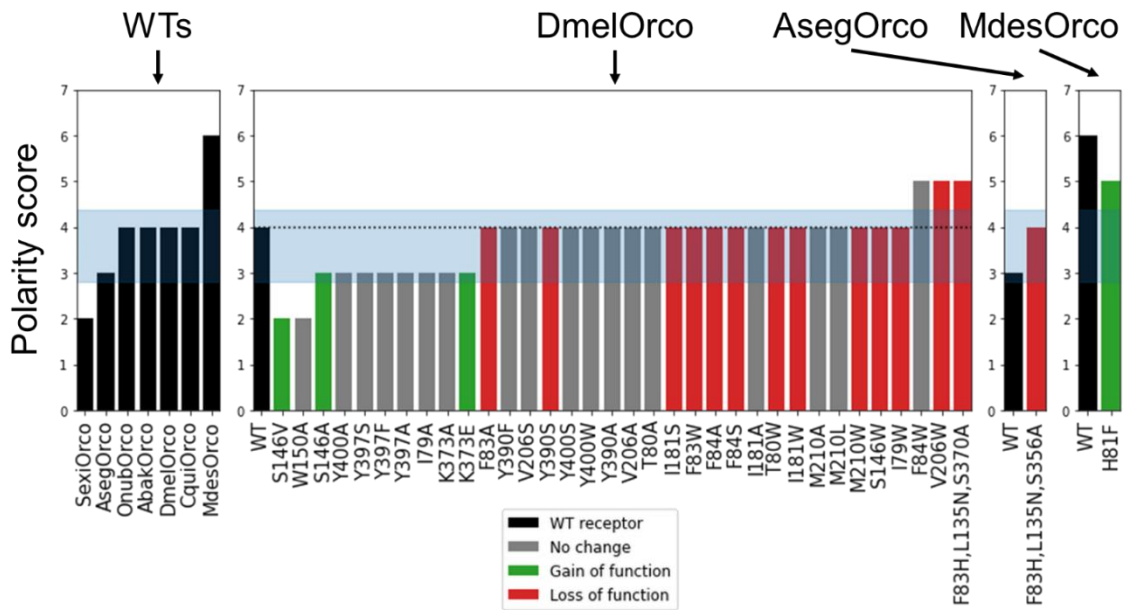


Fig. S5: Evolution of the polarity score computed with fpocket. On the left, the results for WT Orco; for clarity, only the VUAA1-responsive and VUAA1-non-responsive Orcos are shown. SexiOrco and MdesOrco are the receptors with the lowest and highest scores, respectively. The central panel gathers the results for DmelOrco mutants. The two right panel gathers results for AsegOrco and MdesOrco. The blue bands account for the mean and standard deviation among 176 Orcos from 174 species.

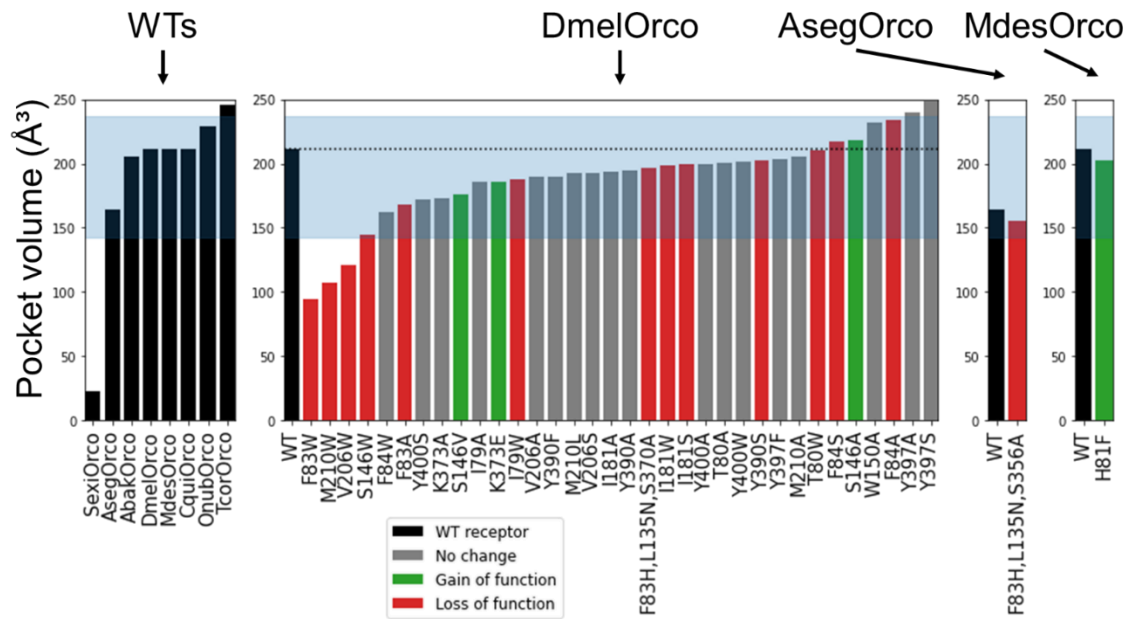


Fig. S6: Evolution of the binding cavity volume computed with fpocket. On the left, the results for WT Orcos, for more clarity, only the VUAA1-responsive and VUAA1-non-responsive Orco are shown. SexiOrco and TcorOrco are the receptors with the lowest and highest volume respectively. The central panel gathers the results for DmelOrco mutants. The two right panel gathers results for AsegOrco and MdesOrco. The blue bands account for the mean and standard deviation among 176 Orcos from 174 species.

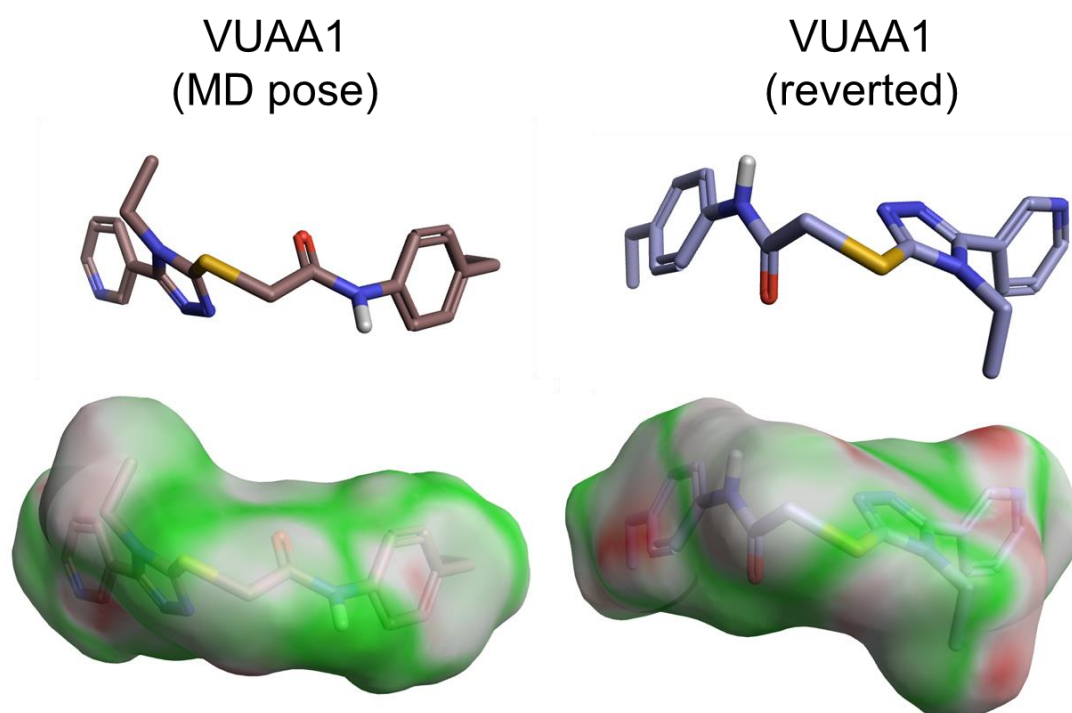


Fig. S7: Comparison of the orientation of VUAA1 in the binding site obtained during the MD simulations (MD pose) and the manually reverted orientation. The electrostatic complementarity of the ligand to the protein appears as a surface. The areas where the protein-ligand electrostatics are favorable or unfavorable are colored from green to red.

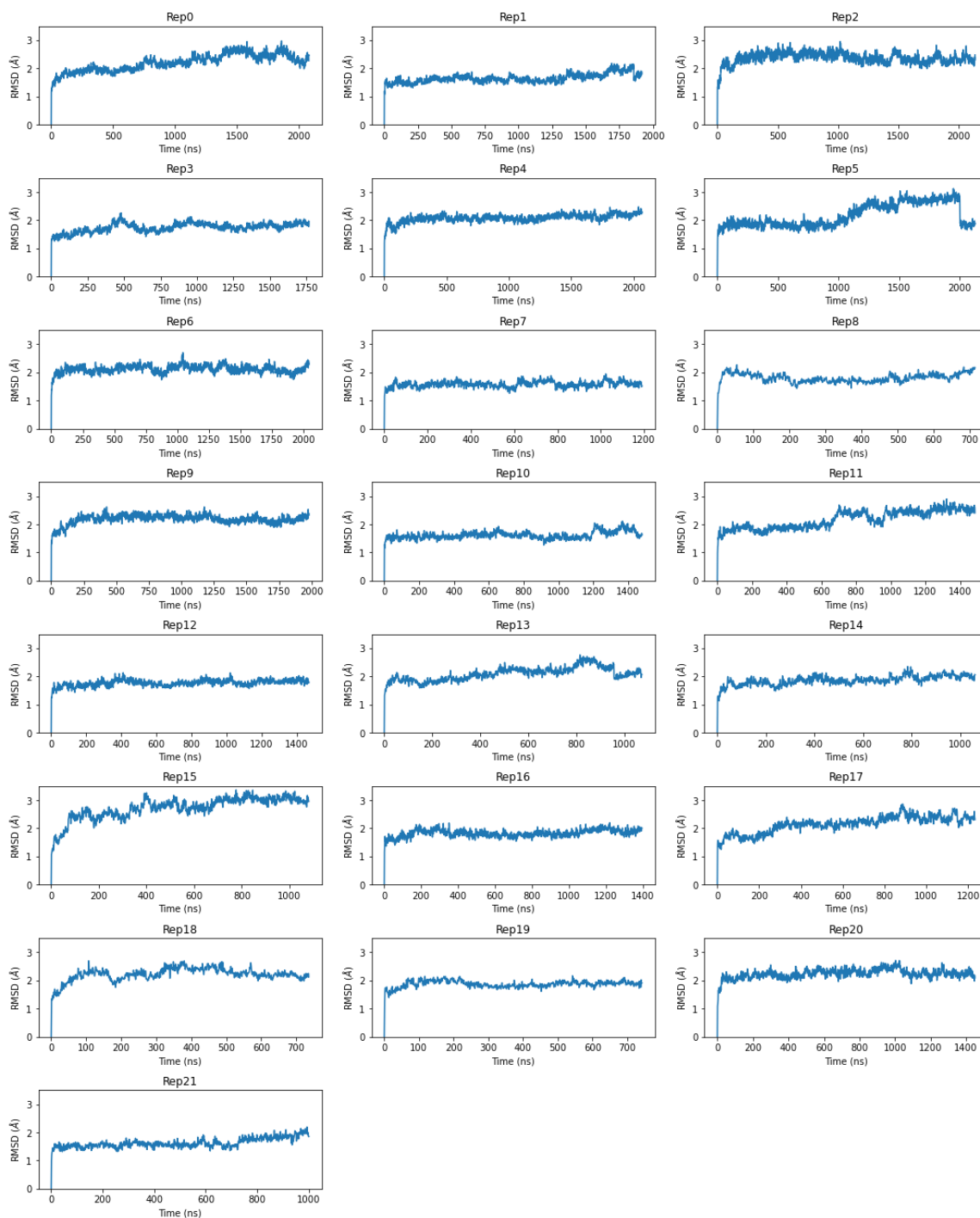


Fig. S8: RMSD of the 22 replicas of DmelOrco studied with respect to time. The RMSD is computed for the receptor backbone (CA, C, N atoms) with respect to the initial model structure.

Table S1. EC₅₀ and I_{max} of VUAA1 from WT and mutated DmelOrco. ND means Not Determined due to the absence of plateau.

	EC ₅₀ (μM)	I _{max} (μA)
WT	94.51	4.71
I181S	101.9	1.45
I79W	ND	0.96
V206W	159.2	0.61
T80W	ND	0.47
S146V	101.9	11.68
M210W	ND	2.00
F84A	159.2	0.61

Table S2: Electrostatic and hydrophobic complementary between the series of VUAA1 analogues and different stable locations in the protein according to Figure 3: vestibule (b), vestibule (c) and binding site (d). The analogues have been ranked according to the EC₅₀ measure by Taylor et. al.(1)

Location	Ligand and EC ₅₀	Hydrophobic complementarity	Electrostatic complementarity
Vestibule (b)	VUAA0.5 (0.11 mM)	0.37	0.26
	VUAA1 (37 μM)	0.41	0.29
	VUAA2 (9.2 μM)	0.44	0.31
	VUAA3 (8.4 μM)	0.46	0.33
	VUAA4 (2.1 μM)	0.45	0.32
Vestibule (c)	VUAA0.5	0.61	0.28
	VUAA1	0.62	0.30
	VUAA2	0.63	0.31
	VUAA3	0.64	0.29
	VUAA4	0.66	0.29
Binding site (d)	VUAA0.5	0.65	0.33
	VUAA1	0.64	0.34
	VUAA2	0.64	0.36
	VUAA3	0.71	0.36
	VUAA4	0.72	0.36
Binding site (d) with reversed ligand	VUAA0.5	0.58	0.28
	VUAA1	0.58	0.31
	VUAA2	0.58	0.31
	VUAA3	0.65	0.30
	VUAA4	0.64	0.32

In all the locations the trends in the hydrophobic and electrostatic complementary match that of the ligand potency. For all ligands considered, we observed an increase in the complementarity of both as the ligand moved deeper into the protein. Finally, when considering the alternative orientation of the ligand in the binding site, we observe a decrease in both complementarities compared to the other orientation.

Table S3: Comparison of the mutation effects on MhOR5, MhOR1 eugenol-induced channel response and DmelOrco VUAA1-induced channel response. A, D, I, n.a. mean Abolished, Decreased, Increased and non-affected, respectively. Results for MhOR5 and MhOR1 are taken from ref(2).

	MhOR5	Effect on Eugenol	Effect on DEET	MhOR1	Effect on Eugenol	DmelOrco	Effect on VUAA1
Ligand diffusion	T87A	D		LA02A	n.a	I79A	n.a.
	L379A	n.a.		L398A	n.a	G399	Not tested
Binding site	V88A	D		T103	A	T80A	n.a.
	Y91A	D		Y106A	A	F83A	D
	F92A	A		I107A	A	F84A	D
	S151A	A		S166A	A	S146A	I
	G154A	A		G169A	I	A149	Not tested
	W158A	A		W173A	A	W150A	n.a.
	M209A	A	D	L227A	D	V206A	n.a.
	M209V	D	I			V206S	n.a
	M209L	D	D			V206W	A
	I213A	A	A	M231A	A	M210A	n.a.
	I231M	D	A			M210W	D
	L379A	n.a.		L398A	n.a	G399	Not tested
	Y380A	A		Y339A	A	Y400A	n.a.
Y383A	A		C402A	n.a	A403	Not tested	

Dataset S1 (Supdata_frequencies.xlsx). Contact frequencies between VUAA1 and amino-acids from Orco during the MD simulations. The first sheet gathers the contact frequencies for the simulations in which the ligand sampled the binding site. The second sheet gather the frequency when the ligand visits cavity b, c or d but do not reach the binding site.

Dataset S2 (Supdata_conservation.xlsx). Sequence alignment of different Orco. The known mutations are indicated in color on the sequence.

Acknowledgments

This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 682286) to C. J. M. This work was supported by the Fondation Roudnitska under the aegis of Fondation de France to J.P. This work was also supported by the CNRS MITI programs: "Modélisation du vivant" and "Biomimétisme".

Authors thank Hervé Pointu, Soumala-maya Bama Toupet and Charlène Caloud for the management and the maintenance of *Xenopus* and acknowledge the platform supported by GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS(ANR-17-EURE-0003). They thank Michel Vivaudou for the development of software for data analysis (63). IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA).

References

1. C. Mora, D. P. Tittensor, S. Adl, A. G. Simpson, B. Worm, How many species are there on Earth and in the ocean? *PLOS Biol.* **9** (2011).
2. B. S. Hansson, M. C. Stensmyr, Evolution of insect olfaction. *Neuron* **72**, 698-711 (2011).
3. J. Bohbot *et al.*, Multiple activities of insect repellents on odorant receptors in mosquitoes. *Med. Vet. Entomol.* **25**, 436-444 (2011).
4. A. Ray, Reception of odors and repellents in mosquitoes. *Curr. Opin. Neurobiol.* **34**, 158-164 (2015).
5. J. A. Butterwick *et al.*, Cryo-EM structure of the insect olfactory receptor Orco. *Nature* **560**, 447-452 (2018).
6. J. Del Marmol, M. A. Yedlin, V. Ruta, The structural basis of odorant recognition in insect olfactory receptors. *Nature* **597**, 126-131 (2021).

7. K. Sato *et al.*, Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* **452**, 1002-1006 (2008).
8. D. Wicher, F. Miazzi, Functional properties of insect olfactory receptors: ionotropic receptors and odorant receptors. *Cell Tissue Res.*, 1-13 (2021).
9. L. B. Vosshall, A. M. Wong, R. Axel, An olfactory sensory map in the fly brain. *Cell* **102**, 147-159 (2000).
10. A. A. Dobritsa, W. v. d. G. van Naters, C. G. Warr, R. A. Steinbrecht, J. R. Carlson, Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* **37**, 827-841 (2003).
11. R. J. Pitts, C. Liu, X. Zhou, J. C. Malpartida, L. J. Zwiebel, Odorant receptor-mediated sperm activation in disease vector mosquitoes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2566-2571 (2014).
12. P. L. Jones, G. M. Pask, D. C. Rinker, L. J. Zwiebel, Functional agonism of insect odorant receptor ion channels. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8821-8825 (2011).
13. H. M. Robertson, C. G. Warr, J. R. Carlson, Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14537-14542 (2003).
14. P. Brand *et al.*, The origin of the odorant receptor gene family in insects. *elife* **7**, e38340 (2018).
15. S. Chen, C. W. Luetje, Identification of new agonists and antagonists of the insect odorant receptor co-receptor subunit. *PLOS ONE* **7**, e36784 (2012).
16. P. L. Jones *et al.*, Allosteric antagonism of insect odorant receptor ion channels. *PLOS ONE* **7**, e30304 (2012).
17. R. W. Taylor *et al.*, Structure–activity relationship of a broad-spectrum insect odorant receptor agonist. *ACS Chem. Biol.* **7**, 1647-1652 (2012).
18. I. M. Romaine *et al.*, Narrow SAR in odorant sensing Orco receptor agonists. *Bioorg. Med. Chem. Lett.* **24**, 2613-2616 (2014).
19. S. Chen, C. W. Luetje, Phenylthiophenecarboxamide antagonists of the olfactory receptor co-receptor subunit from a mosquito. *PLOS ONE* **8**, e84575 (2013).
20. G. M. Pask, Y. V. Bobkov, E. A. Corey, B. W. Ache, L. J. Zwiebel, Blockade of insect odorant receptor currents by amiloride derivatives. *Chem. Senses* **38**, 221-229 (2013).

21. Y. Bobkov, E. Corey, B. Ache, An inhibitor of Na⁺/Ca²⁺ exchange blocks activation of insect olfactory receptors. *Biochem. Biophys. Res. Commun.* **450**, 1104-1109 (2014).
22. T. Nakagawa, M. Pellegrino, K. Sato, L. B. Vosshall, K. Touhara, Amino acid residues contributing to function of the heteromeric insect olfactory receptor complex. *PLOS ONE* **7**, e32372 (2012).
23. B. N. Kumar *et al.*, A conserved aspartic acid is important for agonist (VUAA1) and odorant/tuning receptor-dependent activation of the insect odorant co-receptor (Orco). *PLOS ONE* **8**, e70218 (2013).
24. R. M. Turner *et al.*, Mutational analysis of cysteine residues of the insect odorant co-receptor (Orco) from *Drosophila melanogaster* reveals differential effects on agonist-and odorant-tuning receptor-dependent activation. *J. Biol. Chem.* **289**, 31837-31845 (2014).
25. J. A. Corcoran, Y. Sonntag, M. N. Andersson, U. Johanson, C. Löfstedt, Endogenous insensitivity to the Orco agonist VUAA1 reveals novel olfactory receptor complex properties in the specialist fly *Mayetiola destructor*. *Sci. Rep.* **8**, 1-13 (2018).
26. M. Jaiteh, I. Rodríguez-Espigares, J. Selent, J. Carlsson, Performance of virtual screening against GPCR homology models: Impact of template selection and treatment of binding site plasticity. *PLoS Comput. Biol.* **16**, e1007680 (2020).
27. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
28. R. O. Dror *et al.*, Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13118-13123 (2011).
29. J. Topin, J. Diharce, S. b. Fiorucci, S. Antonczak, J. r. m. Golebiowski, O₂ migration rates in [NiFe] hydrogenases. A joint approach combining free-energy calculations and kinetic modeling. *J. Phys. Chem. B* **118**, 676-681 (2014).
30. R. O. Dror *et al.*, Identification of two distinct inactive conformations of the β 2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4689-4694 (2009).
31. C. A. de March *et al.*, Odorant receptor 7D4 activation dynamics. *Angew. Chem. Int. Ed.* **130**, 4644-4648 (2018).
32. W. Hoffmann *et al.*, An intrinsic hydrophobicity scale for amino acids and its application to fluorinated compounds. *Angew. Chem. Int. Ed.* **58**, 8216-8220 (2019).

33. G. Wang, A. F. Carey, J. R. Carlson, L. J. Zwiebel, Molecular basis of odor coding in the malaria vector mosquito *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4418-4423 (2010).
34. S. A. Kreher, D. Mathew, J. Kim, J. R. Carlson, Translation of sensory input into behavioral output via an olfactory system. *Neuron* **59**, 110-124 (2008).
35. S. M. Boyle, S. McNally, A. Ray, Expanding the olfactory code by in silico decoding of odor-receptor chemical space. *elife* **2**, e01120 (2013).
36. J. Schiebel *et al.*, Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nat. Com.* **9**, 1-15 (2018).
37. L. Gajdos *et al.*, Neutron crystallography reveals mechanisms used by *Pseudomonas aeruginosa* for host-cell binding. *Nat. Com.* **13**, 1-9 (2022).
38. J. Mondal, R. A. Friesner, B. Berne, Role of desolvation in thermodynamics and kinetics of ligand binding to a kinase. *J. Chem. Theory Comput.* **10**, 5696-5705 (2014).
39. A. M. Davis, S. J. Teague, Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew. Chem. Int. Ed.* **38**, 736-749 (1999).
40. R. F. de Freitas, M. Schapira, A systematic analysis of atomic protein–ligand interactions in the PDB. *Medchemcomm* **8**, 1970-1981 (2017).
41. P. Setny, R. Baron, P. M. Kekenus-Huskey, J. A. McCammon, J. Dzubiella, Solvent fluctuations in hydrophobic cavity–ligand binding kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1197-1202 (2013).
42. L. Charlier *et al.*, How broadly tuned olfactory receptors equally recognize their agonists. Human OR1G1 as a test case. *Cell. Mol. Life Sci.* **69**, 4205-4213 (2012).
43. J. Topin *et al.*, Discrimination between Olfactory Receptor Agonists and Non-agonists. *Chem. Eur. J.* **20**, 10227-10230 (2014).
44. X. Zhang, S. Firestein, The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* **5**, 124-133 (2002).
45. H. Yan *et al.*, An engineered orco mutation produces aberrant social behavior and defective neural development in ants. *Cell* **170**, 736-747. e739 (2017).
46. R. A. Fandino *et al.*, Mutagenesis of odorant coreceptor Orco fully disrupts foraging but not oviposition behaviors in the hawkmoth *Manduca sexta*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 15677-15685 (2019).

47. F. Sievers, D. G. Higgins, Clustal omega. *Curr. Protoc. Bioinform.* **48**, 3.13. 11-13.13. 16 (2014).
48. A. Waterhouse *et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296-W303 (2018).
49. M. Varadi *et al.*, AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439-D444 (2022).
50. D. R. Roe, T. E. Cheatham III, PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084-3095 (2013).
51. V. Le Guilloux, P. Schmidtke, P. Tuffery, Fpocket: an open source platform for ligand pocket detection. *BMC Bioinform.* **10**, 1-11 (2009).
52. C. R. Søndergaard, M. H. Olsson, M. Rostkowski, J. H. Jensen, Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K_a values. *J. Chem. Theory Comput.* **7**, 2284-2295 (2011).
53. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370-D376 (2012).
54. S. Schott-Verdugo, H. Gohlke, PACKMOL-memgen: a simple-to-use, generalized workflow for membrane-protein-lipid-bilayer system building. *J. Chem. Inf. Model.* **59**, 2522-2528 (2019).
55. I. Y. B.-S. D.A. Case, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, *et al.*, AMBER 2018, University of California, San Francisco. (2018).
56. M. Frisch *et al.*, Gaussian 09, Revision D. 01, Gaussian, Inc., Wallingford CT. *See also:* URL: <http://www.gaussian.com> (2009).
57. J. Wang, P. Cieplak, P. A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049-1074 (2000).
58. I. N. Shindyalov, P. E. Bourne, Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**, 739-747 (1998).
59. T. V. Pyrkov, A. O. Chugunov, N. A. Krylov, D. E. Nolde, R. G. Efremov, PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes. *Bioinformatics* **25**, 1201-1202 (2009).

60. M. R. Bauer, M. D. Mackey, Electrostatic complementarity as a fast and effective tool to optimize binding and selectivity of protein–ligand complexes. *J. Med. Chem.* **62**, 3036-3050 (2019).
61. R. E. Roberts, J. K. Yuvaraj, M. N. Andersson, Codon optimization of insect odorant receptor genes may increase their stable expression for functional characterization in HEK293 cells. *Front. Cell. Neurosci.* **15** (2021).
62. C. J. Moreau, J. P. Dupuis, J. Revilloud, K. Arumugam, M. Vivaudou, Coupling ion channels to receptors for biomolecule sensing. *Nat Nanotechnol* **3**, 620-625 (2008).
63. M. Vivaudou, eeFit: a Microsoft Excel-embedded program for interactive analysis and fitting of experimental dose–response data. *Biotechniques* **66**, 186-193 (2019).

Chapter II
Human Trace Amine
Associated Receptors (TAAR)

Mammalian olfactory receptors belong to the class A G protein-coupled receptor (GPCR) family. Humans possess about 400 subtypes of ORs able to sense a virtually infinite number of odorants, and 6 trace amine-associated receptors (TAARs) that bind specifically to volatile amines. Genetic variations are particularly prevalent in OR family, and gene mutations can drastically alter our olfactory perception [1, 2]. We studied the case of the TAAR5 receptor which is mutated in high frequency in Nordic countries [3]. Using a combined molecular dynamics simulation and site-directed mutagenesis approach, we predicted that the TAAR5-S95P polymorphism renders the receptor incapable of being activated, through subtle changes in the allosteric communication network.

Contribution: Jérémie Topin and Loïc Briand designed the study. I set up, launched, and analyzed the molecular dynamics simulations with the help of Jérémie Topin and Sébastien Fiorucci. Christine Belloir performed the in vitro assays. Christine Belloir and I contributed equally as first authors.

Reference: *Pacalon J, Belloir C, Fiorucci S, Briand L, Topin J. TAARs response to amine are largely affected by sequence variants. Submitted.*

1. Keller A, Zhuang H, Chi Q, et al (2007) Genetic variation in a human odorant receptor alters odour perception. *Nature* 449:468–472. <https://doi.org/10.1038/nature06162>
2. de March CA, Topin J, Bruguera E, et al (2018) Odorant Receptor 7D4 Activation Dynamics. *Angewandte Chemie International Edition* 57:4554–4558. <https://doi.org/10.1002/anie.201713065>
3. Gisladdottir RS, Ivarsdottir EV, Helgason A, et al (2020) Sequence Variants in TAAR5 and Other Loci Affect Human Odor Perception and Naming. *Current Biology* 30:4643-4653.e3. <https://doi.org/10.1016/j.cub.2020.09.012>

Publication 2

TAARs response to amine are largely affected by sequence variants.

Pacalon J, Belloir C, Fiorucci S, Briand L, Topin J. Submitted.

Abstract

Volatile amines are recognized by a family of chemosensory receptors: the Trace Amine Associated Receptors. Compared to regular olfactory receptors, TAARs are few (6 receptors expressed in the olfactory epithelium) and highly conserved. Thus, polymorphisms in this family can drastically alter our perception of amine compounds. A joint approach of numerical simulations and *in vitro* experiments has revealed the activation mechanisms of hTAAR5. hTAAR5-S95P, a polymorphism found in high frequency in Nordic countries, affects the perception of trimethylamine (rotten fish smell). Our 3D model captures both the inability of hTAAR5-S95P to be activated by trimethylamine *in vitro*, and the activation of the receptor by different agonists. Long-scale molecular dynamics (MD) simulations of the system bound to ligands with different efficacies are performed and recover that the receptor is activated only when stimulated by agonists capturing the features of a prototypical active state of GPCR. Our results also suggest that highly conserved residues hTAAR5-C17, probably involved in a third disulfide bridge in the extracellular region, and hTAAR5-R94, are involved in the activation mechanism.

Introduction

Recent advances in neurobiology have shown that mammal olfactory perception relies on multiple chemosensory receptors expressed at the surface of sensory neurons in the olfactory epithelium [1]. Mammal epithelium expresses a diverse repertoire of receptors which binds characteristic odor molecules, allowing the discrimination of an almost unlimited number of chemical compounds [2]. The mammalian olfactory system contain two families of G-Protein Coupled Receptors (GPCRs) class A : the Olfactory Receptors (ORs) [3], and the Trace amine associated receptors (TAARs) [4].

Until today, few links have been established between a chemical function and chemosensory receptors involved in the sense of smell [5]. ORs polymorphisms alter the odor perception either by modifying the odor threshold or changing the odor associate to a molecule [6]. These genetic variations constitute the greatest diversity of the human genome [7], endowing each individual with a unique combination of OR sequences [8]. Several studies have revealed a clear correlation between the function of ORs *in vitro* and the odor rating or olfactory threshold [6,

9, 10]. From a molecular point of view, we recently demonstrated that the degradation of OR7D4 function originate from the impairment of the receptor to reach its active state [11].

TAARs stands as the exception as they appears to be tuned to respond to amine molecules [12]. TAARs are involved in several physiological process [13]. Their sequence variations could lead to mental or metabolic disorder [14] making them a new pharmaceutical target [14, 15].

Despite their relative low number and their evolutionary conservation compared to regular ORs, polymorphisms in this family can also drastically alter our perception of amine compounds [16]. The first detection of specific anosmia to human metabolite trimethyl amine (TMA) have been made by Moore et. al in 1976 [17]. Since then, it has been shown that human TAAR5, a TAAR subtype with the highest expression level in human epithelium [18], specifically recognizes TMA [19]. A recent study has revealed an association between the smell intensity and quality with a single nucleotide polymorphism (SNP) in hTAAR5 [16]. However the molecular background at the origin of this change in odor perception remains enigmatic.

The authors hypothesise that the serine to proline mutation at position 2.65 (Ballesteros Weinstein numbering [20]) induces either a misfolding of the receptor or impairs the membrane targeting. Residue 2.65 is part of an allosteric sites and called vestibular binding site in aminergic GPCR [21]. Mutations on these residues are known to modify the receptor response to agonist, and in the particular case of OR7D4, the S^{2.64}N mutants leads to a change in the phenotype [6].

In this study, we investigate the importance of conserved residues on odorant-induced activation. We synergise numerical approaches together with site-directed mutagenesis and *in vitro* assays to decipher at the molecular level the basis of the specific anosmia from TAARs. Our model predicts the activation of two human TAARs (TAAR1 and TAAR5), when bound to their agonists, a control non-agonist, or considered in their apo form. The protocol was accurate enough to predict the effect of different mutations on the activation of the receptor, in good agreements with *in vitro* data. We demonstrate that serine^{2.65} to proline polymorphism abrogates TAAR response to different ligands for both hTAAR5 and hTAAR1.

Results

Starting from the same initial AlphaFold modeled structure of TAAR, four unconstrained microsecond MD simulations were performed for height systems: *wt* (TAAR1 and TAAR5) in apo form and bound to their agonist 3-iodothyronamine (T₁AM) [22] and TMA [19] respectively, and S^{2.65}A and S^{2.65}P TAAR1 and TAAR5 mutants bound to their respective agonists. The ligands are interacting with the D^{3.32}. In all simulations, the bundle structures remained stable and do not shift strongly from the initial conformations (see Figures S1-4).

At the molecular level, ligand binding to GPCR triggers a subtle and dynamic mechanism that spreads the allosteric signal between the ligand and G protein binding sites over a distance of more than 20 Å (~5 helix turns) [23]. It results to an opening of a cleft between the intracellular parts of the transmembrane helices (TM) 3 and 6 [24].

Receptor activation was evaluated by the analysis of the intracellular cleft and more particularly by following the C α distance of the so-called ionic lock (D^{3.49} - K^{6.30}) between TM3 and TM6 (see Figure 1A). This metric is well-established as a marker for GPCR activation [11, 25]. In the case of WT apo form, the system samples a structure where the ionic-lock is closed, with a TM3-TM6 inter backbone distance oscillating around 11 Å (Figure 1B, grey color). During very short periods, that could be representative of a constitutive basal activity, receptor samples active state, confirming that the model accurately captures the back and forth between inactive and active states. When docking an agonist within the binding pocket of the initial model, the system dynamics is modified: TM6 is more likely to shift away from TM3 (Figure 1B green color). Our model recovers the agonist induced stabilization of the active state for both TAAR1 and TAAR5. For all the 8 considered system, a correlation between in vitro response and TM6 dynamics is observed.

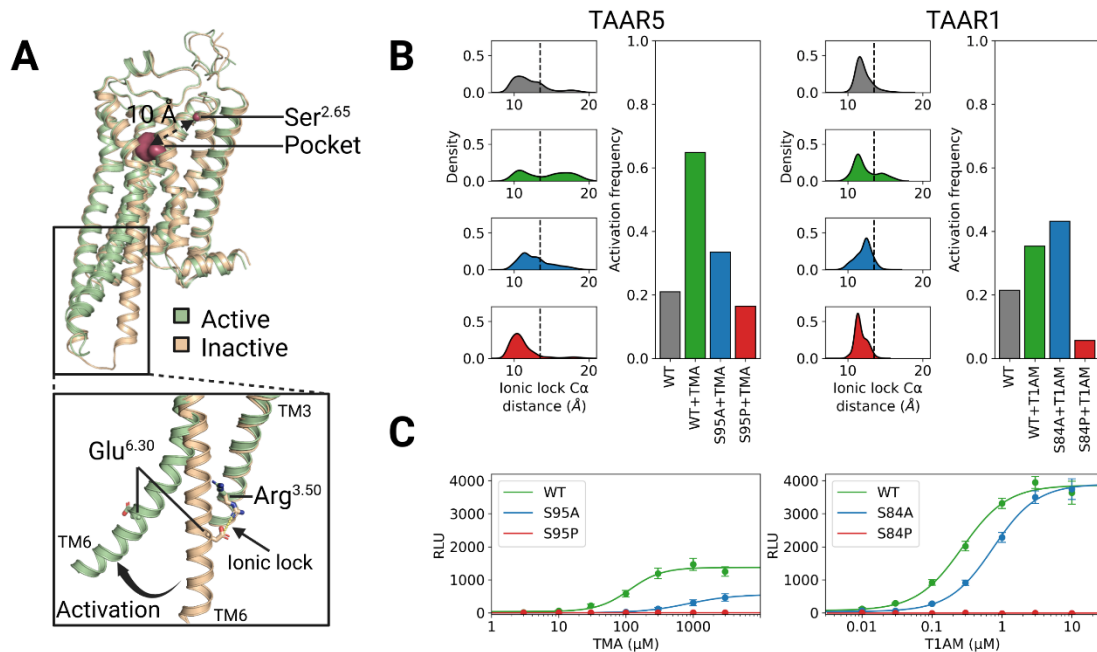


Figure 1: **A**) TAAR model in active (green) and inactive form (beige). The ligand binding pocket appears in red. TAAR activation was deduced from the opening of the intracellular cleft between TM6 and TM3 (measured by the distance between the Ca atoms of the ionic-lock residues). **B**) Results from the multiple molecular dynamics simulations (4 replicas of >2μs long) for the wt (grey) and the different mutants (colored). The limit between active and inactive forms is indicated as a dot line. For both TAARs, simulations with a known agonist (TMA and T1AM for TAAR5 and TAAR1 respectively) show an increase of the sampling of the active form (green) compared to the apo (grey). Mutation of residue S2.65 to an alanine results in a decrease of the sampling of the active form (blue) while the mutation to a proline (red) almost abolish the sampling of the active form. **C**) Normalized responses are shown as means and s.e.m. (n=3).

We further evaluate the importance of mutation S^{2.65} on the receptor response to agonist. The simulations revealed that the receptor is still responding to its ligand and is differentially affected when the serine is mutated to an alanine (Figure 1B, blue color). For TAAR5, the system shows a decrease in sampling the active state compared to *wt* (Figure 1 B, blue color). In the case of TAAR1, a minor increase in the sampling of the active state is observed. The *in vitro* measurements corroborate these results. In the case of TAAR5, the EC₅₀ and the efficacy of TMA are largely affected (114 +/- 20 μM and 848 +/- 63 μM, for TAAR5-WT and TAAR5-S95A respectively). On the contrary, TAAR1 mutant conserves a similar response to T₁AM both in term of efficacy and potency (0.27 +/- 0.04 μM and 0.75 +/- 0.05 μM, for TAAR1-WT, TAAR1-S84A respectively).

Finally, when the S^{2.65} is mutated to a proline, simulations of both TAAR1 and TAAR5 indicate a drastic decrease in the sampling of the active structure. Consistently, the two mutants S^{2.65}P are both insensitive to their respective agonists (Figure 2B and 2C red color). These functional differences may be explained by distinct receptor trafficking and/or dynamics [26]. To control that the differences are not due to a lack of heterologous expression of the mutated receptors and their addressing to the membrane, we verified receptor transfection rates.

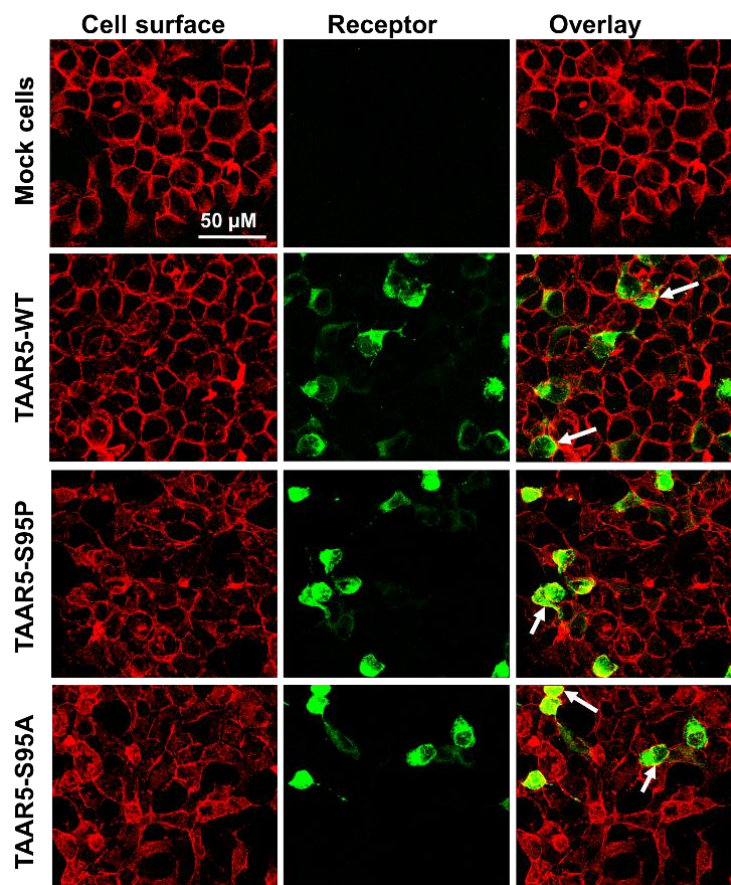


Figure 2: Membrane localization of TAAR5 receptors in Hana3A cells. Confocal fluorescence images of Hana3A cells transfected with TAAR5-WT, TAAR5 SNP variants DNA and mock transfected Hana3A cells are shown. Receptor expression is detected by a primary antibody against the FLAG-tag in combination with an Alexa 488-labeled anti-mouse antibody (green). Plasma membrane is visualized by biotinylated concanavalin A binding to plasma glycoproteins in combination with streptavidin-Alexa 568 (red). Colocalization of the TAAR receptor at the cell surface appear in yellow in the overlay pictures (right panel). Scale bar, 50 μ m.

Here we report for the first time that TAAR1 and TAAR5 and their respective mutants S95P and S84P were localized intracellularly with a strong signal in cytoplasmic region and were also successfully inserted at the cell surface (Figure S5 and S6). As shown in Figure 2 (and

Figure S7), transfection rates confirmed that most TAAR5 receptors were expressed in an average frequency of 19 % of cells. TAAR1 receptor expression was lower with an average frequency of 12.6 % of cells (Figure S8). Small differences observed in membrane transfection rate and targeting of the receptor seems to be unrelated to the functional activity measured by GloSensor assay. Indeed, luminescence signal amplitude is greater for TAAR1-WT construct than for TAAR5-WT whereas transfection rate is higher for TAAR5 than TAAR1. This insensitivity to TMA thus explain the origin of the different phenotype of individual carrying the S95^{2.65}P mutation on TAAR5 previously characterized Gisladdottir *et al.* [16].

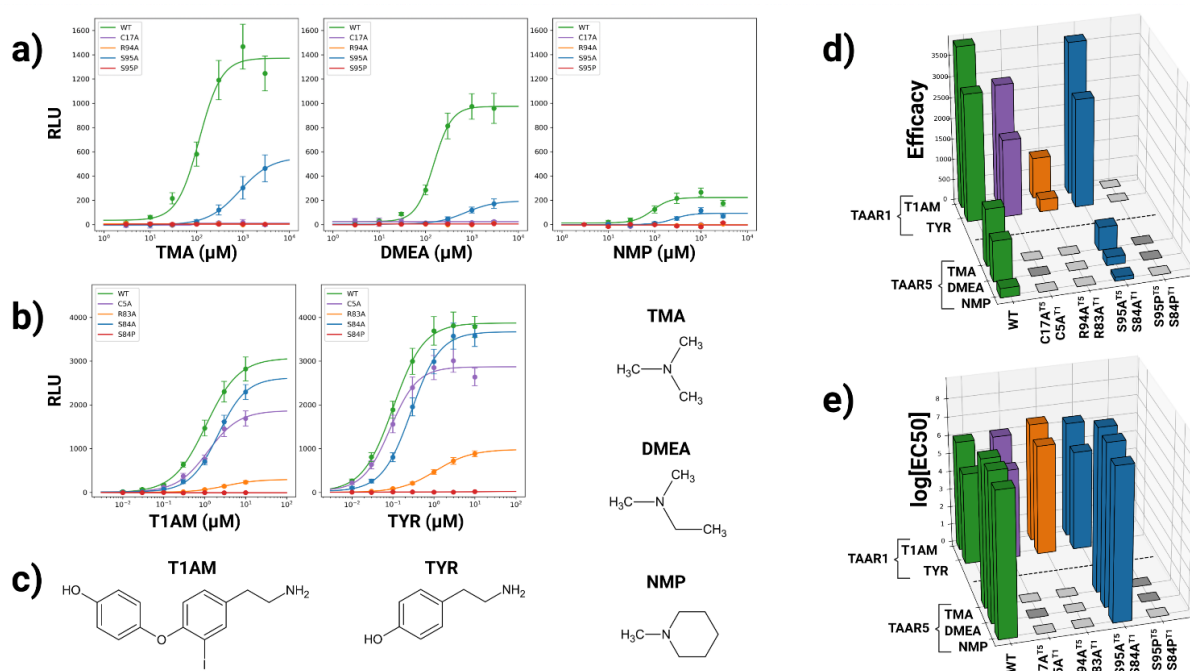


Figure 3: Dose-response curves of TAAR5 (A) and TAAR1 (B) wt and mutants responding to various volatile amines (C). Normalized responses are shown as means and s.e.m. Wt. and mutants receptor efficacy (D) and potency (E) are summarized as bar graphs.

To get new insights into structure-function relationships of the TAAR family, we also studied the role of highly conserved residues surrounding the ligand binding site, especially TAAR1-C5 and R83 aligned with TAAR5-C17 and R94. On the one hand, R94^{2.64} is part of the vestibule binding site and has been hypothesized to stabilize large aromatic TAAR5 ligands through cation- π interactions [27]. On the other hand, AlphaFold models suggested a third disulfide bridge involving TAAR1-C5 or TAAR5-C17 in the N-terminal part and the top of TM2, uncommon in class A GPCR but already observed for the human A2A adenosine receptor. We then extend the functional assays to other volatile amines and measure the receptor response to

single point mutations TAAR1-C5A, TAAR5-C17A, TAAR1-R83A and TAAR5-R94A (Figure 3).

As expected, the TAAR1-S84A/P, TAAR5-S95A/P and WT receptor responses to other agonists are consistent with previous conclusions. However, TAAR1 and TAAR5 are differently affected by the mutations of the cysteine and arginine residues. C17A and R94A mutations abolished TAAR5 response to tested amines while TAAR1-C5A and R83A mutants are still responding with a reduced efficacy (Figure 3). As for previous *in vitro* experiments, the mutants are still expressed and localized at the cell surface (Figure S5-S8). The observations may be due to differences in ligand recognition or in the receptor activation mechanism.

The MMGBSA analysis of MD trajectories (Table S2) and the measure of ligand-receptor per-residue contact frequencies (Table S3) indicates that even if the ligand-receptor recognition is modified by single point mutants, residues R83/94^{2,64}, S84/95^{2,65} and C5/17^{N-term} are not in direct contact with the ligand and do not contribute to stabilize it in the binding site.

Extracting dynamical information from the MD simulation bring details on the role of these mutations. Degree centrality measures the influence of a given residue in the protein network (Figure S9). The allosteric mechanism occurring during receptor activation is highlighted by the variation in degree centrality. A consequence of the mutations is reflected by a modification of the protein network, especially in the TM2 and TM7 helix (Figure S9). We hypothesize that residues C5/17^{N-term}, R83/94^{2,64} and S84/95^{2,65} play a functional role in the activation mechanism and not in the ligand recognition. This seems consistent with experimental results, especially for C5^{N-term} and R83^{2,64} TAAR1 mutants, showing a significant decrease in ligand efficacy without changing their potency.

In this study, we illustrate how computational methods combined with site directed mutagenesis experiments and functional assays are state of the art approaches for the study of GPCR structure-function relationships. In total, the dynamics of TAAR1 and TAAR5 variants were explored for ~0.1 ms and the models captures the typical features of GPCR active state in agreement with *in vitro* experiments. We demonstrate that TAAR5-S95P polymorphism abolishes the receptor response to various amines, even if the receptor is still expressed at the cell surface, explaining how this variant affects the perception of fish odor containing trimethylamine.

Supporting Information

Experimental Section

Chemicals.

Ligands (Trimethylamine (TMA) CAS 75-50-3, Dimethylethylamine (DMEA) CAS 598-56-1, Tyramine (TYR) CAS 51-67-2, 3-iodothyronamine (T1AM) CAS 78824-64-6) were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). Initial solubilization were made in water at 100 mM for TMA and DMEA and 25 mM for TYR and T1AM. Further concentrations were freshly prepared by dilution in C1 buffer (130 mM NaCl, 5 mM KCl, 10 mM Hepes, 2 mM CaCl₂, 5 mM sodium pyruvate, pH 7.4). Cell culture media were purchased from Invitrogen (Life Technologies). For immunocytochemistry, biotin-conjugated concanavalin A (C2272) and mouse monoclonal anti-FLAG M2 antibody (F1804) were purchased from Sigma-Aldrich while streptavidin Alexa Fluor 568 conjugate (S11226), goat anti-mouse Alexa 488 secondary antibody (A11001) and Prolong gold antifade reagent with DAPI were purchased from Life Technologies (St Aubin, France).

Design of TAARs expression constructs.

The cDNA sequence encoding Homo sapiens TARR1 (UniProtKB accession Q96RJ0) and TAAR5 (accession O14805) were cloned into pcDNA4 expression vector generating pcDNA4-hTAAR plasmids. Codon were optimized for expression in mammalian cells and FLAG-tag epitope was added to C-terminus of each construct to measure receptor expression. TAAR1 and TAAR5 variants were generated through the introduction of point mutations using PCR-based direct mutagenesis (Azenta Life Sciences). Plasmids were amplified after transformation in E.coli Top10F' (Life technologies) and purified by QIAfilter Plasmid Midi kit (Qiagen, Courtaboeuf, France). Sequence of each amplified plasmid was confirmed by Sanger sequencing before use (Azenta Life Sciences).

Heterologous expression and GloSensor cAMP assay.

TAARs functional activities were experimented on Hana3A cells kindly provided by Dr. H. Matsunami (Duke University, Durham, USA). The Hana3A cells derived from HEK293T cells after stable transfection with the G alpha protein $G\alpha_{olf}$, and accessory proteins RTP1L, RTP2, Reep1 known to enhanced cell surface expression of odorant receptor [28, 29]. For GloSensor cAMP assay (Promega), cells were seeded into 96-well white walled, clear bottom, poly-D-lysine treated microtiter plates (Corning), at density of 0.35×10^6 cells per well in high-glucose DMEM supplemented with 2 mM GlutaMAX, 10% dialyzed foetal bovine serum, penicillin/streptomycin and puromycin (1 $\mu\text{g}/\text{mL}$) at 37°C and 6.3% CO₂ in a humidified atmosphere. Then, 24 h later, using Fugene HD (Promega), cells were transiently transfected with hTAARs constructs, the pCI-RTP1S vector and the pGlo™-22F cAMP plasmid (Promega) used as a genetically encoded firefly luciferase biosensors allowing accurate detection of the intracellular second messenger cAMP. Hana3A cells were also transfected with empty plasmid pcDNA4 (mock cells) as negative control. After another 24 h incubation, the cells were loaded with GloSensor cAMP reagents diluted in CO₂-independent medium during 2 h at room temperature, before being stimulated by chemicals compounds. Changes in luminescence intensity were measured for 20 min after the addition of a range of ligand solution, using FlexStation 3 system (Molecular Devices). We calculated a fold response of luciferase activity using ratio of the basal luciferase value before addition of ligand upon luminescence value measured at 10 min after stimulation. The dose-response data were fitted using a four-parameter logistic equation. The median effective concentrations (EC₅₀ values) were generated using SigmaPlot software (SystatSoftware). The assays were performed in duplicate, and each experiment was repeated at least three times.

Immunocytochemistry.

Immunological detection of hTAARs were performed as previously described with slight modification. We seeded Hana3A on 4-well culture slides (Corning) coated with Corning Cell-Tak adhesive at density of 0.15×10^6 cells per well and transiently transfected pcDNA4-hTAAR plasmids using Fugene HD after 24 h. The next day after transfection, we washed the cells with HBSS HEPES solution, cooled them at 4°C for 30 min. For plasma membrane staining, we incubated cells for 1 h with 20 $\mu\text{g}/\text{mL}$ biotin-labeled concanavalin A (Sigma), which binds

cell-surface glycoproteins. Then, we fixed the cells for 5 min in ice cold methanol:acetone (1:1). Blocking step was done in Dako antibody diluent buffer supplemented by 5% goat serum to reduce non-specific binding. We added mouse antiserum against FLAG epitope (Sigma, 1:500) to detect the TAAR receptors and Alexa Fluor 568 streptavidin conjugate (Life Technologies, dilution 1:500) to stain the cell surface. Then, we incubated cells with the secondary antibody Alexa Fluor 488-conjugated goat antiserum against mouse IgG (Life Technologies, 1:400) to visualize TAAR receptor. Each incubation was performed at room temperature and lasted one hour with intermediates washes with PBS between each step. Finally, we embedded the cells in Fluorescent Mounting medium (Life Technologies) with DAPI allowing detection of cell nucleus and analyzed them using an epi-fluorescence inverted microscope (Eclipse TiE, Nikon, Champigny sur Marne, France) equipped with an 20× objective lens and a LucaR EMCCD camera (Andor Technology, Belfast, UK) or a confocal laser-scanning microscope (Leica TCS SP8) equipped with an 63× objective lens (DIMAcell platform, University of Burgundy, Dijon, France).

The acquired images allowed us to calculate the proportion of cells expressing recombinant TAAR receptors (number of cells with green fluorescence divided by total cell number in the microscope field) and to visualize the expression of TAAR receptors at the plasma membrane (cells with colocalization of green and red fluorescence).

Computational Methods

Molecular dynamics simulations.

The initial structures of TAAR5 and TAAR1 receptors were generated using AlphaFold [1]. 8 systems were considered: TAAR5-WT apo, TAAR5-WT+TMA, TAAR5-S95A+TMA, TAAR5-S95P+TMA, TAAR1-WT apo, TAAR1-WT+T1AM, TAAR1-S84A+T1AM, TAAR1-S84P+T1AM (cf. Table S1). Mutations were performed using the Pymol mutagenesis tool [30]. Ligands were docked with Autodock Vina 1.1.2 [31] into their respective binding sites, with the best pose selected by visual inspection. Propka3 [32] was used to predict the protonation states of the proteins at a target pH of 6.5. Orientation in the membranes were determined using the OPM server [33]. The systems were integrated into a POPC-only model membrane using PACKMOL-memgen [34]. The simulation boxes were completed using TIP3P water molecules and neutralized using K^+ and Cl^- ions with a final concentration of 0.15 M.

Molecular dynamics simulations were performed with the AMBER18 [35] sander.MPI and pmemd.cuda modules, with the ff14SB force field for proteins, the lipid14 force field for the membrane [36]. The ligand parameters were generated by calculating the partial atomic charges with the HF/6-31G* basis set using Gaussian 09 [37]. The resulting electrostatic potential was fitted by the RESP program [ref]. The other parameters were taken from the general amber force field 2 (gaff2) [38]. Bonds involving hydrogen atoms were constrained using the SHAKE algorithm and long-range electrostatic interactions were treated using Particle Mesh Ewald. The threshold for unbound interactions was set at 10 Å. Each system was first minimized with the AMBER18 [35] sander.MPI module with 5000 steps of the steepest descent algorithm followed by 5000 steps of conjugate gradient with a harmonic potential constraint of 10 kcal·mol⁻¹·Å² on the protein part of the system. A second minimization run was performed without constraints. The systems were then thermalized from 100 to 310 K for 10000 steps with constraints on the protein and ligands with a harmonic potential of 10 kcal·mol⁻¹·Å². Each system underwent 50000 steps of equilibration in the NPT and 1 bar ensemble constraining the protein and ligands with a harmonic potential of 10 kcal·mol⁻¹·Å², before the production phase. During the equilibration and production phase, the temperature was kept constant in the system at 310 K using a Langevin thermostat with a collision frequency of 5 ps⁻¹. 4 replicates were made for each system, for a simulation time of 2.5 to 3 μs per replicate.

Analysis of Molecular dynamics simulations.

The stability of the system was assessed from the evolution of the root mean square deviation (RMSD) calculated: 1) on the transmembrane helix backbone (Figure S1 and S2) 2) on the ligand only (Figure S3 and S4), using the RMSD module of pytraj 2.0.2 [39]. For all systems, the receptors underwent small fluctuations (RMSD < 3 Å) showing that they remained correctly folded during the microsecond simulations (Figure S1 and S2). For all systems with a ligand, the ligands remained in their respective orthosteric cavities during the simulations (RMSD < 2.0 Å). Along the MD trajectories, the frequency of interaction between 1) the receptors and their respective ligands, 2) the amino acids of the ionic lock were calculated using get_contact [40] with the default options. The minimum distance between the amino acids of the ionic lock were calculated with the mindist module of pytraj 2.0.2 [39]. Dynamical correlation analysis was performed using Correlation Plus 0.2.1 [41], starting with the "calculate" module and the "ndcc" option with default settings. The data was then evaluated using the "analyze" module

and the "degree" option with default settings. MM-PBSA calculations were performed using the AMBER18 [35] sander.MPI module, on snapshots sampled every 25 ns and covering the entire respective trajectory, with a per-residue decomposition.

Table S1: summary of the MD simulations.

Receptor	Wild type / mutant	Apo / ligand	Total simulation length (μs)
TAAR5	WT	apo	13.1
TAAR5	WT	TMA	13.0
TAAR5	S95A	TMA	12.8
TAAR5	S95P	TMA	8.6
TAAR1	WT	apo	13.3
TAAR1	WT	T1AM	12.7
TAAR1	S84A	T1AM	10.7
TAAR1	S84P	T1AM	10.4

Table S2: MMGBSA per-residue energy decomposition. Only the top 10 residues positively and negatively contributing to the binding energy are summarized in the table.

Residue	T5WT+TMA	Residue	T5S95A+TMA	Residue	T5S95P+TMA	Residue	T1WT+T1AM	Residue	T1S84A+T1AM	Residue	T1S84P+T1AM
LIG 323	-11,81	LIG 323	-9,84	LIG 323	-10,56	LIG 323	-13,22	ASP 103	-13,41	ASP 103	-11,81
ASP 114	-5,62	ASP 114	-4,64	ASP 114	-8,97	ASP 80	-11,89	LIG 322	-9,20	LIG 322	-11,06
ASP 80	-3,77	ASP 80	-4,46	ASP 80	-3,47	ASP 114	-4,15	ASP 69	-3,98	TRP 264	-3,42
ASP 288	-1,49	ASP 288	-1,67	ASP 288	-1,94	ILE 291	-3,06	ASP 287	-3,32	ILE 104	-3,09
ASP 271	-1,30	ASP 271	-1,51	TYR 295	-1,69	ASP 271	-2,69	TRP 264	-2,90	ASP 69	-2,75
ILE 291	-1,15	PHE 268	-1,21	ILE 291	-1,67	TYR 295	-2,61	ILE 104	-2,55	SER 107	-2,60
TYR 295	-0,98	ASP 275	-0,84	ASP 271	-1,46	ASP 288	-2,45	SER 107	-2,46	THR 100	-2,23
ASP 275	-0,68	TRP 265	-0,79	PHE 268	-0,96	PHE 290	-2,16	PHE 268	-1,18	ASP 287	-1,96
THR 115	-0,63	ILE 291	-0,69	HID 110	-0,78	CYS 118	-2,01	THR 100	-1,07	SER 198	-1,22
PHE 287	-0,49	THR 115	-0,61	ASP 275	-0,74	THR 115	-1,04	PHE 195	-1,05	PHE 268	-0,91
HID 110	-0,42	HID 110	-0,40	TRP 265	-0,68	ASP 275	-1,02	PHE 267	-1,04	PHE 154	-0,77
HID 124	0,30	THR 120	0,34	HID 124	0,32	SER 91	0,24	CYS 270	0,31	CYS 270	0,20
ASN 204	0,30	SER 91	0,35	ILE 122	0,34	LEU 195	0,26	LEU 106	0,31	VAL 76	0,20
THR 120	0,35	ARG 132	0,36	ASN 301	0,34	HID 124	0,27	TYR 294	0,32	LEU 106	0,22
PRO 267	0,36	MET 1	0,36	TYR 165	0,35	ARG 108	0,31	SER 80	0,38	SER 110	0,26
ARG 2	0,40	ARG 2	0,36	LEU 195	0,36	ASN 204	0,34	HID 113	0,39	ASN 296	0,26
ASN 301	0,47	ASN 301	0,37	THR 120	0,40	THR 120	0,35	SER 110	0,44	ASN 113	0,27
LYS 198	0,48	HID 124	0,39	ASN 204	0,46	ASN 301	0,38	SER 183	0,44	TYR 294	0,33
SER 121	0,49	ASN 204	0,39	SER 121	0,46	SER 121	0,38	GLY 293	0,45	ASN 300	0,35
ASN 297	0,62	ASN 297	0,67	LYS 198	0,60	ASN 297	0,50	ASN 300	0,49	GLY 293	0,42
ARG 94	2,49	ARG 94	2,83	ARG 94	3,22	LYS 198	3,95	ARG 83	4,32	ARG 83	3,44

Table S3: Per-residue ligand-receptor contact frequency analysis. Top 20 interacting residues are reported.

Number	TAAR5						TAAR1					
	T5WT+TMA		T5S95A+TMA		T5S95P+TMA		T1WT+T1AM		T1S84A+T1AM		T1S84P+T1AM	
	Residue	Frequency	Residue	Frequency	Residue	Frequency	Residue	Frequency	Residue	Frequency	Residue	Frequency
1	ASP 114	0,836	ASP 114	0,83	ASP 114	1	ASP 103	1	ASP 103	1	ASP 103	0,999
2	CYS 118	0,575	PHE 268	0,664	TYR 295	0,963	SER 107	0,998	SER 107	0,999	SER 107	0,993
3	ALA 294	0,566	CYS 118	0,576	ALA 294	0,915	TRP 264	0,979	TRP 264	0,946	SER 198	0,969
4	TYR 295	0,529	ILE 291	0,332	PHE 268	0,797	SER 198	0,969	PHE 268	0,934	ILE 104	0,961
5	ILE 291	0,394	TYR 295	0,278	ILE 291	0,704	PHE 267	0,853	SER 198	0,877	TRP 264	0,929
6	LEU 196	0,227	TRP 265	0,249	CYS 118	0,645	ILE 104	0,821	ILE 104	0,833	PHE 268	0,9
7	PHE 268	0,224	ALA 294	0,193	TRP 265	0,286	PHE 268	0,734	PHE 267	0,82	SER 108	0,856
8	LEU 194	0,146	ASP 80	0,155	LEU 194	0,075	SER 108	0,685	SER 108	0,659	PHE 195	0,628
9	ASP 271	0,142	SER 121	0,15	THR 115	0,073	TYR 294	0,528	PHE 199	0,655	PHE 154	0,592
10	ASP 275	0,113	ASN 301	0,128	ASN 204	0,024	THR 194	0,514	PHE 186	0,535	PHE 186	0,526
11	PHE 287	0,103	TYR 261	0,118	LEU 83	0,019	PHE 154	0,478	PHE 195	0,532	ILE 111	0,491
12	LEU 195	0,085	LEU 194	0,108	LEU 195	0,018	PHE 195	0,46	ILE 111	0,415	PHE 267	0,47
13	THR 115	0,083	LEU 76	0,104	SER 91	0,012	SER 297	0,435	THR 194	0,397	THR 194	0,455
14	PHE 290	0,05	THR 115	0,103	LEU 196	0,012	ASN 296	0,409	ILE 290	0,39	MET 158	0,314
15	LEU 83	0,049	PHE 290	0,085	HID 110	0,011	PHE 199	0,384	THR 271	0,331	PHE 199	0,298
16	THR 282	0,049	SER 298	0,079	PHE 287	0,01	PHE 186	0,34	PHE 260	0,306	THR 100	0,269
17	THR 272	0,048	ASN 297	0,076	THR 111	0,008	ILE 290	0,301	SER 297	0,293	PHE 260	0,209
18	ASN 197	0,041	SER 91	0,065	PHE 208	0,006	ILE 111	0,282	TYR 294	0,287	ILE 290	0,203
19	TRP 265	0,026	ASP 271	0,054	TYR 261	0,006	PHE 260	0,25	ASN 296	0,264	VAL 184	0,197
20	ILE 281	0,024	TYR 305	0,054	ARG 94	0,006	VAL 184	0,244	LEU 261	0,222	ARG 83	0,145

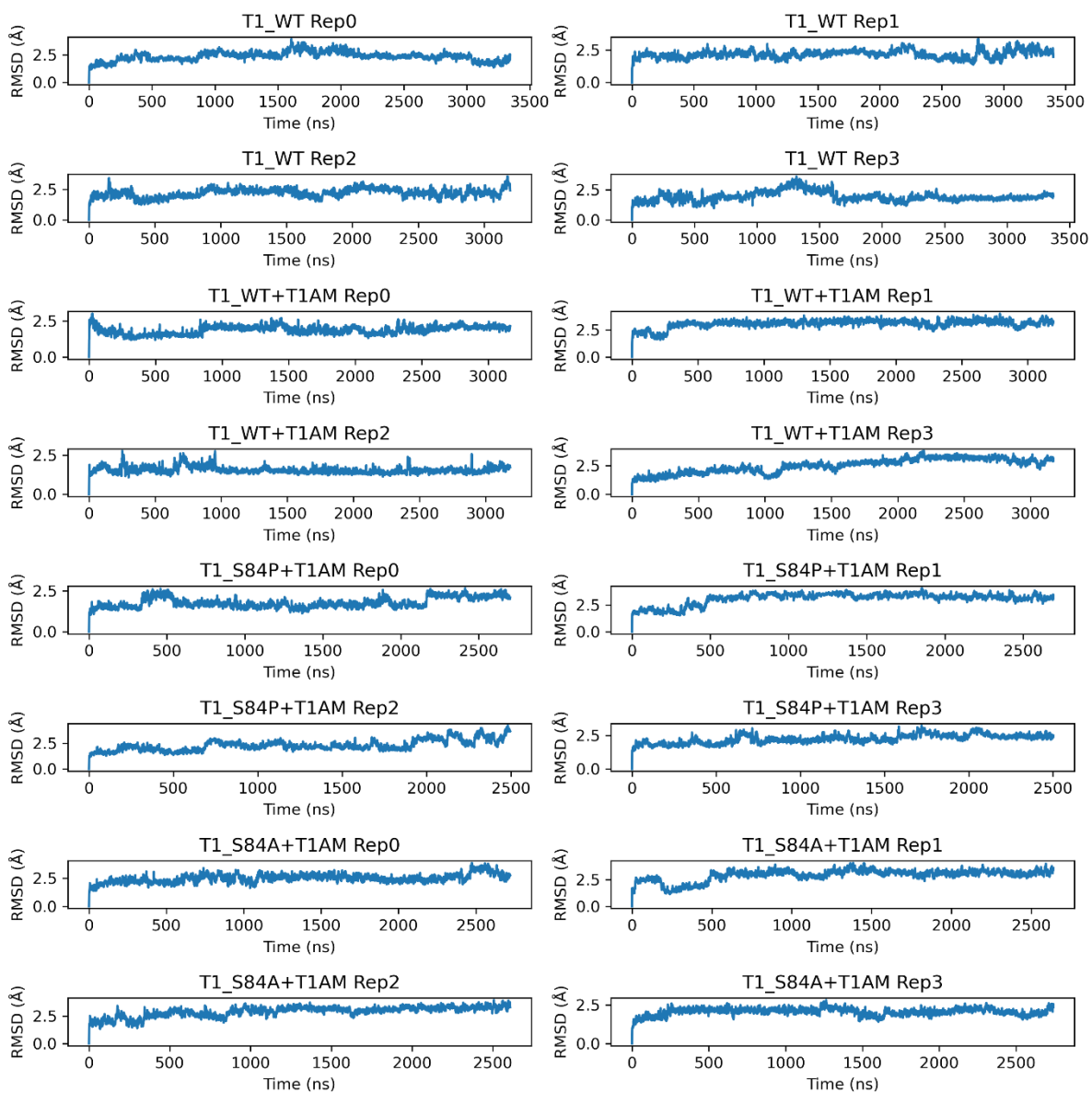


Figure S1: Root Mean Square Deviation (RMSD) of TAAR1 molecular dynamics simulations.

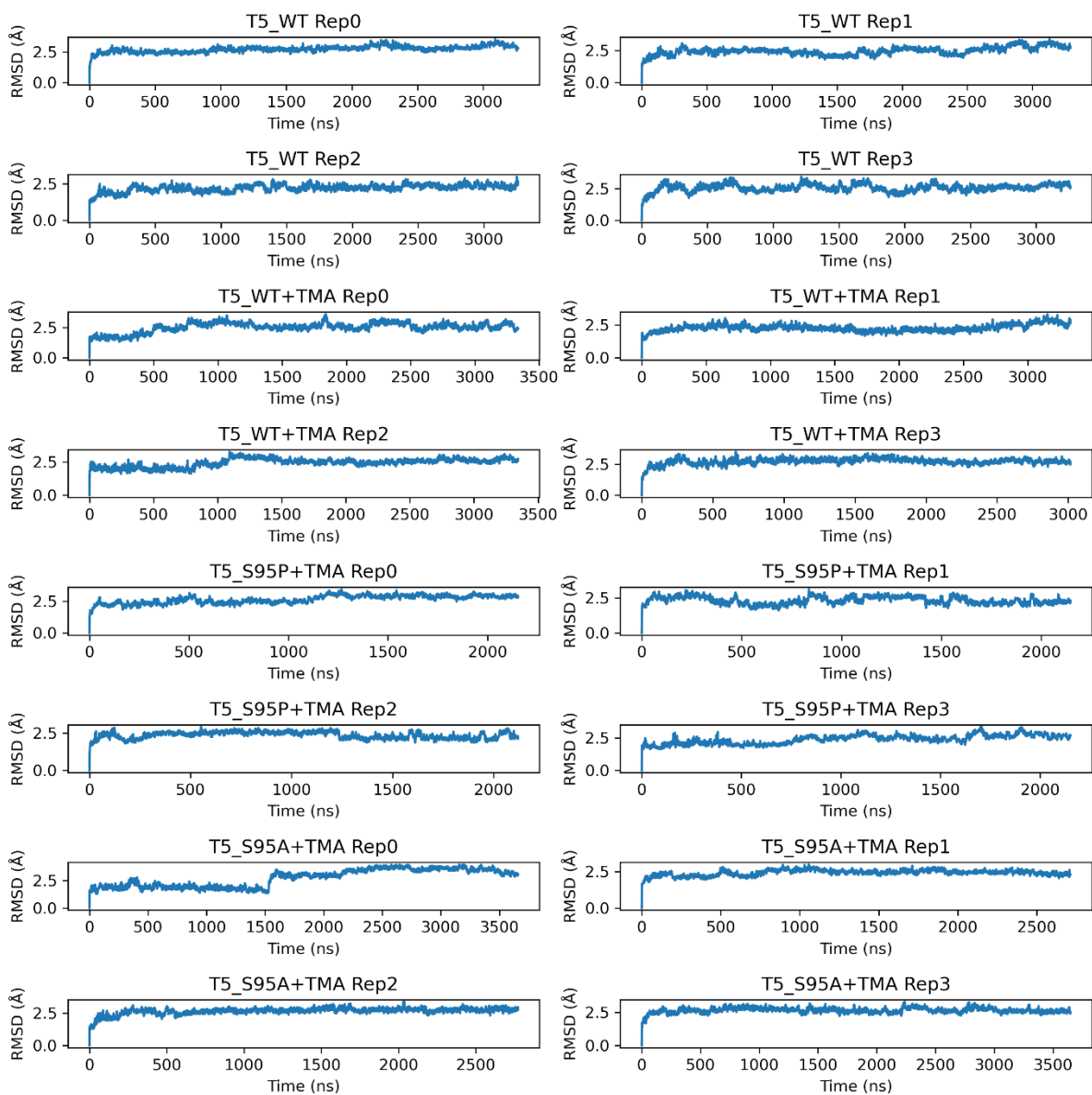


Figure S2: Root Mean Square Deviation (RMSD) of TAAR5 molecular dynamics simulations.

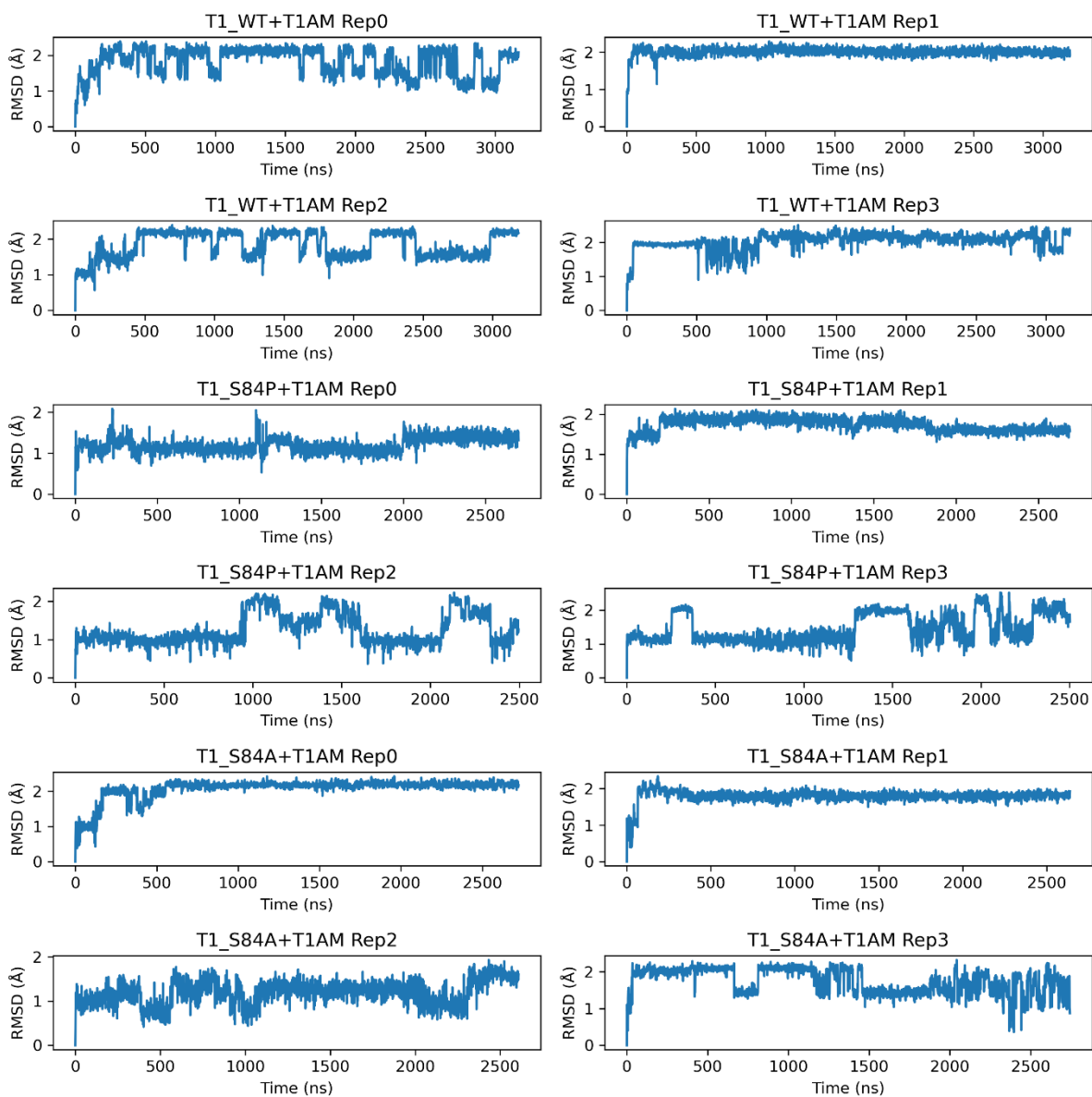


Figure S3: Root Mean Square Deviation (RMSD) of T₁AM interacting with TAAR1.

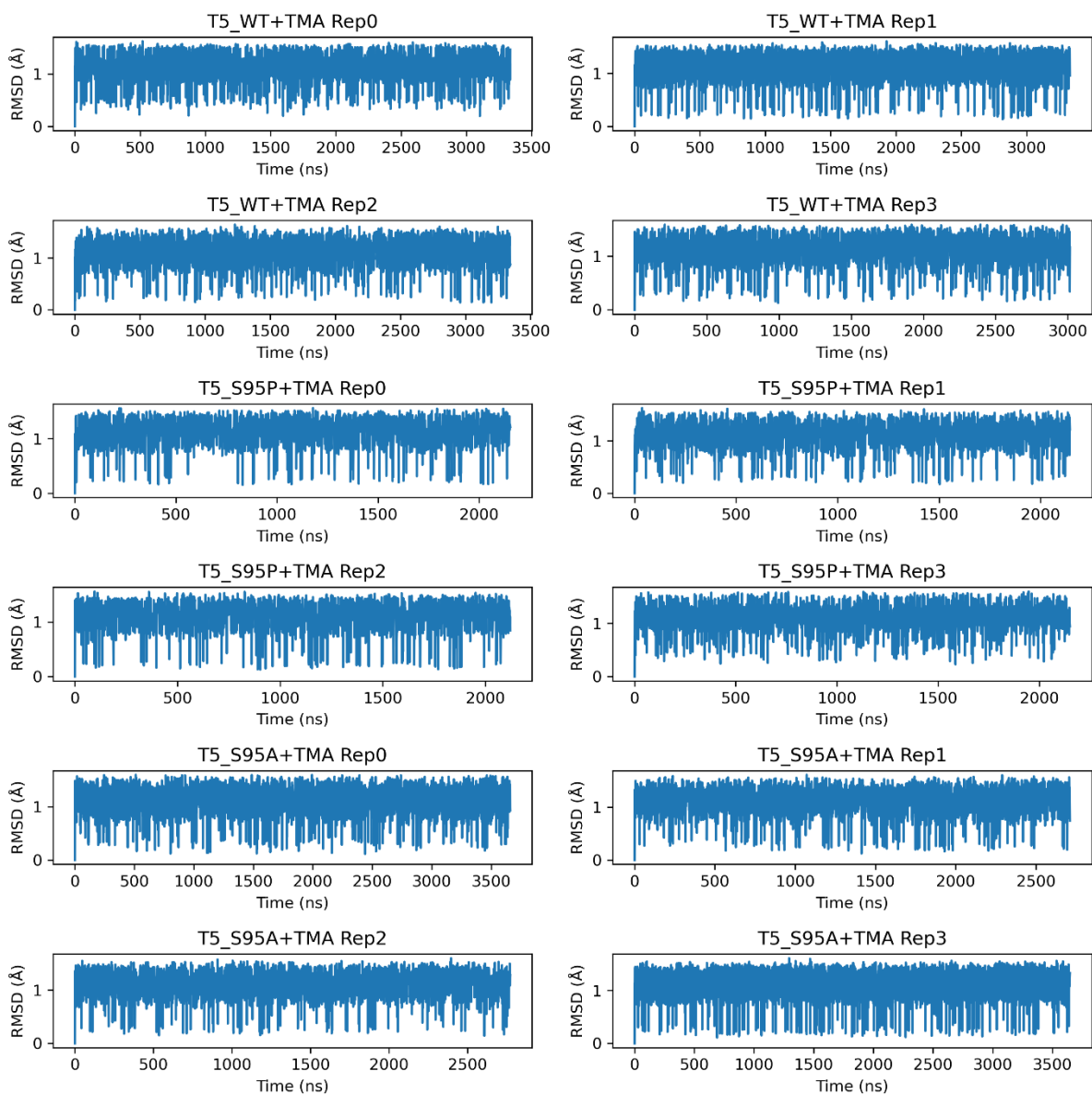


Figure S4: Root Mean Square Deviation (RMSD) of TMA interacting with TAAR5.

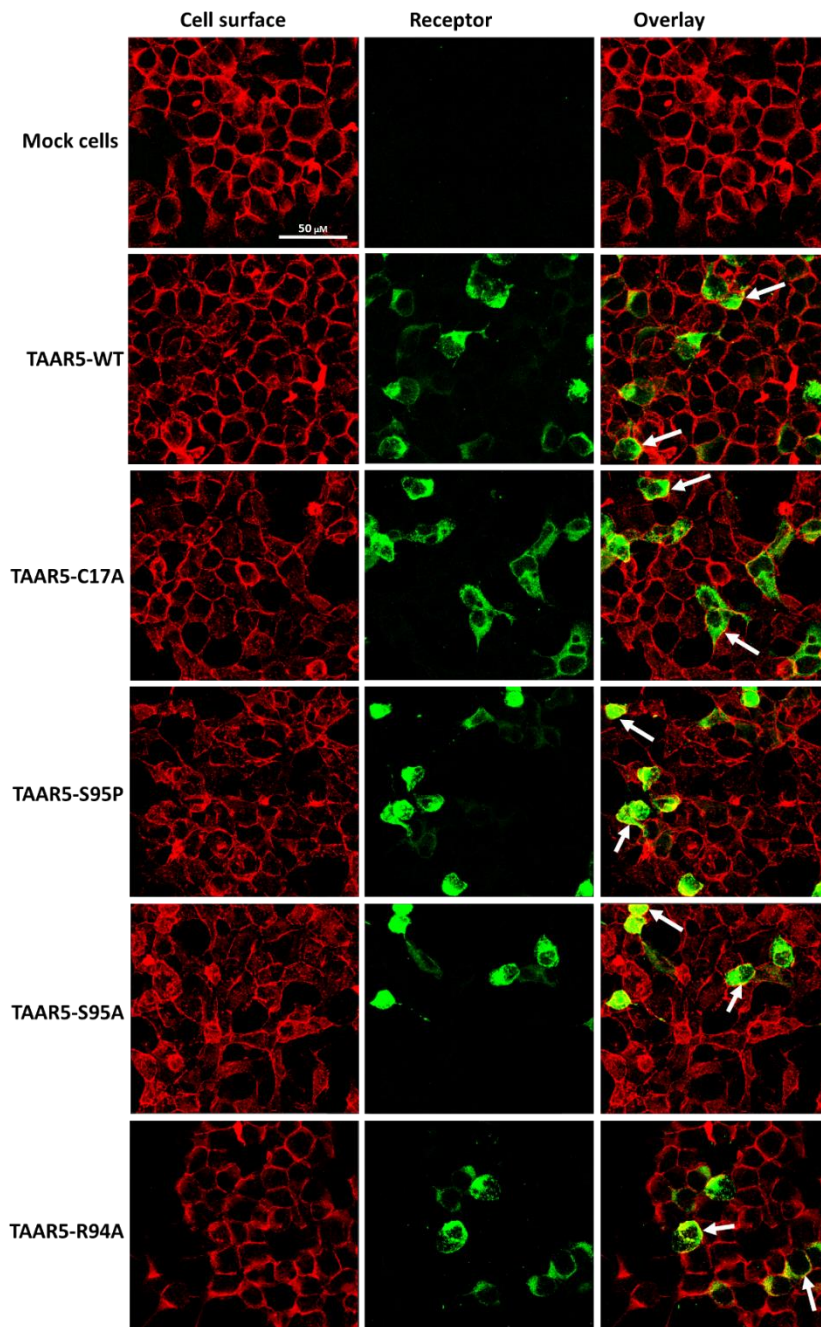


Figure S5: Membrane localization of TAAR5 receptors in Hana3A cells. Confocal fluorescence images of Hana3A cells transfected with hTAAR5-WT, hTAAR5 SNP variants DNA and mock transfected Hana3A cells are shown. Receptor expression is detected by a primary antibody against the FLAG-tag in combination with an Alexa 488-labeled anti-mouse antibody (green). Plasma membrane is visualized by biotinylated concanavalin A binding to plasma glycoproteins in combination with streptavidin-Alexa 568 (red). Colocalization of the TARR receptor at the cell surface appear in yellow in the overlay pictures (right panel). Scale bar, 50 μm .

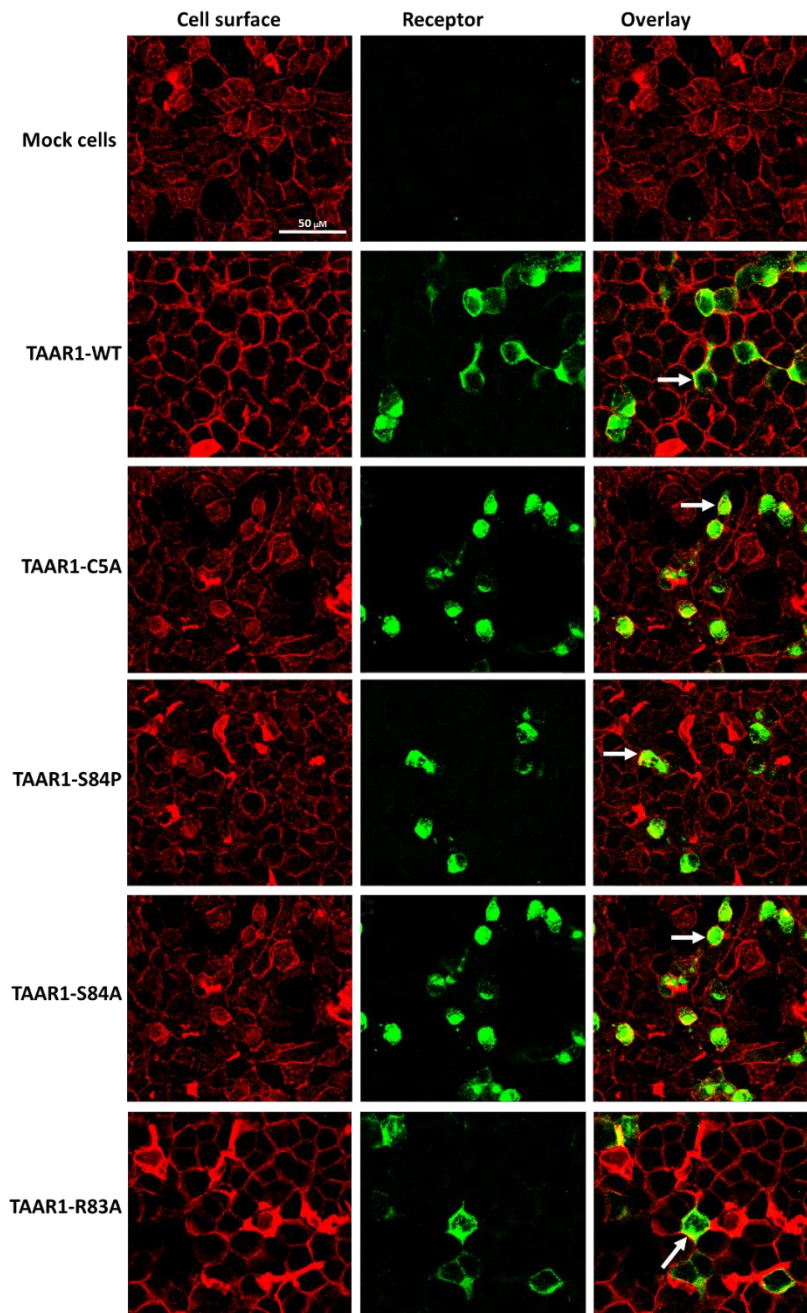


Figure S6: Membrane localization of TAAR1 receptors in Hana3A cells. Confocal fluorescence images of Hana3A cells transfected with hTAAR1-WT, hTAAR1 SNP variants DNA and mock transfected Hana3A cells are shown. Receptor expression is detected by a primary antibody against the FLAG-tag in combination with an Alexa 488-labeled anti-mouse antibody (green). Plasma membrane is visualized by biotinylated concanavalin A binding to plasma glycoproteins in combination with streptavidin-Alexa 568 (red). Colocalization of the TARR receptor at the cell surface appear in yellow in the overlay pictures (right panel). Scale bar, 50 μm .

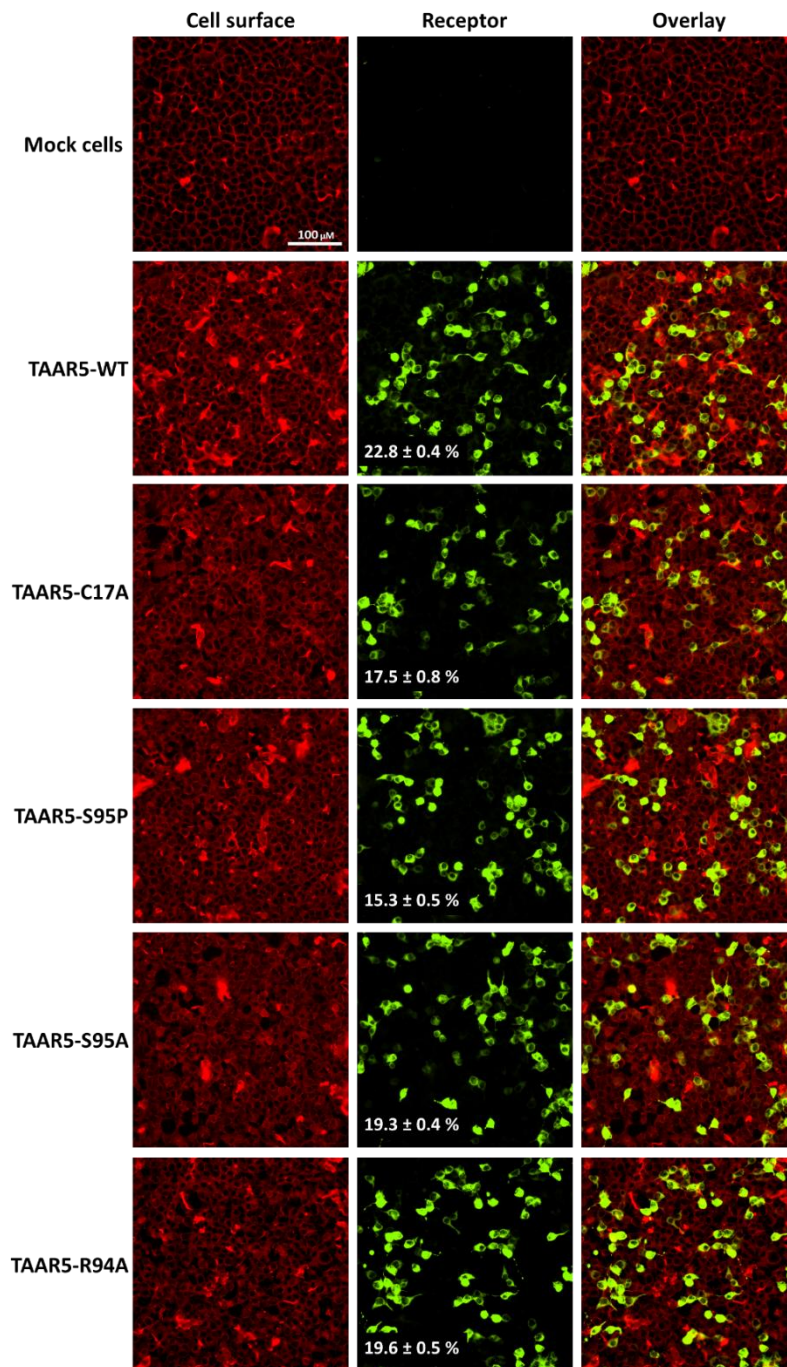


Figure S7: Expression of hTAAR5 after transient transfection into Hana3A cells. Epifluorescence images of and hTAAR5-WT, hTAAR5 SNP variants and mock transfected Hana3A cells are shown. Receptor expression is detected by a primary antibody against the FLAG-tag in combination with an Alexa 488-labeled anti-mouse antibody (green). Plasma membrane is visualized by biotinylated concanavalin A binding to plasma glycoproteins in combination with streptavidin-Alexa 568 (red). Overlay pictures of the green and red channels are shown to measure the transfection efficiency. Scale bar, 100 μm .

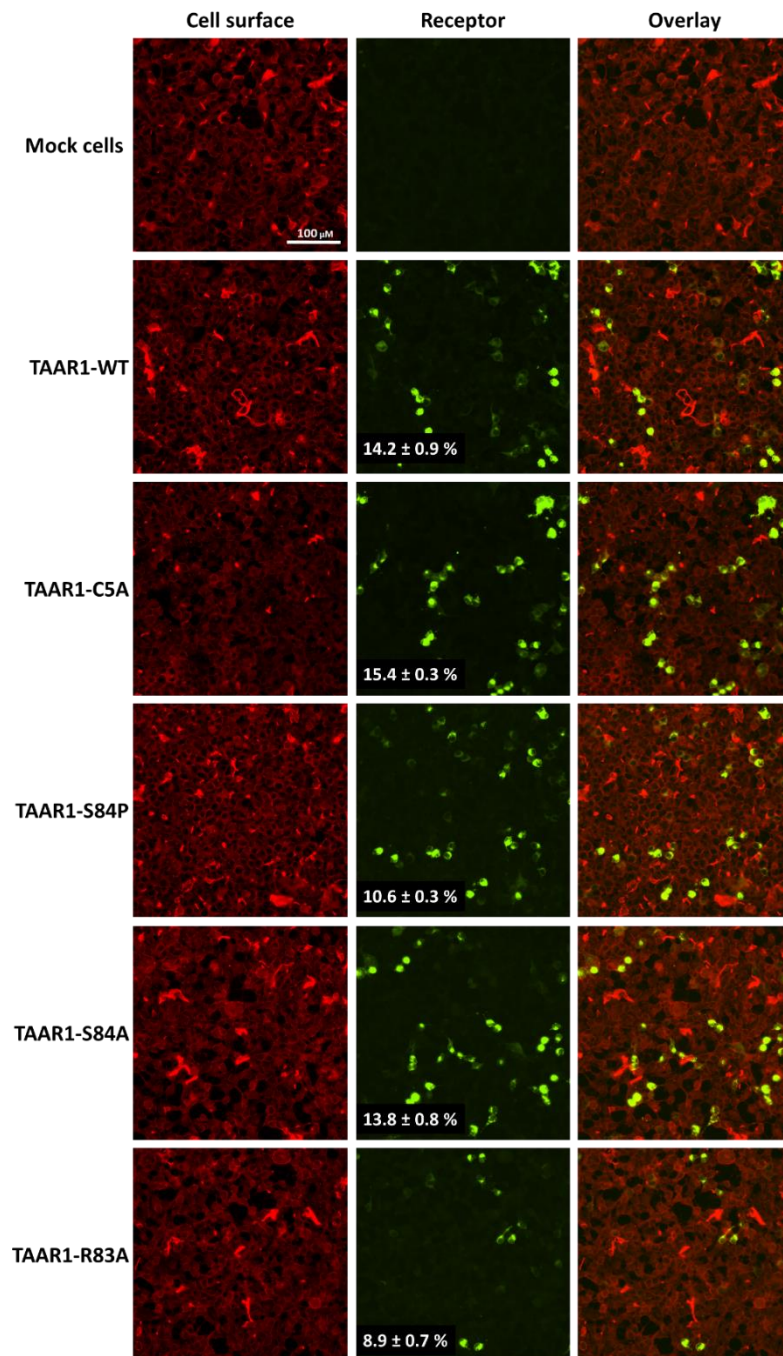


Figure S8: Expression of hTAAR1 after transient transfection into Hana3A cells. Epifluorescence images of and hTAAR1-WT, hTAAR5 SNP variants and mock transfected Hana3A cells are shown. Receptor expression is detected by a primary antibody against the FLAG-tag in combination with an Alexa 488-labeled anti-mouse antibody (green). Plasma membrane is visualized by biotinylated concanavalin A binding to plasma glycoproteins in combination with streptavidin-Alexa 568 (red). Overlay pictures of the green and red channels are shown to measure the transfection efficiency.

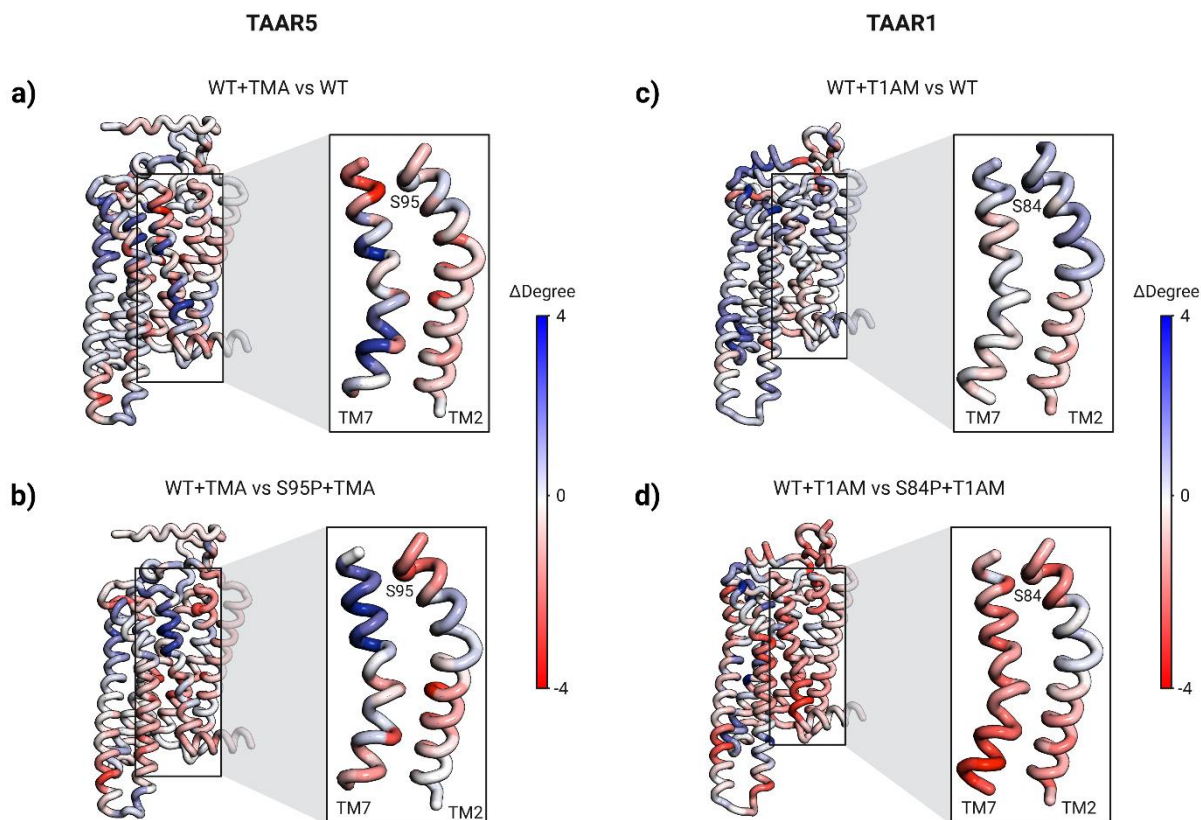


Figure S9: Network analysis of TAAR5 and TAAR1 MD simulations. Differences in degree centrality emphasize how residues are involved in the activation mechanism (**a** and **c** for TAAR5 and TAAR1 respectively) and how residues are affected by single point mutation (**b** and **d** for TAAR5 and TAAR1 respectively). Blue and red colors mean respectively that a residue is more or less central in the network compared to the reference.

References

1. Fleischer J, Breer H, Strotmann J (2009) Mammalian olfactory receptors. *Frontiers in Cellular Neuroscience* 3:
2. Bushdid C, Magnasco MO, Vosshall LB, Keller A (2014) Humans Can Discriminate More than 1 Trillion Olfactory Stimuli. *Science* 343:1370–1372. <https://doi.org/10.1126/science.1249168>
3. Buck L, Axel R (1991) A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* 65:175–187. [https://doi.org/10.1016/0092-8674\(91\)90418-X](https://doi.org/10.1016/0092-8674(91)90418-X)
4. Liberles SD, Buck LB (2006) A second class of chemosensory receptors in the olfactory epithelium. *Nature* 442:645–650. <https://doi.org/10.1038/nature05066>

5. Dunkel A, Steinhaus M, Kotthoff M, et al (2014) Nature's Chemical Signatures in Human Olfaction: A Foodborne Perspective for Future Biotechnology. *Angewandte Chemie International Edition* 53:7124–7143. <https://doi.org/10.1002/anie.201309508>
6. Keller A, Zhuang H, Chi Q, et al (2007) Genetic variation in a human odorant receptor alters odour perception. *Nature* 449:468–472. <https://doi.org/10.1038/nature06162>
7. Menashe I, Man O, Lancet D, Gilad Y (2003) Different noses for different people. *Nat Genet* 34:143–144. <https://doi.org/10.1038/ng1160>
8. Hasin-Brumshtein Y, Lancet D, Olender T (2009) Human olfaction: from genomic variation to phenotypic diversity. *Trends in Genetics* 25:178–184. <https://doi.org/10.1016/j.tig.2009.02.002>
9. Mainland JD, Keller A, Li YR, et al (2014) The missense of smell: functional variability in the human odorant receptor repertoire. *Nat Neurosci* 17:114–120. <https://doi.org/10.1038/nn.3598>
10. McRae JF, Mainland JD, Jaeger SR, et al (2012) Genetic Variation in the Odorant Receptor OR2J3 Is Associated with the Ability to Detect the “Grassy” Smelling Odor, cis-3-hexen-1-ol. *Chemical Senses* 37:585–593. <https://doi.org/10.1093/chemse/bjs049>
11. de March CA, Topin J, Bruguera E, et al (2018) Odorant Receptor 7D4 Activation Dynamics. *Angewandte Chemie International Edition* 57:4554–4558. <https://doi.org/10.1002/anie.201713065>
12. Gainetdinov RR, Hoener MC, Berry MD (2018) Trace Amines and Their Receptors. *Pharmacol Rev* 70:549–620. <https://doi.org/10.1124/pr.117.015305>
13. Freyberg Z, Saavedra JM (2020) Trace Amines and Trace Amine-Associated Receptors: A New Frontier in Cell Signaling. *Cell Mol Neurobiol* 40:189–190. <https://doi.org/10.1007/s10571-020-00800-x>
14. Rutigliano G, Zucchi R (2020) Molecular Variants in Human Trace Amine-Associated Receptors and Their Implications in Mental and Metabolic Disorders. *Cell Mol Neurobiol* 40:239–255. <https://doi.org/10.1007/s10571-019-00743-y>
15. Berry MD, Gainetdinov RR, Hoener MC, Shahid M (2017) Pharmacology of human trace amine-associated receptors: Therapeutic opportunities and challenges. *Pharmacology & Therapeutics* 180:161–180. <https://doi.org/10.1016/j.pharmthera.2017.07.002>
16. Gisladdottir RS, Ivarsdottir EV, Helgason A, et al (2020) Sequence Variants in TAAR5 and Other Loci Affect Human Odor Perception and Naming. *Current Biology* 30:4643–4653.e3. <https://doi.org/10.1016/j.cub.2020.09.012>
17. Amoore JE, Forrester LJ (1976) Specific anosmia to trimethylamine: The fishy primary odor. *J Chem Ecol* 2:49–56. <https://doi.org/10.1007/BF00988023>
18. Carnicelli V, Santoro A, Sellari-Franceschini S, et al (2010) Expression of Trace Amine-Associated Receptors in Human Nasal Mucosa. *Chem Percept* 3:99–107. <https://doi.org/10.1007/s12078-010-9075-z>
19. Wallrabenstein I, Kuklan J, Weber L, et al (2013) Human Trace Amine-Associated Receptor TAAR5 Can Be Activated by Trimethylamine. *PLOS ONE* 8:e54950. <https://doi.org/10.1371/journal.pone.0054950>
20. Ballesteros JA, Weinstein H (1995) [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In: Sealfon SC (ed) *Methods in Neurosciences*. Academic Press, pp 366–428

21. Kooistra AJ, Kuhne S, de Esch IJP, et al (2013) A structural chemogenomics analysis of aminergic GPCRs: lessons for histamine receptor ligand design. *British Journal of Pharmacology* 170:101–126. <https://doi.org/10.1111/bph.12248>
22. Scanlan TS, Suchland KL, Hart ME, et al (2004) 3-Iodothyronamine is an endogenous and rapid-acting derivative of thyroid hormone. *Nat Med* 10:638–642. <https://doi.org/10.1038/nm1051>
23. Rosenbaum DM, Cherezov V, Hanson MA, et al (2007) GPCR Engineering Yields High-Resolution Structural Insights into β 2-Adrenergic Receptor Function. *Science* 318:1266–1273. <https://doi.org/10.1126/science.1150609>
24. Gether U, Kobilka BK (1998) G Protein-coupled Receptors: II. MECHANISM OF AGONIST ACTIVATION *. *Journal of Biological Chemistry* 273:17979–17982. <https://doi.org/10.1074/jbc.273.29.17979>
25. Deupi X, Standfuss J (2011) Structural insights into agonist-induced activation of G-protein-coupled receptors. *Current Opinion in Structural Biology* 21:541–551. <https://doi.org/10.1016/j.sbi.2011.06.002>
26. Martínez-Morales JC, Solís KH, Romero-Ávila MT, et al (2022) Cell Trafficking and Function of G Protein-coupled Receptors. *Archives of Medical Research* 53:451–460. <https://doi.org/10.1016/j.arcmed.2022.06.008>
27. Cichero E, Espinoza S, Tonelli M, et al (2016) A homology modelling-driven study leading to the discovery of the first mouse trace amine-associated receptor 5 (TAAR5) antagonists. *Med Chem Commun* 7:353–364. <https://doi.org/10.1039/C5MD00490J>
28. Saito H, Kubota M, Roberts RW, et al (2004) RTP Family Members Induce Functional Expression of Mammalian Odorant Receptors. *Cell* 119:679–691. <https://doi.org/10.1016/j.cell.2004.11.021>
29. Wu L, Pan Y, Chen G-Q, et al (2012) Receptor-transporting Protein 1 Short (RTP1S) Mediates Translocation and Activation of Odorant Receptors by Acting through Multiple Steps *. *Journal of Biological Chemistry* 287:22287–22294. <https://doi.org/10.1074/jbc.M112.345884>
30. Yuan S, Chan HCS, Hu Z (2017) Using PyMOL as a platform for computational drug design. *WIREs Computational Molecular Science* 7:e1298. <https://doi.org/10.1002/wcms.1298>
31. Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31:455–461. <https://doi.org/10.1002/jcc.21334>
32. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* 7:525–537. <https://doi.org/10.1021/ct100578z>
33. Lomize MA, Pogozheva ID, Joo H, et al (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res* 40:D370-376. <https://doi.org/10.1093/nar/gkr703>
34. Schott-Verdugo S, Gohlke H (2019) PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. *J Chem Inf Model* 59:2522–2528. <https://doi.org/10.1021/acs.jcim.9b00269>
35. D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N.

- Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J., et al (2018) Amber18
36. Dickson CJ, Madej BD, Skjerveik ÅA, et al (2014) Lipid14: The Amber Lipid Force Field. *J Chem Theory Comput* 10:865–879. <https://doi.org/10.1021/ct4010307>
 37. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A., et al., (2009) Gaussian 09, Revision A.02. Wallingford CT,
 38. Wang J, Wolf RM, Caldwell JW, et al (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174. <https://doi.org/10.1002/jcc.20035>
 39. Roe DR, Cheatham TE (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* 9:3084–3095. <https://doi.org/10.1021/ct400341p>
 40. <https://getcontacts.github.io/>
 41. Tekpinar M, Neron B, Delarue M (2021) Extracting Dynamical Correlations and Identifying Key Residues for Allosteric Communication in Proteins by correlationplus. *J Chem Inf Model* 61:4832–4838. <https://doi.org/10.1021/acs.jcim.1c00742>

Chapter III
Olfactory Receptor
Extracellular Loops (ECLs)

G protein-coupled receptors have a fold composed of 7 transmembrane helices forming a bundle in which a binding cavity can be found usually 10 Å from the extracellular surface. The transmembrane helices are connected by several intra- and extracellular loops that serve to maintain the receptor structure. However, their role does not end there. Loops are often exposed outside or inside the cell, having the potential to interact with solvents, molecules, and proteins [1]. They therefore often have key roles in protein function. Extremely variable in shape and length (even when comparing different subtypes of the same receptor family), these parts have been found to be involved in several functions of class A GPCRs including: ligand binding, receptor selectivity, activation capacity and basal activity [2, 3].

We propose to further improve our knowledge of ECLs in the vertebrate ORs family through 2 studies on ECL2 and ECL3. The first study focuses on ECL2, and we highlight its critical role in the promiscuity and specificity of the ORs through a study mixing homology modeling, molecular dynamics, and functional tests on 2 mouse ORs with high homology but different recognition spectra. The second article deals with ECL3 in class II ORs. Using a molecular dynamics approach and *in vitro* assays, we identify the importance of ECL3 in the passage of a ligand to its active site and thus its importance in ligand recognition and receptor activation.

Reference 1: Yu Y, Ma Z, Pacalon J, Xu L, Li W, Belloir C, Topin J, Briand L, Golebiowski J, Cong X. *Extracellular loop 2 of G protein-coupled olfactory receptors is critical for odorant recognition.* *J Biol Chem.* **2022**, 298(9):102331. <https://doi.org/10.1016/j.jbc.2022.102331>

Contribution: Yiqun Yu, Jérôme Golebiowski and Xiaojing Cong designed the study. I helped curating the data with all the co-authors. I helped Xiaojing Cong setting up the molecular dynamics simulations and analyzing them. The research group of Yiqun Yu and Loïc Briand performed the *in vitro* assays.

Reference 2: Shim T, Pacalon J, Kim WC, Cong X, Topin J, Golebiowski J, Moon C. *The third extracellular loop of mammalian odorant receptors is involved in ligand binding.* **2022**, *IJMS*, under review

Contribution: Jérôme Golebiowski, Jérémie Topin and Cheil Moon designed the study. I set up, launched, and analyzed the molecular dynamics simulations with the help of Xiaojing Cong. Tammy Shim and WonCheol Kim performed the in vitro assays. Tammy Shim and I contributed equally as first authors.

1. Stevens AO, He Y (2022) Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction. *Biomolecules* 12:985. <https://doi.org/10.3390/biom12070985>
2. Wheatley M, Wootten D, Conner MT, et al (2012) Lifting the lid on GPCRs: the role of extracellular loops. *Br J Pharmacol* 165:1688–1703. <https://doi.org/10.1111/j.1476-5381.2011.01629.x>
3. Lawson Z, Wheatley M (2004) The third extracellular loop of G-protein-coupled receptors: more than just a linker between two important transmembrane helices. *Biochem Soc Trans* 32:1048–1050. <https://doi.org/10.1042/BST0321048>

Publication 3

Extracellular loop 2 of G protein-coupled olfactory receptors is critical for odorant recognition.

Yu Y, Ma Z, Pacalon J, Xu L, Li W, Belloir C, Topin J, Briand L, Golebiowski J, Cong X. Extracellular loop 2 of G protein-coupled olfactory receptors is critical for odorant recognition. J Biol Chem. 2022, 298(9):102331. <https://doi.org/10.1016/j.jbc.2022.102331>

Abstract

G protein–coupled olfactory receptors (ORs) enable us to detect innumerable odorants. They are also ectopically expressed in nonolfactory tissues and emerging as attractive drug targets. ORs can be promiscuous or highly specific, which is part of a larger mechanism for odor discrimination. Here, we demonstrate that the OR extracellular loop 2 (ECL2) plays critical roles in OR promiscuity and specificity. Using site-directed mutagenesis and molecular modeling, we constructed 3D OR models in which ECL2 forms a lid over the orthosteric pocket. We demonstrate using molecular dynamics simulations that ECL2 controls the shape and volume of the odorant-binding pocket, maintains the pocket hydrophobicity, and acts as a gatekeeper of odorant binding. Therefore, we propose the interplay between the specific orthosteric pocket and the variable, less specific ECL2 controls OR specificity and promiscuity. Furthermore, the 3D models created here enabled virtual screening of new OR agonists and antagonists, which exhibited a 70% hit rate in cell assays. Our approach can potentially be generalized to structure-based ligand screening for other G protein–coupled receptors that lack high-resolution 3D structures.

Introduction

G protein–coupled receptors (GPCRs) are the largest family of membrane proteins in the human genome, comprising over 800 members. Half of the human GPCR genes code for olfactory receptors (ORs) [1], which can discriminate an astonishing number of different odors [2]. ORs are also ectopically expressed in nonolfactory tissues, emerging as appealing drug targets [3–8]. GPCRs detect diverse ligands and control most of the cell signaling. Despite their diverse functions, GPCRs conserve a seven transmembrane helical (TM) architecture (TM1—TM7), connected by three extracellular loops (ECL1—ECL3) and three intracellular loops (ICL1—ICL3). ORs belong to class A GPCRs, which account for ~85% of the human GPCR genes. The orthosteric ligand-binding pocket in class A GPCRs is located within the extracellular half of the TM bundle, extending ~15 Å deep into the cell membrane [9]. The pocket may be solvent accessible (e.g., in receptors for peptides or soluble molecules) or shielded by ECL2 (e.g., in lipid receptors and rhodopsin) [10]. ECL2 is often the longest extracellular loop, which is highly variable in length, sequence, and structure [11, 12]. A disulfide bond between ECL2 and TM3

is conserved in 92% of human GPCRs [13]. It is important for ligand binding and receptor activation [10]. Peptide-activated GPCRs mostly contain an ECL2 in the form of a β -hairpin lying on the rim of the orthosteric pocket. ECL2 of GPCRs that are modulated by small-molecule endogenous ligands exhibits diverse shapes. They are often unstructured and cover partially or fully the pocket entrance [10]. Rhodopsin is a case in-between, in which a β -hairpin-shaped ECL2 inserts deep into the orthosteric pocket [14]. It has been suggested that rhodopsin ECL2 represents an evolutionary transition between peptide receptors and small-molecule receptors [12]. In small-molecule receptors, ECL2 may have evolved to mimic the peptide ligands and occupy part of the pocket, which renders the pocket suitable for binding small molecules. ECL2 plays important roles in ligand binding and activation of class A GPCRs [11]. It may act as a gateway to the orthosteric pocket [15–19], bind allosteric modulators [20, 21], or participate in receptor activation [22, 23].

ECL2 of ORs are among the longest in class A GPCRs. ORs can be promiscuous or highly specific, in which ECL2 may play a central role. However, the lack of high-resolution OR structures hampers the study of OR-odorant recognition. Homology modeling combined with site-directed mutagenesis have shed light on the structure and ligand specificity of the orthosteric pocket of various ORs [24–28]. Yet, the role and structure of ECL2 remain mostly elusive. In this work, we studied the role of ECL2 in two prototypical mouse ORs (mORs) of the same subfamily, mOR256-3 and mOR256-8, which share 54% sequence identity. Our previous work indicated that mOR256-3 is promiscuous for a series of commonly encountered odorants, whereas mOR256-8 is rather specific [29]. In this study, we found that ECL2 properties strongly modulate OR-odorant recognition. We performed site-directed mutagenesis along ECL2 and built 3D OR models that are in concordance with the mutagenesis data. Virtual screening using the 3D models identified new mOR256-3 ligands, including an antagonist that inhibited some of the agonists. The 3D models provide structural explanations to the promiscuity of mOR256-3 and the selective antagonism.

Results

Sequence analysis of OR ECL2

Sequence alignment of 1521 human and mORs showed that their ECL2 mostly contain 34 to 35 amino acids (Fig. S1). They are longer than ECL2 in most class A GPCRs. Three cysteines are highly conserved (C169, C179, and C189 in mOR256-3, conserved in 93.4%, 99.5%, and 95.0% of human and mORs, respectively). C179 forms the classic disulfide bond with TM3, whereas C169 and C189 have been suggested to form a second disulfide bond within ECL2 [30]. A few residues around the two disulfide bonds are highly conserved, whereas the rest of the OR ECL2 sequence displays low conservation (Fig. S1). It is plausible that the two disulfide bonds are important for ECL2 structuring and OR functions.

Nonspecific roles of ECL2 in OR responses to odorants

In our previous work, we screened diverse odorants at a near-saturating concentration (300 μ M) on several ORs in the heterologous Hana3A cells. We found a wide range of potential ligands for mOR256-3 but only two for mOR256-8 [29]. Yet, one or few point mutations in mOR256-8 could significantly expand its ligand spectrum [29]. Here, we reexamined 20 of these odorants at various concentrations in Hana3A cells expressing mOR256-3 or mOR256-8. Ten odorants activated mOR256-3 in a dose-dependent manner, including cyclic and acyclic alcohols, aldehydes, acids, ketones, and esters: R-carvone, coumarin, 1-octanol, allyl phenylacetate, benzyl acetate, citral, geraniol, 2-heptanone, octanal, and octanoic acid (Table S1 and Fig. S2A). mOR256-8 responded only to 1-octanol and geraniol in a dose-response manner, which are two primary acyclic alcohols of similar lengths (Table S1 and Fig. S2B).

Focusing on the role of ECL2, we performed site-directed mutagenesis to probe the residues that are responsible for the functional differences between mOR256-3 and mOR256-8. Based on the 3D models in our previous work [29, 31–34], we mutated 14 residues on TM3–TM6 around the orthosteric pocket, as well as 15 residues in ECL2 of mOR256-8 that differ from mOR256-3. In the narrowly tuned mOR256-8, these residues were mutated one by one into their counterpart in the broadly tuned mOR256-3. We then tested the response of the mutant

receptors to R-carvone and coumarin, two reference ligands of mOR256-3. While wild-type (wt) mOR256-8 does not respond to these odorants, 14 of the mutants showed dose-dependent responses to R-carvone, and some of them also responded to coumarin (Fig. 1A). Four of the mutations were in ECL2, R173I, N175D, L181V, and L184M (Fig. 1A). These residues flank the ECL2–TM3 disulfide bond, suggesting that this region (residues 173–184) is important for the receptor function. Five residues in this region are conserved in mOR256-8 and mOR256-3 (H176, F177, E180, P182, and A183). Therefore, we mutated these five residues in mOR256-3 to evaluate their role in this promiscuous receptor. They were mutated into alanine, except for A183, which was mutated into a bulky isoleucine. While F177A impaired receptor expression on the cell surface (Fig. S3), the other four mutations systematically diminished the receptor’s response to R-carvone and coumarin (Fig. 1B). The aforementioned mutations in the two receptors had less drastic impacts on the response to geraniol (Fig. S4), which suggest that geraniol interacts with the receptors in a different manner.

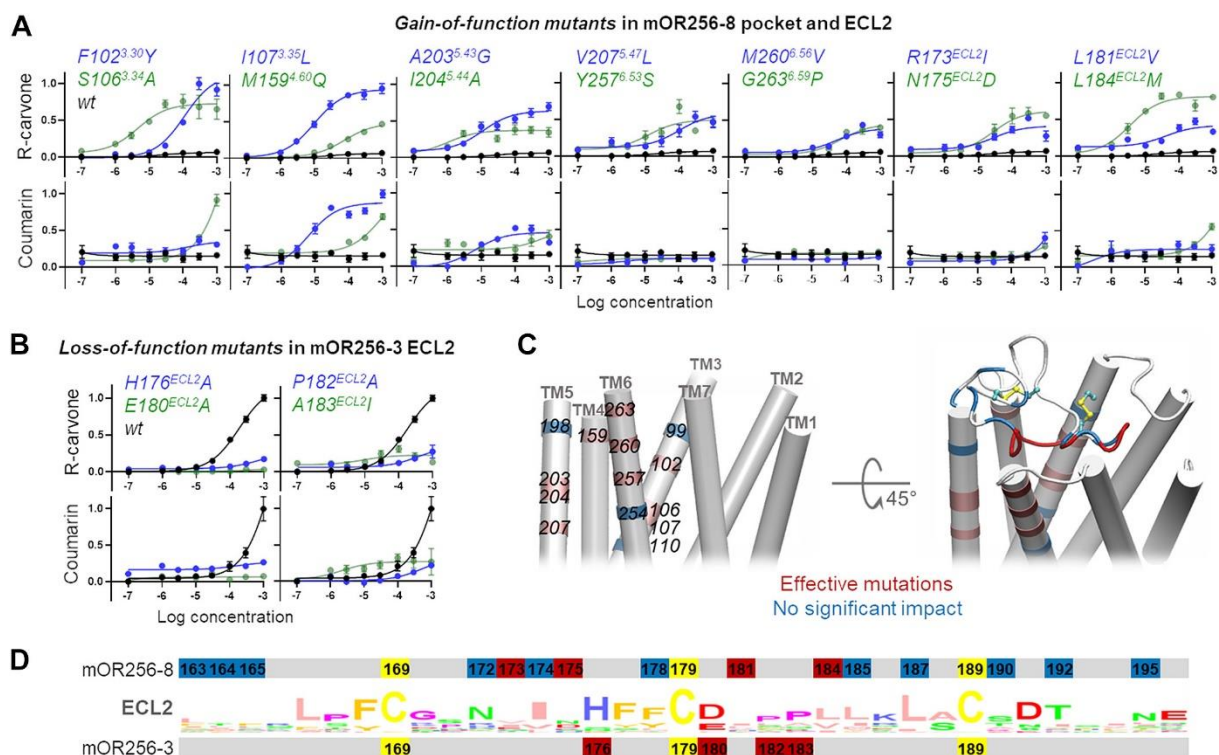


Figure 1: Site-directed mutagenesis and location of the mutation sites. Mutations in (A) mOR256-8 pocket and ECL2 and (B) mOR256-3 ECL2 affected the response to various odorants. Data are mean \pm SEM of three independent experiments. C, homology model of mOR256-3 selected according to the data in A and B. D, consensus ECL2 sequence of human and mouse ORs and location of the mutation sites. Effective mutations are colored in pink (in the pocket) or red (in ECL2). Noneffective mutations are colored in blue, including V99^{3.27}A, V110^{3.38}T, L198^{5.38}E, S254^{6.50}T, R172^{ECL2}N, I174^{ECL2}L,

L178^{ECL2}F, I185^{ECL2}L, M187^{ECL2}L, V190^{ECL2}T, A192^{ECL2}T, and V195^{ECL2}N in mOR256-8. In the 3D models, consistently, the noneffective mutation sites (blue) do not constitute the ligand-binding site or the pathway to the binding site. ECL2, extracellular loop 2.

We also generated a chimeric mOR256-8 in which ECL2 was replaced with that of mOR256-3. However, it did not gain response to the ligands of mOR256-3. The aforementioned data highlight that residues 173–184 in ECL2 are critical but not solely responsible for ligand recognition or receptor promiscuity. This is in line with the notion that in class A GPCRs, ECL2 acts as a vestibule or a molecular sieve of ligand binding and/or an allosteric site of receptor activation. Since residues 173–184 in ORs surround the conserved ECL2–TM3 disulfide bond, they are likely important in most, if not all, mammalian ORs. For instance, mutations in this region have dramatic impact on the response of mOR-EG to its odorants [28]. This region has also been found to interact with the orthosteric ligands in several nonolfactory class A GPCRs [11].

3D modeling explains OR promiscuity

To date, there are no high-resolution OR structures or structural information on the structural fold of OR ECL2. We generated three types of 3D models using AlphaFold 2 (DeepMind Technologies) [35], Modeller (University of California San Francisco) [36], and SWISS-MODEL (Swiss Institute of Bioinformatics) [37]. The three models displayed distinct structures in ECL2 (Figs. 1C and S5). We evaluated the predictivity of the models using site-directed mutagenesis data and docking. The model that best matched these data was generated by Modeller based on our hand-curated multiple sequence alignment (Fig. S6). In this model, ECL2 appears as an unstructured coil, in which residues 173–184 form a lid of the orthosteric pocket (Fig. 1C). Residues 180–183 may interact directly with the ligands (Fig. 1C). The model also suggests that the pocket of mOR256-3 is much larger than mOR256-8, showing two connected cavities (Fig. 2A). This may allow mOR256-3 to accommodate odorants of diverse size and shape. Molecular docking suggests that the upper cavity can accommodate the cyclic ligands, whereas the deeper cavity accommodates the acyclic ones (Figs. 2A and S7). The pocket of mOR256-8 shows only one small cavity for its acyclic ligands. We estimated the pocket volume of all the human and mORs by summing up the side-chain volume of the residues outlining the pocket with or without ECL2. We found that the pocket size of mOR256-

3 is ranked in the 47th and 46th percentile with and without ECL2, respectively, whereas that of mOR256-8 is at the 26th and 22nd, respectively (Fig. S8). Thus, the larger pocket volume of mOR256-3 than mOR256-8 may provide a structural explanation to the promiscuity of the former. In order to assess this hypothesis and the model predictivity, we use the model to virtually screen for new mOR256-3 ligands by molecular docking.

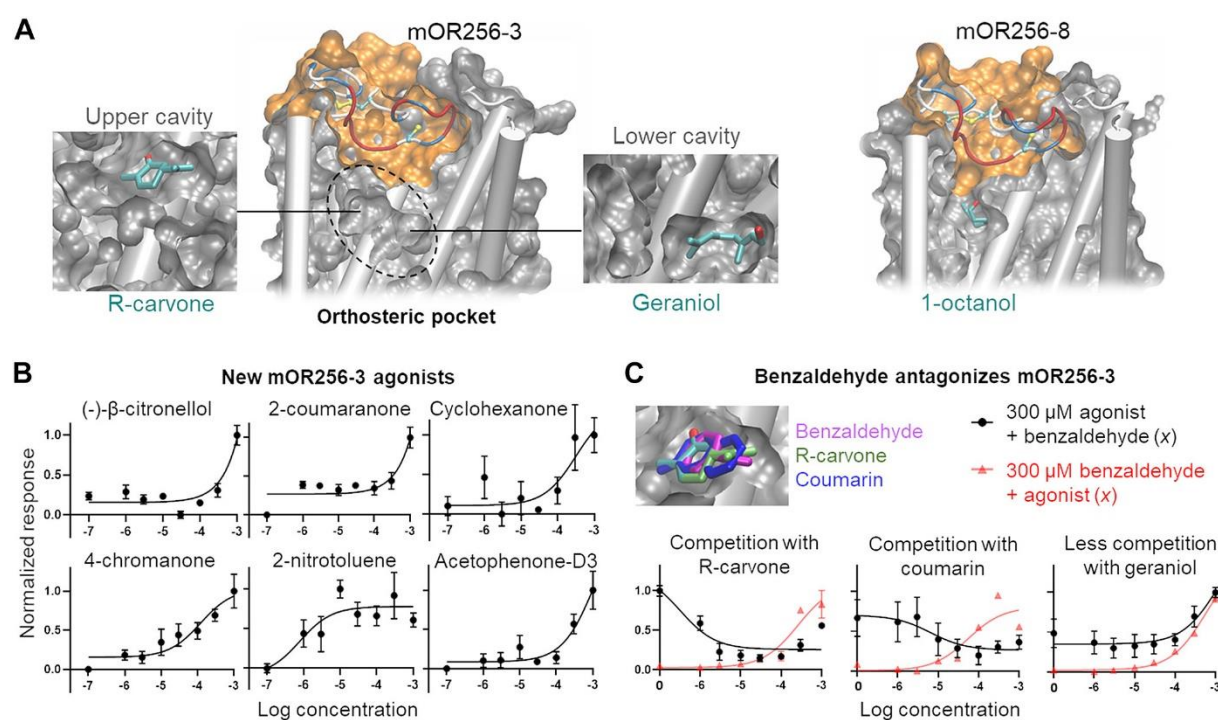


Figure 2: Selected 3D models and new mOR256-3 ligands discovered by virtual screening. A, cross-section of the best model of mOR256-3 and mOR256-8, illustrating ECL2 as the pocket lid. mOR256-3 displays two connected cavities in the pocket, in which the upper cavity binds cyclic ligands and the lower one accommodates acyclic molecules. B, dose-dependent curves of new mOR256-3 agonists from virtual screening. C, benzaldehyde binds in the same cavity as R-carvone and coumarin. It inhibits R-carvone, coumarin, and geraniol. Data are mean \pm SEM of three independent experiments. ECL2, extracellular loop 2.

Docking benchmarks were first performed with 52 compounds, including 10 known ligands of mOR256-3 and 42 decoys (Table S2) [29]. An ensemble-docking protocol (Fig. S9) was used to account for the conformational flexibility of the OR. Namely, enhanced sampling molecular dynamics (MD) simulations were performed on the initial model of mOR256-3 to sample the receptor conformations (see the Experimental procedures section for details). Ten receptor conformers (snapshots) were extracted from a clustering analysis of the MD trajectory. The 52

benchmark compounds were docked to each of the 20 conformers using AutoDock Vina (The Scripps Research Institute) [38] and ranked by their Vina scores for the given conformer. The “best” conformers were chosen as those that could best separate the ligands from the decoys by the Vina scores (Fig. S9). We performed this benchmarking process for our in-house model as well as for the models generated by AlphaFold 2 and SWISS-MODEL. The in-house model—generated by Modeller and selected according to site-directed mutagenesis data—gave the best predictions on the benchmark compounds (Table 1). Removing ECL2 from this model significantly reduced the predictivity (Table 1).

Table 1: Docking benchmark using different 3D models of mOR256-3 and 52 compounds

Model	MD snapshot ^a	MCC	Hit rate ^b (precision)	Recall ^c	True positive	True negative	False positive	False negative
In-house model	1	0.50	0.60	0.60	6	38	4	4
	2	0.26	0.40	0.40	4	36	6	6
In-house model without ECL2	1	0.26	0.40	0.40	4	36	6	6
	2	0.13	0.30	0.30	3	35	7	7
SWISS-MODEL	1	0.23	0.36	0.40	4	35	7	6
	2	0.20	0.33	0.40	4	34	8	6
AlphaFold 2	1	0.38	0.50	0.50	5	37	5	5
	2	0.26	0.40	0.40	4	36	6	6

^a Two snapshots that gave the best Matthew’s correlation coefficient (MCC) as a statistical measure of the model’s predictivity [39]. MCC returns a value between -1 (total disagreement between prediction and observation) and $+1$ (perfect prediction).

^b Hit rate or precision, the fraction of true ligands among the model predicted ones.

^c Recall indicates the fraction of true ligands retrieved by the model out of all the true ligands in the benchmark compounds.

Finally, we chose two best conformers of the aforementioned in-house model to virtually screen a library of 80 odorants in our laboratory (Tables S3 and S4). The screening returned 10 candidate compounds (Table S3), which were tested in functional assays in Hana3A cells. Six of them turned out to be mOR256-3 agonists and one (benzaldehyde) was an antagonist, giving

70% hit rate (Fig. 2B and C and Table S3). Benzaldehyde antagonized R-carvone, coumarin, and geraniol (Fig. 2C). Docking predicted that benzaldehyde may bind in the upper cavity of the mOR256-3 pocket for cyclic ligands, similar to R-carvone and coumarin (Fig. 2C).

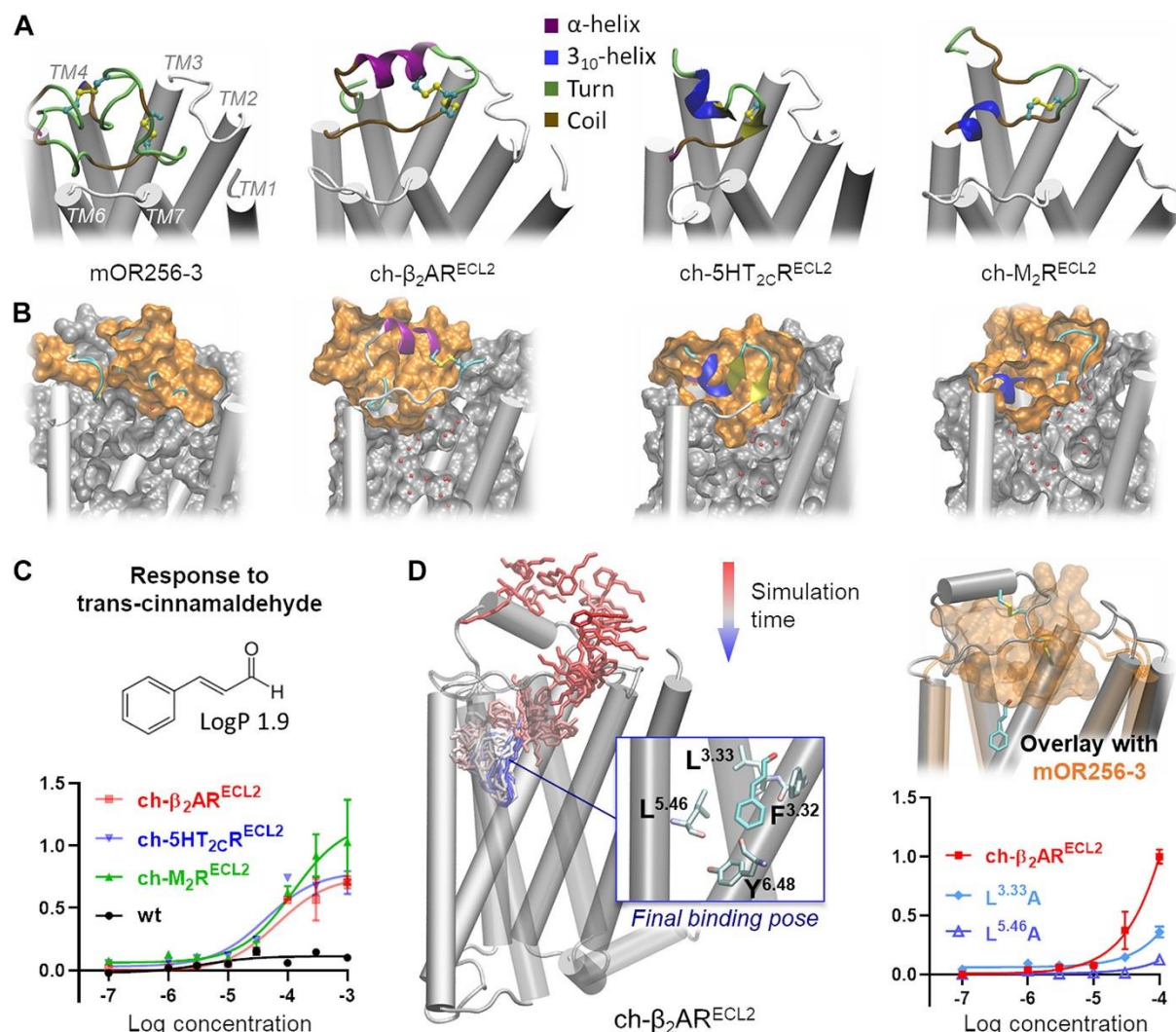


Figure 3: Structure models and functional assays of mOR256-3 chimeras. A, homology models of mOR256-3 variants with different ECL2 sequences and structures (in cartoon presentation, colored by secondary structure). B, the pocket of the chimeras was hydrated during molecular dynamics (MD) simulations without ligand in the pocket, whereas that of wt mOR256-3 remained dehydrated during the same simulation course. Shown here is the final MD simulation frame in cross-section. The water molecules within the pocket are shown in red balls, and the surface of ECL2 is shown in orange. C, dose-dependent responses of the three chimeras to trans-cinnamaldehyde. D, trans-cinnamaldehyde entered the pocket of $\beta_2\text{AR}^{\text{ECL2}}$ via the ECL2–TM7 gap during MD simulations. It adopted a binding pose that interacts with the toggle switch Y6.48. Mutating the trans-cinnamaldehyde-binding residues L3.33 and L5.46 diminished the receptor response to this ligand. An overlay with wt mOR256-3 (orange) shows a steric clash of trans-cinnamaldehyde with ECL2, which is likely the reason why wt

mOR256-3 does not respond to this odorant. Data are mean \pm SEM of three independent experiments. ECL2, extracellular loop 2; TM, transmembrane.

ECL2 controls pocket shape and hydrophobicity

To further examine the role of ECL2 in odorant recognition, we constructed three mOR256-3 chimeras, by replacing its ECL2 with that of M2 muscarinic receptor, β_2 adrenergic receptor, and 5HT serotonin 2C receptor, respectively (denoted as ch- β_2 AR^{ECL2}, ch-M2R^{ECL2}, and ch-5HT_{2C}R^{ECL2}). ECL2 of these receptors exhibit distinct structures (Fig. 3A). In Hana3A cells, the chimeras showed no significant response to the mOR256-3 ligands (Fig. S10). Nevertheless, they all displayed specific dose-dependent response to transcinnamaldehyde (Fig. 3B), whereas wt mOR256-3 does not respond to this odorant [29]. To understand how the chimeric mOR256-3 became specific receptors of transcinnamaldehyde, we built homology models for the chimeras and performed all-atom MD simulations in an explicit membrane–water environment. The homology models were built by assuming that ECL2 of the chimeras preserve the same fold as in β_2 AR, M₂R, and 5HT_{2C}R, respectively. The models illustrated that ECL2 of the chimeras only partly covered the ligand entrance. The orthosteric pocket of the chimeras was hydrated during the MD, whereas that of wt mOR256-3 was shielded from hydration by ECL2 (Fig. 3A). This might be the reason why the chimeras did not respond to the hydrophobic ligands of mOR256-3. Rather, they responded to the less hydrophobic transcinnamaldehyde (Fig. 3C).

We then added transcinnamaldehyde in the MD simulations of wt mOR256-3 and the chimeras to monitor the ligand binding. The ligand was initially placed at 10 Å above ECL2 and was restrained within a 15 Å radius around ECL2. Each system underwent 30 independent MD runs of 200 ns. We observed two binding events in ch- β_2 AR^{ECL2}, in which transcinnamaldehyde entered the orthosteric pocket near the toggle switch residue Y6.48 (Fig. 3D). It caused the side chain of Y6.48 to flip toward TM5, which is likely an early step of OR activation [32]. In the case of wt mOR256-3, ch-M₂R^{ECL2}, and ch-5HT_{2C}R^{ECL2}, transcinnamaldehyde associated with ECL2 but could not enter the pocket. The binding pose of transcinnamaldehyde in ch- β_2 AR^{ECL2} suggests that wt mOR256-3 cannot accommodate this ligand, since ECL2 occupies part of its pocket (Fig. 3D). Indeed, mOR256-3 ligands are generally smaller or more flexible than transcinnamaldehyde. The lack of ligand binding in ch-M₂R^{ECL2} and ch-5HT_{2C}R^{ECL2} was likely because of insufficient sampling of the ECL2 conformations in these very short simulations.

The entrance to the pocket is narrower in the initial models in ch-M₂R^{ECL2} and ch-5HT_{2c}R^{ECL2} than that in ch-β₂AR^{ECL2}. To verify the binding pose of transcinnamaldehyde observed in the MD simulations, we mutated three pocket residues that are in close contact with the ligand. Mutations L3.33A and L5.46A abolished the receptor response to transcinnamaldehyde (Fig. 3D). F3.32A impaired the receptor expression on the cell surface (Fig. S3) and is thus not discussed. The results suggest that the recognition of transcinnamaldehyde is specific to the orthosteric pocket, whereas ECL2 served as an unspecific molecular sieve for the ligand entrance.

Discussion

Mammalian OR sequences have highly diversified during evolution to detect and discriminate a vast spectrum of odorants. Specific (or narrowly tuned) ORs may be responsible for the detection of specific odorants or endogenous ligands when ectopically expressed in nonolfactory tissues [3–6]. Promiscuous (or broadly tuned) ORs may play important functions in olfaction, such as expanding the detection spectrum, diversifying the combinatorial code, and acting as general odor detectors or odor intensity analyzers [29]. Promiscuous ORs feature mostly nonpolar interactions in the orthosteric pocket with odorants, which are more adaptable to different odorant structures [33, 40]. Here, we showed that ECL2 is indispensable for OR promiscuity. ECL2 acts as a pocket lid to maintain the pocket hydrophobicity and also forms the upper part of the pocket to control its shape and volume. Its structural flexibility and mostly hydrophobic nature may tolerate diverse odorants, resulting in promiscuity. Indeed, in class A GPCRs, ECL2 may change conformations upon ligand binding and adopt different forms for different ligands [11]. The evolution of ECL2 in class A GPCRs is strongly coupled to that of the orthosteric pocket [12]. Therefore, class A GPCR–ligand recognition relies on the interplay between ECL2 and the orthosteric pocket. ECL2 may also take part in receptor activation via allosteric coupling with the receptor movements on the intracellular side [11]. However, this aspect is beyond the scope of the current study. Note that the 3D models reported here are not to present the exact structural fold of ECL2. Rather, they are to illustrate the approximate position of the ECL2 residues according to the mutagenesis data. Since mOR256-3 ECL2 features mostly nonpolar interactions with the odorants, such approximate models serve as suitable structural basis for ligand discovery, as demonstrated by the virtual screening

performance. The MD simulations based on these models are insufficient to sample the ECL2 conformational changes upon ligand binding. High-resolution OR structures may enable further investigations on this challenging question. Nevertheless, the models provide an explanation to competitive antagonism, which has been shown to be essential for the perception of odor mixtures [41]. Therefore, the models and the virtual screening approach established here may serve the design of biosensors with wide odor detection spectrum or specific odor maskers and/or drug candidates targeting ectopic ORs in nonolfactory tissues.

Experimental procedures

Chemicals and OR constructs

Odorants were purchased from Sigma–Aldrich. They were dissolved in dimethyl sulfoxide to make stock solutions at 1 mM and then diluted freshly in optimal MEM (Thermo Fisher Scientific) to prepare the odorant stimuli. The OR constructs were kindly provided by Dr Hiroaki Matsunami (Duke University). Site-directed mutants were constructed using the Quikchange site-directed mutagenesis kit (Agilent Technologies). The sequences of all plasmid constructs were verified by both forward and reverse sequencing (Sangon Biotech).

Chimera construction

All chimeras were constructed by three PCR steps with modification [42]. Briefly, two fragments were amplified from the mOR256-3, whereas ECL2 of β_2 AR, M₂R, and 5HT_{2c}R was synthesized by Sangon Biotech Co. The primers were partially complementary at their 5' ends to the adjacent fragments, necessary to fuse the different fragments together. Three fragments were purified and fused together in a second PCR step. Equal amount of each fragment was mixed with dNTP and Phusion High-Fidelity DNA Polymerase (NEB) in the absence of primers. The PCR program consisted of 10 repetitive cycles with a denaturation step at 98 °C for 10 s, an annealing step at 55 °C for 30 s, and an elongation step for 30 s at 72 °C. The third step corresponded to the PCR amplification of the fusion product using the primers of mOR256-

3. The PCR product was purified and ligated into PCI vector. The sequences of all chimeras were verified by both forward and reverse sequencing.

Cell culture and transfection

We used Hana3A cells, a human embryonic kidney 293T–derived cell line that stably expresses receptor-transporting proteins (RTP1L and RTP2), receptor expression–enhancing protein 1 (REEP1), and olfactory G protein ($G\alpha_{olf}$) [43]. The cells were grown in MEM (Corning) supplemented with 10% (v/v) fetal bovine serum (FBS; Thermo Fisher Scientific) plus 100 $\mu\text{g}/\text{ml}$ penicillin–streptomycin (Thermo Fisher Scientific), 1.25 $\mu\text{g}/\text{ml}$ amphotericin (Sigma–Aldrich), and 1 $\mu\text{g}/\text{ml}$ puromycin (Sigma–Aldrich).

All constructs were transfected into the cells using Lipofectamine 2000 (Thermo Fisher Scientific). Before the transfection, the cells were plated on 96-well plates (NEST) and incubated overnight in MEM with 10% FBS at 37 °C and 5% CO₂. For each 96-well plate, 2.4 μg of pRL-SV40 (simian virus 40), 2.4 μg of CRE-Luc, 2.4 μg of mouse RTP1S, and 12 μg of receptor plasmid DNA were transfected. The cells were subjected to a luciferase assay 24 h after transfection.

Luciferase assay

The luciferase assay was performed with the Dual-Glo Luciferase Assay Kit (Promega) following the protocol [43]. OR activation triggers the $G\alpha_{olf}$ -driven AC-cAMP-PKA signaling cascade and phosphorylates cAMP response element–binding protein. Activated cAMP response element–binding protein induces luciferase gene expression, which can be quantified luminometrically (measured here with a bioluminescence plate reader [MD SPECTRAMAX L]). Cells were cotransfected with firefly and Renilla luciferases where firefly luciferase served as the cAMP reporter. Renilla luciferase is driven by a constitutively active SV40 promoter (pRL-SV40; Promega), which served as a control for cell viability and transfection efficiency. The ratio between firefly luciferase versus Renilla luciferase was measured. Normalized OR activity was calculated as $(LN - L_{\min}) / (L_{\max} - L_{\min})$, where LN is the luminescence in response to the odorant, and L_{\min} and L_{\max} are the minimum and maximum luminescence values on a plate, respectively. The assay was carried out as follows: 24 h after transfection, medium was

replaced with 100 μ l of odorant solution (at different doses) diluted in Optimal MEM, and cells were further incubated for 4 h at 37 °C and 5% CO₂. After incubation in lysis buffer for 15 min, 20 μ l of Dual-Glo Luciferase Reagent was added to each well of 96-well plate, and firefly luciferase luminescence was measured. Next, 20 μ l Stop-Glo Luciferase Reagent was added to each well, and Renilla luciferase luminescence was measured. Data analysis followed the published procedure [43]. Three-parameter dose–response curves were fitted with GraphPad Prism 9 (GraphPad Software, Inc).

Flow cytometry analysis

Hana3A cells were seeded in 35 mm dishes. The cells were cultured overnight to >80% confluence and transfected with 0.3 μ g RTP1S, 0.3 μ g GFP, and 0.8 μ g OR plasmid by Lipofectamine 2000. At 24 h after transfection, the cells were stripped with TrypLE Express Enzyme (Thermo Fisher Scientific) and then kept in round bottom polystyrene tubes on ice. The cells were spun down at 200g for 3 min at 4 °C and resuspended in PBS containing 2% FBS and 15 mM NaN₃. They were incubated with primary antibody mouse antirhodopsin for 45 min and then with phycoerythrin (PE)-conjugated donkey antimouse immunoglobulin G (Jackson ImmunoResearch; catalog no.: 715-116-150) in the dark for 30 min on ice. After washing twice, the cells were analyzed using Beckman Coulter CytoFLEX with gating for GFP positive, single, viable cells. The measured PE fluorescence intensities were analyzed and visualized using FlowJo (BD), version 10. The PE fluorescence intensity was normalized to the average value of wt ORs for statistical analysis.

Molecular modeling

The in-house models of mOR256-3 and mOR256-8 were generated with Modeller 9.21 [36] using our hand-curated sequence alignment to four structure templates: human α 2AR (Protein Data Bank [PDB] ID: 2YDV), human CXCR1 (PDB ID: 2LNL), human CXCR4 (PDB ID: 3ODU), and bovine rhodopsin (PDB ID: 1U19). The N and C termini were excluded. The template structures are all in inactive state. The sequence similarity between the templates and the two target ORs ranged from 31% to 38%. In the TM regions, the sequence similarity was 38–44%. For the three chimeras, the ECL2 structure of β ₂AR (PDB ID: 2RH1), M₂R (PDB ID:

3UON), and 5HT_{2C}R (PDB ID: 6BQH), respectively, was used as templates for the ECL2. For each receptor, 2500 models were generated and ranked by the DOPE score [44]. The 250 top ranked models were selected and clustered using the k-means algorithm. We obtained five clusters for each receptor and selected a representative model that was the most compatible with the mutagenesis data. The SWISS-MODELS were generated using the SWISS-MODEL webserver [37] and the target OR sequence. Template search and model building were performed using default settings of the webserver. The AlphaFold 2 models [35] were generated using the API hosted at the Söding Laboratory based on the MMseqs2 server [45]. Using the target OR sequence as input, the models were generated using the parameters [35]. Docking was performed with AutoDock Vina [38]. The receptors were prepared with AutodockTools to add nonpolar hydrogens and Gasteiger charges. A grid box was set to encompass the pocket and the lid, with a 0.375 Å grid point spacing. Initial 3D coordinates of the ligands were generated using Balloon (Åbo Akademi University) [46] and converted by AutoDock Raccoon (The Scripps Research Institute) for the docking [47]. Pocket residues and ligand rotatable bonds were set flexible. For virtual screening, however, pocket residues were kept rigid and multiple receptor conformers were used. Other parameters for the docking were left as their default values.

MD simulations

The receptor N and C termini were truncated at residues 23 and 305, respectively. Protonation state of titratable residues in the receptors were predicted at pH 7 using the H++ server [48]. The receptors or receptor-odorant complexes were embedded in a bilayer of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine using PACKMOL-Memgen (Heinrich Heine University Düsseldorf) [49]. Each system was solvated in a periodic $75 \times 75 \times 105 \text{ \AA}^3$ box of explicit water and neutralized with 0.15 M of Na⁺ and Cl⁻ ions. Effective point charges of the ligands were obtained by restrained electrostatic potential fitting [50] of the electrostatic potentials calculated with the HF/6-31G* basis set using Gaussian 09 [51]. The Amber 14SB [52], lipid 14 [53], and GAFF [54] force fields were used for the proteins, lipids, and ligands, respectively. The TIP3P model [55] and the Joung–Cheatham parameters [56] were used for the water and the ions, respectively.

The process of ligand binding was simulated with 30 runs of 200 ns of all-atom brute-force MD for each OR–ligand pair using Amber18. The ligand was initially placed 10 Å above ECL2.

After energy minimization, each system was gradually heated to 310 K with a restraint of 200 kcal/mol on the receptor and ligand. This was followed by 5 ns of pre-equilibration with a restraint of 5 kcal/mol and 5 ns of unrestrained equilibration. Bonds involving hydrogen atoms were constrained using the SHAKE algorithm [57], allowing for a 2-fs time step. van der Waals and short-range electrostatic interactions were cut off at 12 Å. Long-range electrostatic interactions were computed using particle mesh Ewald [58] method with a Fourier grid spacing of 1.2 Å. During the production run, when the ligand exceeded 15 Å from the center of ECL2, a distance restraint of 10 kcal/mol was applied to drive the ligand toward the center. Finally, the trajectories were visualized with VMD 1.9.2 (University of Illinois Urbana-Champaign) to inspect the binding events.

To thoroughly sample the conformations of mOR256-3 for ensemble docking, we used an enhanced sampling technique, replica exchange with solute scaling 2 (REST2) [59]. REST2 MD was performed with 48 replicas in the NVT ensemble using Gromacs 5.1 (University of Groningen, Uppsala Universitet) [60] patched with the PLUMED 2.3 plugin (the PLUMED consortium) [61]. The protein and ligands were considered as “solute” in the REST2 scheme. The force constants van der Waals, electrostatic, and dihedral terms of the protein and ligands were scaled down to facilitate conformational changes. The effective temperatures used for generating the REST2 scaling factors ranged from 310 to 700 K, following a distribution calculated with the Patriksson–van der Spoel approach [62]. Exchange between replicas was attempted every 1000 simulation steps. This setup resulted in an average exchange probability of ~40%. A total of 60 ns × 48 replicas of REST2 MD was carried out. The first 10 ns were discarded for equilibration, and only the original unscaled replica (at 310 K effective temperature) was collected. The Gromacs clustering tool was used to analyze the simulation trajectory. An RMSD-based clustering was performed on the C α atoms using the GROMOS method [63] and a 1 Å cutoff. The representative frames of the top 20 clusters (covering 97% of the trajectory) were extracted for ensemble docking.

Data availability

All data generated or analyzed during this study are included in this published article and its supporting information files.

Supporting information

This article contains supporting information including Tables S1—S4 and Figures S1—S10. MD simulation trajectories of mOR256-3 and transcinamaldehyde binding to ch- β_0 AR^{ECL2} are available at <https://mycore.core-cloud.net/index.php/s/MkGBt36XDBbMeGz>. Only the protein and ligand are shown for clarity [29, 34, 64].

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions

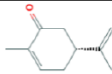

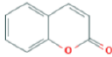

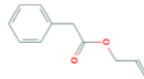

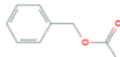
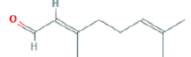

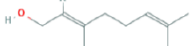
Y. Y., J. G., and X. C. conceptualization; J. P., J. G., and X. C. software; Y. Y., J. P., J. G., and X. C. validation; Y. Y., Z. M., J. P., L. X., W. L., C. B., J. T., L. B., and X. C. formal analysis; Y. Y., Z. M., J. P., L. X., W. L., C. B., J. T., L. B., and X. C. investigation; Y. Y., J. P., and X. C. resources; Y. Y., Z. M., J. P., L. X., W. L., C. B., J. T., L. B., and X. C. data curation; Y. Y., J. G., and X. C. writing—original draft; Y. Y., J. G., and X. C. writing—review & editing; Y. Y. and X. C. visualization; Y. Y., J. G., and X. C. supervision; J. G. and X. C. project administration; Y. Y., J. P., J. G. and X. C. funding acquisition.

Funding and additional information

This work was supported by the National Natural Science Foundation of China (grants 32070996 and 31771155; to Y. Y.), the NeuroMod Institute of the University of Côte d'Azur (2019–2020 project grant), the Roudnitska Foundation in France (2018–2021 research fellowship; to J. P.), GIRACT in Switzerland (2019 research fellowship; to J. P.), the Science and Technology Commission of Shanghai Municipality (grant 21140900600; to Y. Y.), and the Eye, Ear, Nose, and Throat Hospital, Fudan University, Excellent Doctors-Excellent Clinical Researchers Program (grant no.: SYB202002; to Y. Y.).

Supporting information

Table S1: Structure, hydrophobicity and potency of mOR256-3-wt ligands.

Odorant	Structure	LogP ^a	Odorant	Structure	LogP ^a
R-carvone		2.7	1-octanol		3.0
Coumarin		2.4	Octanal		3.5
Allyl phenylacetate		~2.4 ^b	Octanoic acid		3.1
Benzyl acetate		2.0	Citral		3.5
2-heptanone		2.0	Geraniol		3.6

^a data from PubChem

^b computed by XLogP3

Table S2: Decoy^a compounds used for docking benchmark.

Odorant	PubChem CID	Odorant	PubChem CID
R-(+)-Pulegone	442495	Diethyl sebacate	8049
1-Butanol	263	d-limonene	440917
2,3-Hexanedione	19707	Ethanol	702
2,5-Dimethylpyrazine	31252	Ethyl acetate	8857
2,5-Dimethylpyrrole	12265	Eugenol	3314
Ambrette	6753	Furfural	7362
2-Methoxy-4-methylphenol	7144	Hexyl octanoate	14228
2-Octanone	8093	Isobutylamine	6558
3-Methyl-2-butanol	11732	Isobutyraldehyde	6561
Acetaldehyde	177	Isobutyric acid	6590
Acetophenone	7410	Lilial	228987
Allyl hexanoate	31266	linalyl	91604
Ammonium hydroxide	14923	m-Cresol	342
Amyl butyrate	10890	Musk ketone	6669
Amyl laurate	62571	Propyl acetate	7997
α -Phellandrene	7460	Pyridine	1049
Benzyl alcohol	244	Pyrrolidine	31268
Benzyl salicylate	8363	Thymol	6989
Phenethylamine	1001	Toluene	1140
Cyclohexylamine	7965	Cinnamaldehyde	637511
Diacetyl	650	Triethylamine	8471

^a Non-effective compounds from the single-dose screening in ref. [29], except for 5 compounds which showed activities in dose-dependent assays in ref. [39] and this work.

Table S3: Ten candidate compounds selected from virtual screening for functional assays. The compounds were selected using the procedure shown in Fig. S9.


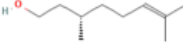
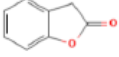
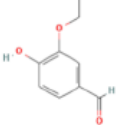
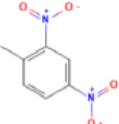

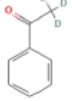
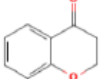
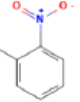

Name	PubChem CID	Structure	SMILES
Cyclohexanone	7967		<chem>C1CCC(=O)CC1</chem>
(-)-β-Citronellol	7793		<chem>C[C@@H](CCC=C(C)C)CCO</chem>
2-Coumaranone	68382		<chem>C1C2=CC=CC=C2OC1=O</chem>
Ethyl vanillin	8467		<chem>CCOC1=C(C=CC(=C1)C=O)O</chem>
2,4-DNT	8461		<chem>CC1=C(C=C(C=C1)[N+](=O)[O-])[N+](=O)[O-]</chem>
2-Ethyl fenchol	106997		<chem>CCC1(C(C2CCC1(C2)C)(C)C)O</chem>
Acetophenone-D3	140244		<chem>[2H]C([2H])([2H])C(=O)C1=CC=CC=C1</chem>
4-chromanone	68110		<chem>C1COC2=CC=CC=C2C1=O</chem>
2-Nitrotoluene	6944		<chem>CC1=CC=CC=C1[N+](=O)[O-]</chem>
Benzaldehyde	240		<chem>C1=CC=C(C=C1)C=O</chem>

Table S4: Compounds used in virtual screening in addition to those in Table S3.

Name	Pubchem CID	Name	Pubchem CID
Nonanoic acid	8158	Isopropyl Mercaptan	6364
Decanoic acid	2969	Undecanoic acid	8180
Heptanoic acid	8094	Pyrene	31423
Benzophenone	3102	Fluoranthene	9154
Ethyl isobutyrate	7342	Phenethyl acetate	7654
2-Methyl-1-propanethiol	10558	Undecanedioic acid	15816
Butyl formate	11614	β -ionone	638014
Alpha-terpinyl acetate	111037	Sebacid	5192
Butyl butyryl lactate	24114	1-Methylcyclopropene	151080
Methyl salicylate	4133	β -nadh tetrasodium salt	5884
Shoyu pyrazine	27458	1-Heptanethiol	15422
Isoamyl octanoate	16255	Oxoazelaic acid	269945
Dimethyl sulfide	1068	1-(methylthio)-octane	77289
Dextro-sorbitol	61431	Decane-1-10-dithiol	14494
2-Butanone	6569	2-methylbutane-1-thiol	15877
4-Methylvaleric acid	12587	Nonane-19-dithiol	248488
Ethylene brassylate	61014	Pentane-15-dithiol	70236
Guaiacol	460	3-Methyl-2-butanethiol	519823
Isovaleric acid	10430	1-Propanol	1031
Methanethiol	878	Isopropanol	3776
Pyrazine	9261	2-Methyl-2-propanol	6386
Terpineol	17100	2-butanol	6568
1-Hexanethiol	8106	(R)-(+)-1-Phenylethanol	637516
MTMT	122370	Phenetole	7674
NN-Dimethylethylamine	11723	Pyrrole	8027
2-Mercaptopyrimidine	1550489	Piperidine	8082
thioacetic acid	10484	β -Damascone	5374527
Trans-cyclo octene	5463599	Isopropyl tiglate	5367745
1-(methylthio)ethanethiol	525462	TNT	8376
(methylsulfanyl)methane	93236	26-bis(trimethylsilyl)benzenethiol	15376349
Cis-cyclooctene	638079	2-(Naphthyl)ethylamine hydrochloride	16218122
Hexanoic-66-D3 acid	12222599	N-methyl-2-phenylethylamine	11503
Methyl DL-lactate	11040	Sodium hydrosulfide	28015
Testosterone	6013	Curcumin	969516
2-Methyl-2-pentanethiol	74213	Trimethylamine	16387

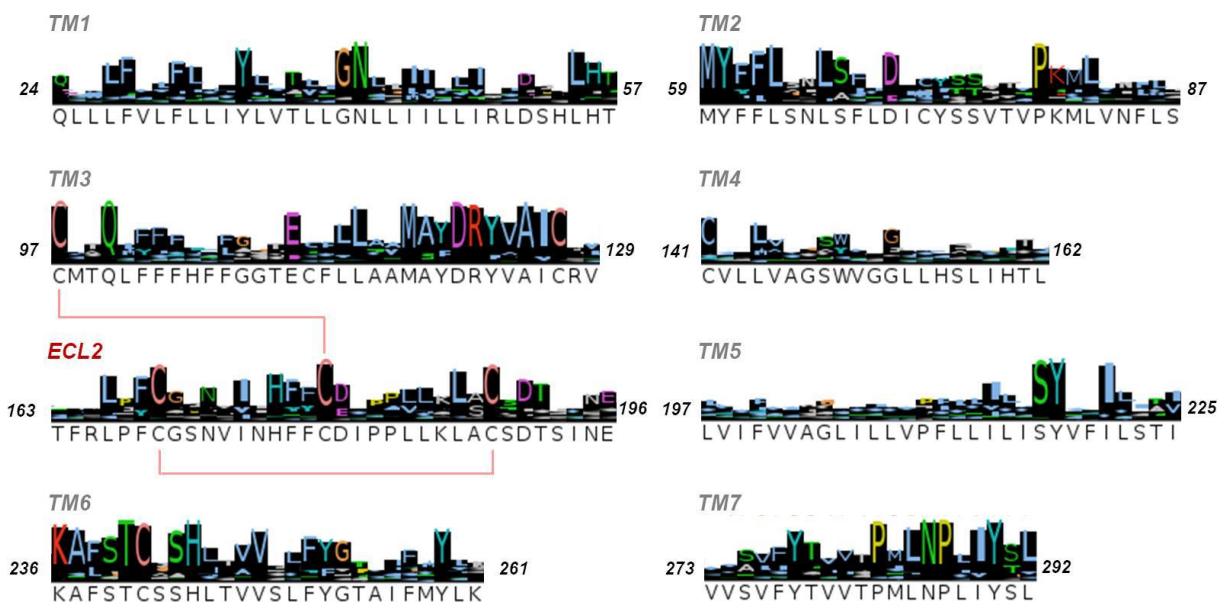


Figure S1: Consensus sequence of the TM regions and ECL2 in human and mouse ORs. Residue numbers in mOR256-3 are labeled on both sides of each region. Histogram indicates sequence conservation.

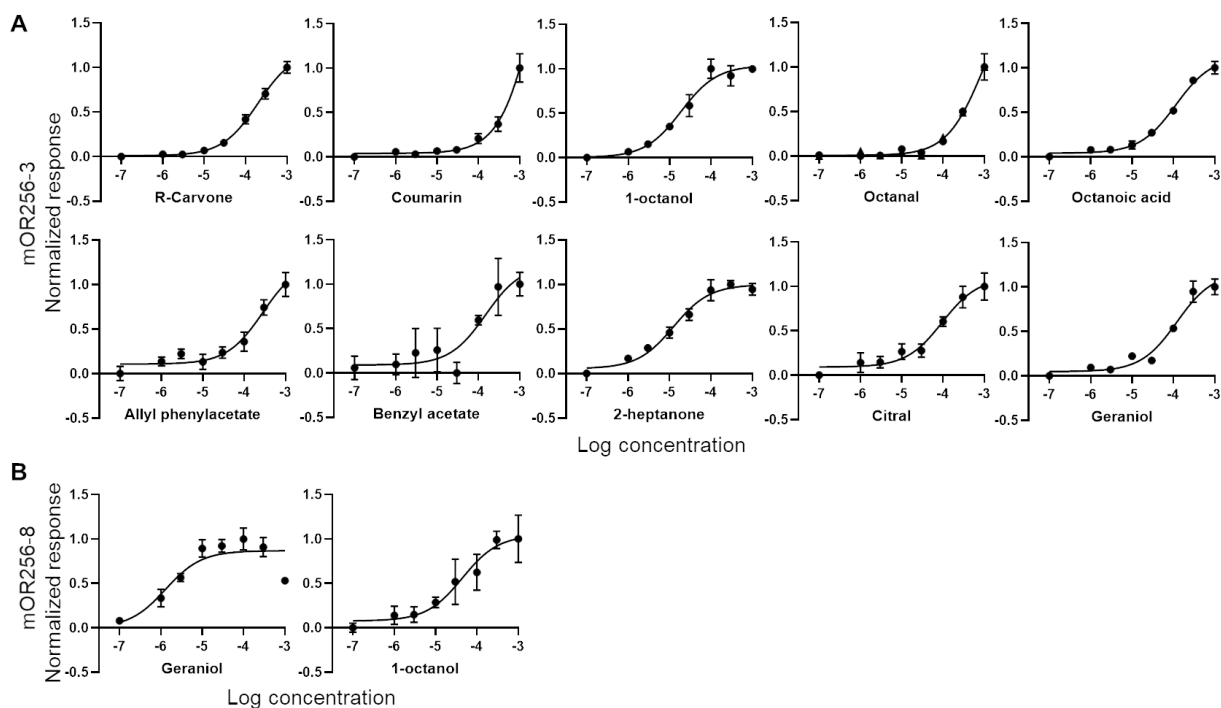


Figure S2: Dose-dependent response curves of (A) mOR256-3 and (B) mOR256-8 to their ligands. Data are mean \pm SEM of 3 technical repeats.

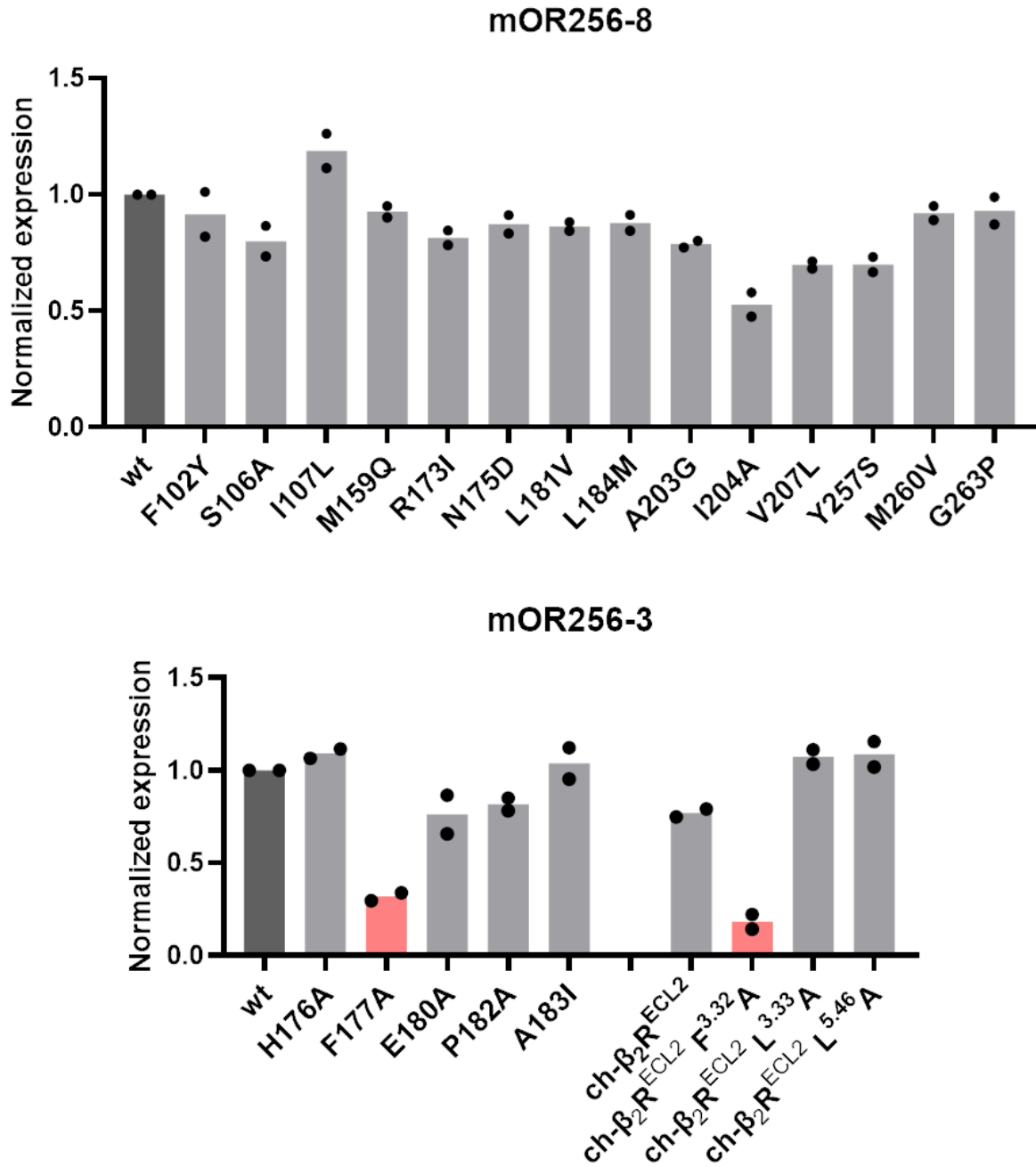


Figure S3: Cell-surface expression level of mOR256-3 and mOR256-8 variants relative to the wt receptor. Two technical repeats were performed.

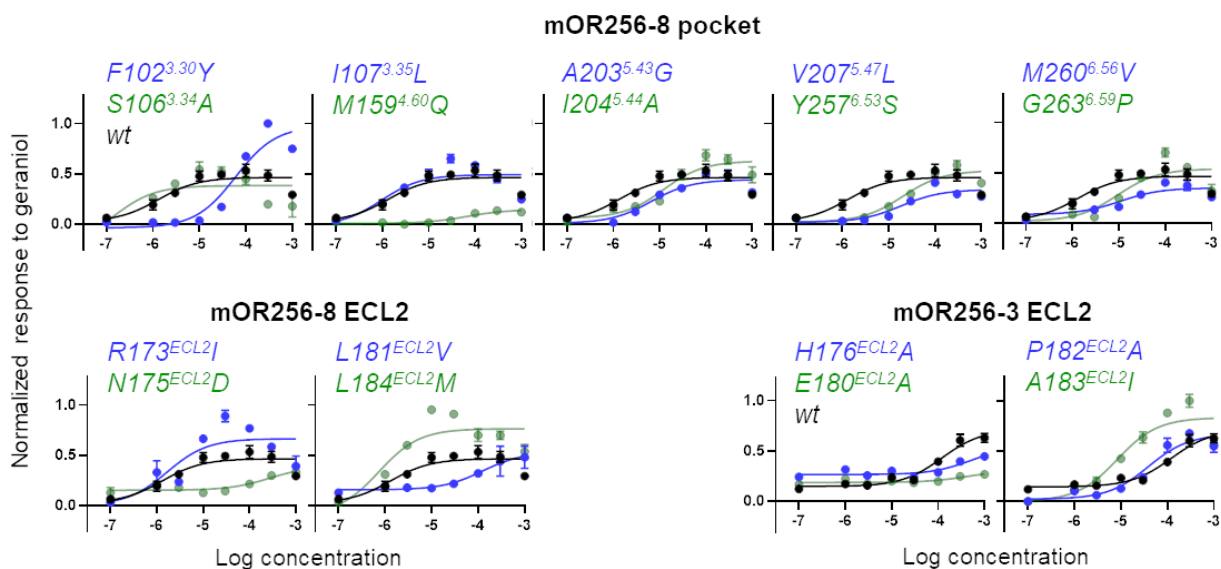


Figure S4: Dose-dependent response curves of mOR256-3 and mOR256-8 variants to geraniol. Data are mean \pm SEM of 3 technical repeats.

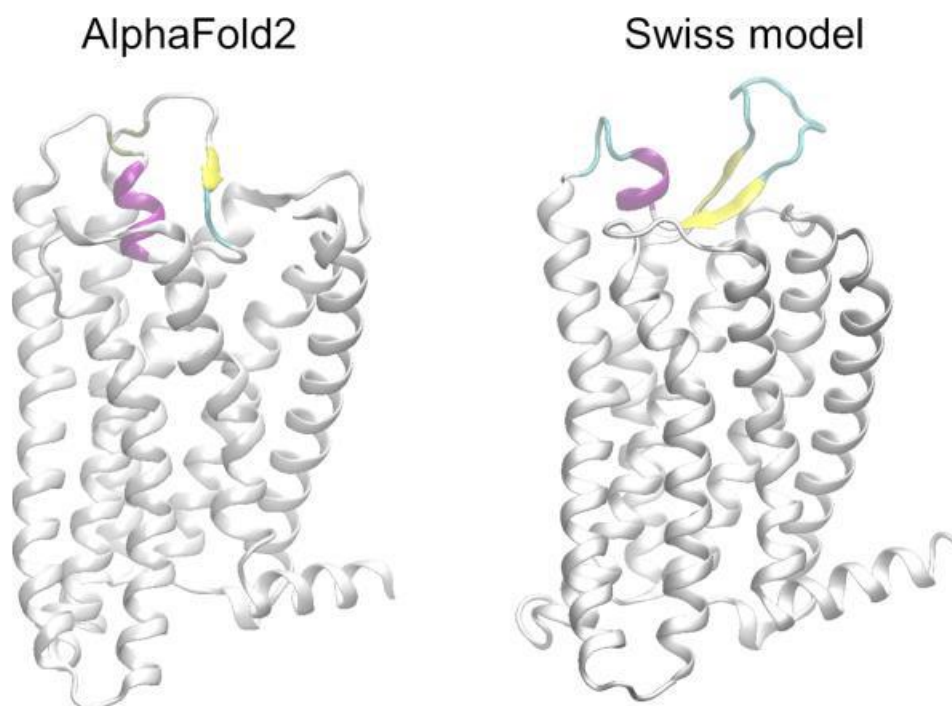


Figure S5: mOR256-3 models built by AlphaFold2 and Swiss model. ECL2 is colored by secondary structures. The N- and C-termini are neglected.

```

bRho      YYLAEPWQFSMLAAYMFLIMLGFPINFLTLTYVTVQHK-KLRTPLNYILLNLAVADLFMFVGGFTTTLTSLHG YFV----FG
CXCR1     TETLNKYVVI IAYALVFLLSLLGNLSLVM LVLILYSR----VGRSVTDVYLLNLALADLLF-ALTLP IWAASKVN-GWI----FG
A2a       MPIMGSSVYITVELAIAVLA I LGNVLVCWAVWLSN----NLQNVNTYFVVSAAAADI LVGVLAIPFAIAISTG--FC----AA
CXCR4     NANFNKIFLPTIYSIIFLTGIVGNGLVILVMGYQK----KLRSMTDKYRLHLSVADLLF-VITLPFWAVDAVA-NWY----FG
mOR256-3  DRPWLETPLFVIFLVAYIFALFGNISII LVSRLDP----QLDSPMYFFVSNLSLLDL CYTTSTVPQMLVNLRGPEKT----IS
mOR256-8  DRPRLEMVLFIVNFTLYSVAVLGNITII LVCILDP----RLHTPMYFFLANLSFLDL LCFSTSCI PQMLVNLWGPDKT----IS

bRho      PTG C NLEGGFFATLGGEIALWSLVVLA I ERYVVVCKPMSNFR-FGENHAIMGVAFTWVMALACAAPPLVGWSR----YIPEG--
CXCR1     TFLCKVVSLLKEVNFYSGILL LACISVD RYLAIVHATRRTL-QKRHLVKFVCLG CWGLSMNLSLPPFLFRQA----YHPNN--
A2a       CHGCLF IACFVLVLTASSIFSL LAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWVLSFA IGLTPMLGWNN----CGQPKE-
CXCR4     NFLCKAVHV IYTVNLYSSVWI LAFISLDRYLAIVHATNSQRPRKLLAEKVYVGVWIPALLLTI PDFIFANV----SEADD--
mOR256-3  YGGCVAQLYIFLALGSTECIL LAIMAFDRFAAICRPLHYPIIMNQKRCIHMATGTWISGFANSLVQSTL-TVVAPRCGQR---
mOR256-8  YAGC VVQLF SFLSIGSVECI L LAVMAYDRYA AVCKPLHYMVI MHPQLCVRLMAVAWGVGLANAIMSPL-AMTLPRCGRR---

bRho      ---MQCS C GIDYYTPHE-----ETNNSFVIYMFVVFHFI I PL I V I FFCY GQLVFTVKEAAAQ-----QESATTQKAEKV
CXCR1     ---SSPVC Y EVLGNDT-----AKWRMVLRI LPHTFGFI V PL FVM L FCYGFTLRTL FKAH-----MGQKHRA
A2a       GK AHSQGC GEGQVACL F---EDVVP MNYMVFYFNFFACV L V P L L M L G V Y L R I F L A A R R Q L K Q M E S Q L P G E R A R S T L Q K E V H A
CXCR4     ---RYI C D R F Y P N D-----LWV V V F Q F Q H I M V G L I L P G I V I L S C Y C I I I S K L S H S-----K G H Q K R K A
mOR256-3  -VIDHFF C E V P A L L K L A--CTD T S V N E A E L N V L G A L L L L V P L S L I L G T Y V F I A Q A V L K L R S A-----E S R R K A
mOR256-8  -RINHFL C E L P A L I K M A--CVDAR P V E M L S F T L A I L I V L L P L T L I L V S Y G Y I A A A V L R I K S A-----A G R W K A

bRho      TRMVI I M V I A F L I C W L P Y A G V A F Y I F T H O G S D F-----G P I F M T I P A F F A K T S A V Y N P V I Y I M N K Q F R N C M V T T L C C-
CXCR1     MRV I F A V V L I F L C W L P Y N L V L L A D T L M R T Q V I Q E S C E R R N N I G R A L D A T E I L G F L H S C L N P I Y A F I G Q N F R H G F L K I L A M H
A2a       A K S L A I I V G L F A L C W L P L H I I N C F T F F C P D C S H A-----P L W L M Y L A I V L S H T N S V V N P F I Y A Y R I R E F R Q T F R K I I R S H
CXCR4     L K T T V I L I L A F F A C W L P Y Y I G I S I D S F I L L E I I K Q G C E F E N T V H K W I S I T E A L A F F H C C L N P I L Y A F L G A K F K T S A Q H A L T S G
mOR256-3  F N T C A S H L L V V S L F Y F T A I S M Y V Q P P S S Y S-----H E R G K I M A L F Y G I V T P T L N P F I Y T L R N K D V K A A L R R A L T K E
mOR256-8  F N T C S S H L T V V S L F Y G S I I Y M Y M Q P G N S S S-----Q D Q G K F L T L F Y N L V T P M L N P L I Y T L R N K E M K G A L K K V C G R H

```

Figure S6: Sequence alignment for the homology modeling of mOR356-3 and mOR256-8.

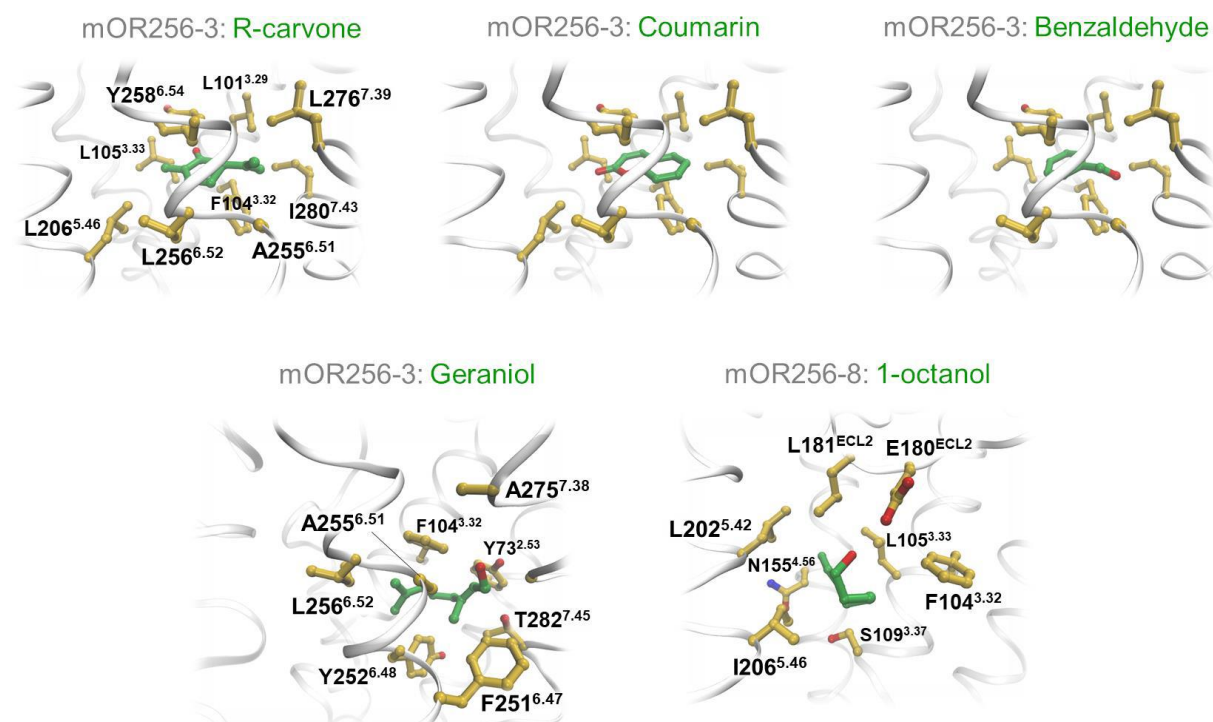


Figure S7: Predicted ligand interactions with mOR256-3 and mOR256-8.

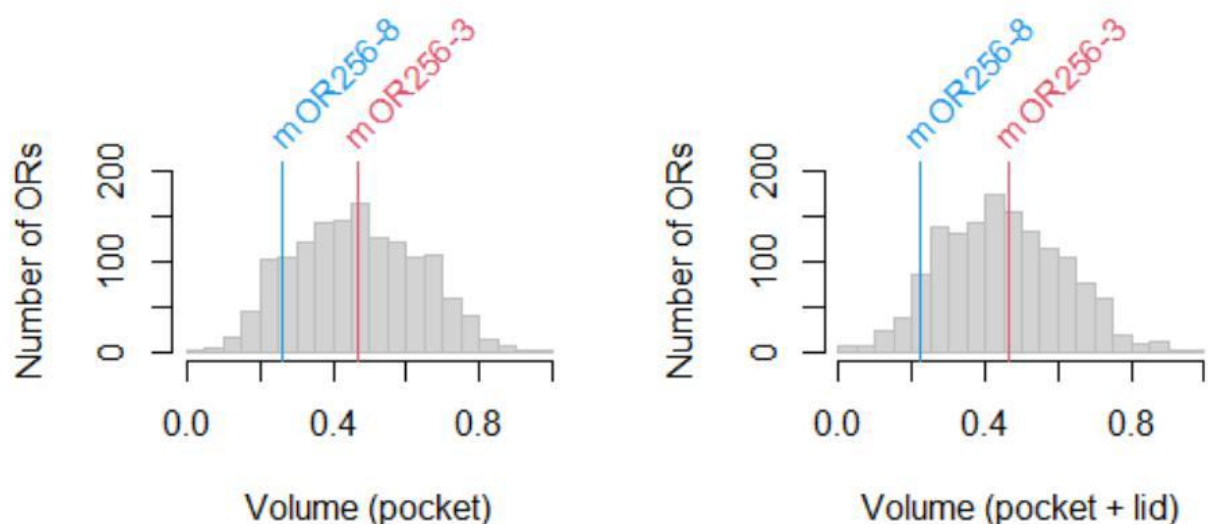


Figure S8: Histogram of normalized pocket volume of human and mouse ORs. The pocket volume was calculated from the sum of the side-chain volume of the residues forming the pocket and the lid. We used the 17 pocket residues identified in our previous work [34], in addition to the lid residues chosen according to the homology model in this work. The same residues were used for all the ORs according to the sequence alignment. It is a coarse estimation assuming similar shape and side-chain orientations in the pocket, without considering the 3D stacking of the residues.

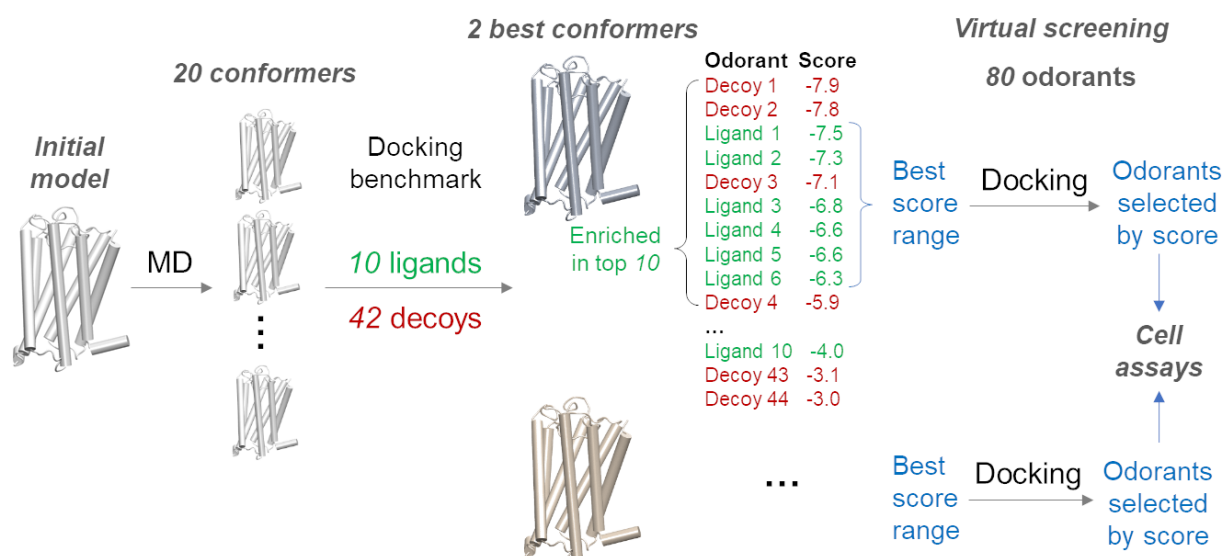


Figure S9: Virtual screening protocol. MD simulations were first performed on the initial model to obtain 20 conformers. Benchmark compounds were docked to each of the conformers and ranked by docking scores for the given conformer. The best conformers were chosen as those that returned the most ligands in the 10 top-ranked compounds. The range of scores that best separated the ligands from the decoys was used in the subsequent virtual screening to select hits. We used 2 best conformers for virtual screening and the common hits were tested in cell assays.

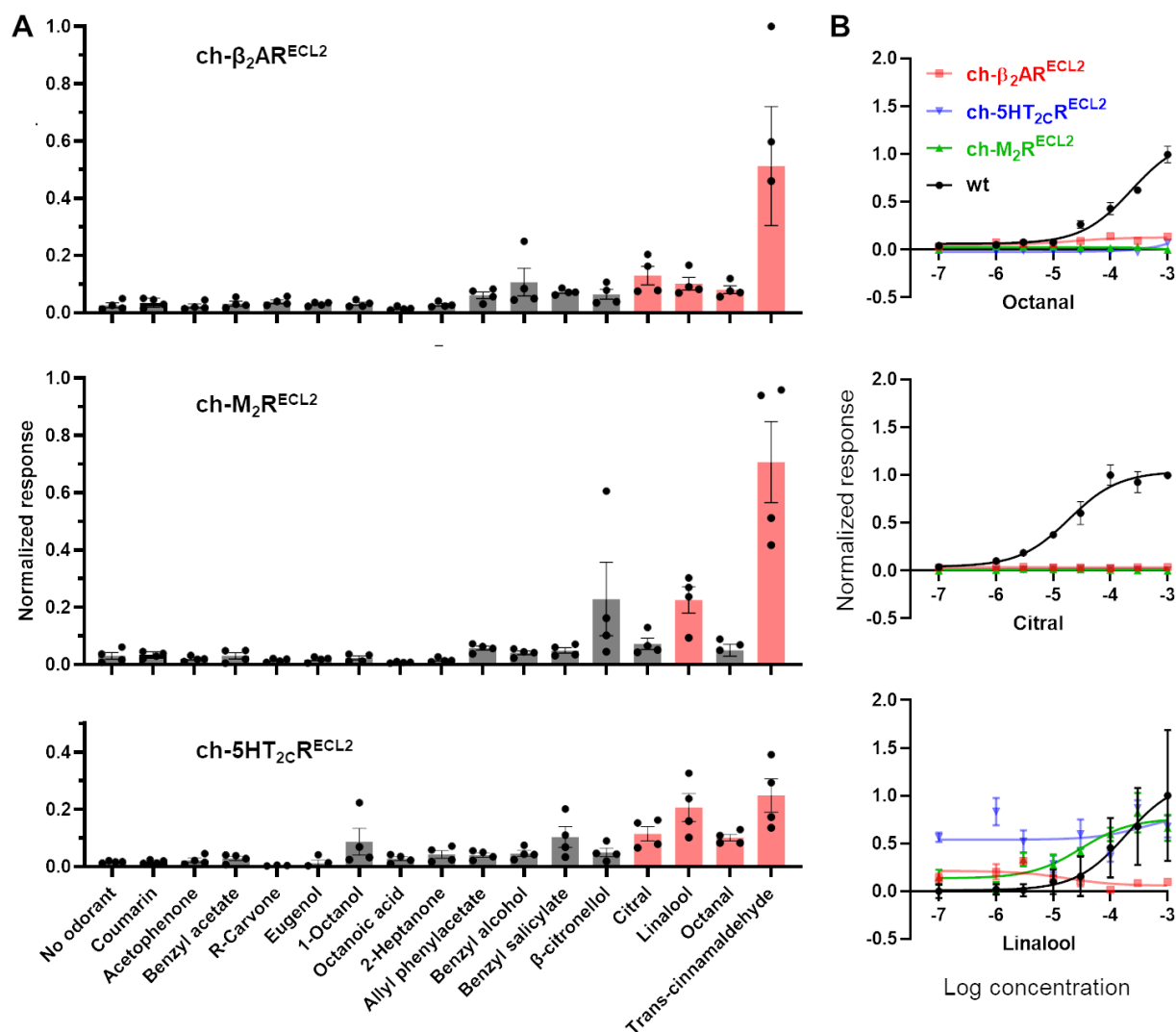


Figure S10: Functional assays of mOR256-3 chimeras. (A) Screening of 16 odorants at 300 μ M concentration. Significant responses are colored in red, which were tested in dose-dependent assays in (B). Data are mean \pm SEM of 3-4 technical repeats.

References

1. Malnic B, Godfrey PA, Buck LB (2004) The human olfactory receptor gene family. *Proc Natl Acad Sci U S A* 101:2584–2589. <https://doi.org/10.1073/pnas.0307882100>
2. Bushdid C, Magnasco MO, Vosshall LB, Keller A (2014) Humans Can Discriminate More than 1 Trillion Olfactory Stimuli. *Science* 343:1370–1372. <https://doi.org/10.1126/science.1249168>
3. Ferrer I, Garcia-Esparcia P, Carmona M, et al (2016) Olfactory Receptors in Non-Chemosensory Organs: The Nervous System in Health and Disease. *Frontiers in Aging Neuroscience* 8:

4. Kang N, Koo J (2012) Olfactory receptors in non-chemosensory tissues. *BMB Reports* 45:612–622. <https://doi.org/10.5483/BMBRep.2012.45.11.232>
5. Kang N, Kim H, Jae Y, et al (2015) Olfactory Marker Protein Expression Is an Indicator of Olfactory Receptor-Associated Events in Non-Olfactory Tissues. *PLOS ONE* 10:e0116097. <https://doi.org/10.1371/journal.pone.0116097>
6. Lee S-J, Depoortere I, Hatt H (2019) Therapeutic potential of ectopic olfactory and taste receptors. *Nat Rev Drug Discov* 18:116–138. <https://doi.org/10.1038/s41573-018-0002-3>
7. Li E, Shan H, Chen L, et al (2019) OLFR734 Mediates Glucose Metabolism as a Receptor of Asprosin. *Cell Metab* 30:319–328.e8. <https://doi.org/10.1016/j.cmet.2019.05.022>
8. Shepard BD, Koepsell H, Pluznick JL (2019) Renal olfactory receptor 1393 contributes to the progression of type 2 diabetes in a diet-induced obesity model. *Am J Physiol Renal Physiol* 316:F372–F381. <https://doi.org/10.1152/ajprenal.00069.2018>
9. Venkatakrisnan AJ, Deupi X, Lebon G, et al (2013) Molecular signatures of G-protein-coupled receptors. *Nature* 494:185–194. <https://doi.org/10.1038/nature11896>
10. Woolley MJ, Conner AC (2017) Understanding the common themes and diverse roles of the second extracellular loop (ECL2) of the GPCR super-family. *Mol Cell Endocrinol* 449:3–11. <https://doi.org/10.1016/j.mce.2016.11.023>
11. Wheatley M, Wootten D, Conner MT, et al (2012) Lifting the lid on GPCRs: the role of extracellular loops. *Br J Pharmacol* 165:1688–1703. <https://doi.org/10.1111/j.1476-5381.2011.01629.x>
12. Wolf S, Grünewald S (2015) Sequence, Structure and Ligand Binding Evolution of Rhodopsin-Like G Protein-Coupled Receptors: A Crystal Structure-Based Phylogenetic Analysis. *PLOS ONE* 10:e0123533. <https://doi.org/10.1371/journal.pone.0123533>
13. Karnik SS, Gogonea C, Patil S, et al (2003) Activation of G-protein-coupled receptors: a common molecular mechanism. *Trends Endocrinol Metab* 14:431–437. <https://doi.org/10.1016/j.tem.2003.09.007>
14. Palczewski K, Kumasaka T, Hori T, et al (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289:739–745. <https://doi.org/10.1126/science.289.5480.739>
15. Dror RO, Pan AC, Arlow DH, et al (2011) Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci U S A* 108:13118–13123. <https://doi.org/10.1073/pnas.1104614108>
16. Yuan S, Chan HCS, Vogel H, et al (2016) The Molecular Mechanism of P2Y1 Receptor Activation. *Angew Chem Int Ed Engl* 55:10331–10335. <https://doi.org/10.1002/anie.201605147>
17. Sandal M, Behrens M, Brockhoff A, et al (2015) Evidence for a Transient Additional Ligand Binding Site in the TAS2R46 Bitter Taste Receptor. *J Chem Theory Comput* 11:4439–4449. <https://doi.org/10.1021/acs.jctc.5b00472>

18. Sharma K, Balfanz S, Baumann A, Korsching S (2018) Full rescue of an inactive olfactory receptor mutant by elimination of an allosteric ligand-gating site. *Sci Rep* 8:9631. <https://doi.org/10.1038/s41598-018-27790-7>
19. Bushdid C, de March CA, Topin J, et al (2019) Mammalian class I odorant receptors exhibit a conserved vestibular-binding pocket. *Cell Mol Life Sci* 76:995–1004. <https://doi.org/10.1007/s00018-018-2996-4>
20. Bock A, Merten N, Schrage R, et al (2012) The allosteric vestibule of a seven transmembrane helical receptor controls G-protein coupling. *Nat Commun* 3:1044. <https://doi.org/10.1038/ncomms2028>
21. Thal DM, Glukhova A, Sexton PM, Christopoulos A (2018) Structural insights into G-protein-coupled receptor allostery. *Nature* 559:45–53. <https://doi.org/10.1038/s41586-018-0259-z>
22. Xu J, Hu Y, Kaindl J, et al (2019) Conformational Complexity and Dynamics in a Muscarinic Receptor Revealed by NMR Spectroscopy. *Mol Cell* 75:53-65.e7. <https://doi.org/10.1016/j.molcel.2019.04.028>
23. Kruse AC, Ring AM, Manglik A, et al (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* 504:101–106. <https://doi.org/10.1038/nature12735>
24. Yuan S, Dahoun T, Brugarolas M, et al (2019) Computational modeling of the olfactory receptor Olfr73 suggests a molecular basis for low potency of olfactory receptor-activating compounds. *Commun Biol* 2:1–10. <https://doi.org/10.1038/s42003-019-0384-8>
25. Gelis L, Wolf S, Hatt H, et al (2012) Prediction of a ligand-binding niche within a human olfactory receptor by combining site-directed mutagenesis with dynamic homology modeling. *Angew Chem Int Ed Engl* 51:1274–1278. <https://doi.org/10.1002/anie.201103980>
26. Geithe C, Protze J, Kreuchwig F, et al (2017) Structural determinants of a conserved enantiomer-selective carvone binding pocket in the human odorant receptor OR1A1. *Cell Mol Life Sci* 74:4209–4229. <https://doi.org/10.1007/s00018-017-2576-z>
27. Abaffy T, Malhotra A, Luetje CW (2007) The molecular basis for ligand specificity in a mouse olfactory receptor: a network of functionally important residues. *J Biol Chem* 282:1216–1224. <https://doi.org/10.1074/jbc.M609355200>
28. Baud O, Etter S, Spreafico M, et al (2011) The mouse eugenol odorant receptor: structural and functional plasticity of a broadly tuned odorant binding pocket. *Biochemistry* 50:843–853. <https://doi.org/10.1021/bi1017396>
29. Yu Y, de March CA, Ni MJ, et al (2015) Responsiveness of G protein-coupled odorant receptors is partially attributed to the activation mechanism. *Proceedings of the National Academy of Sciences* 112:14966–14971. <https://doi.org/10.1073/pnas.1517510112>
30. Cook BL, Steuerwald D, Kaiser L, et al (2009) Large-scale production and study of a synthetic G protein-coupled receptor: human olfactory receptor 17-4. *Proc Natl Acad Sci U S A* 106:11925–11930. <https://doi.org/10.1073/pnas.0811089106>

31. de March CA, Kim S-K, Antonczak S, et al (2015) G protein-coupled odorant receptors: From sequence to structure. *Protein Science : A Publication of the Protein Society* 24:1543–1548. <https://doi.org/10.1002/pro.2717>
32. de March CA, Yu Y, Ni MJ, et al (2015) Conserved Residues Control Activation of Mammalian G Protein-Coupled Odorant Receptors. *J Am Chem Soc* 137:8611–8616. <https://doi.org/10.1021/jacs.5b04659>
33. Charlier L, Topin J, Ronin C, et al (2012) How broadly tuned olfactory receptors equally recognize their agonists. Human OR1G1 as a test case. *Cell Mol Life Sci* 69:4205–4213. <https://doi.org/10.1007/s00018-012-1116-0>
34. Cong X, Ren W, Pacalon J, et al (2022) Large-Scale G Protein-Coupled Olfactory Receptor-Ligand Pairing. *ACS Cent Sci* 8:379–387. <https://doi.org/10.1021/acscentsci.1c01495>
35. Jumper J, Evans R, Pritzel A, et al (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
36. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815. <https://doi.org/10.1006/jmbi.1993.1626>
37. Waterhouse A, Bertoni M, Bienert S, et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46:W296–W303. <https://doi.org/10.1093/nar/gky427>
38. Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31:455–461. <https://doi.org/10.1002/jcc.21334>
39. Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure* 405:442–451. [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9)
40. Katada S, Hirokawa T, Oka Y, et al (2005) Structural Basis for a Broad But Selective Ligand Spectrum of a Mouse Olfactory Receptor: Mapping the Odorant-Binding Site. *J Neurosci* 25:1806–1815. <https://doi.org/10.1523/JNEUROSCI.4723-04.2005>
41. Reddy G, Zak JD, Vergassola M, Murthy VN (2018) Antagonism in olfactory receptor neurons and its implications for the perception of odor mixtures. *eLife* 7:e34958. <https://doi.org/10.7554/eLife.34958>
42. Wurch T, Lestienne F, Pauwels PJ (1998) A modified overlap extension PCR method to create chimeric genes in the absence of restriction enzymes. *Biotechnology Techniques* 12:653–657. <https://doi.org/10.1023/A:1008848517221>
43. Zhuang H, Matsunami H (2008) Evaluating cell-surface expression and measuring activation of mammalian odorant receptors in heterologous cells. *Nat Protoc* 3:1402–1413. <https://doi.org/10.1038/nprot.2008.120>
44. Shen M-Y, Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15:2507–2524. <https://doi.org/10.1110/ps.062416606>

45. Mirdita M, Steinegger M, Söding J (2019) MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* 35:2856–2858. <https://doi.org/10.1093/bioinformatics/bty1057>
46. Vainio MJ, Johnson MS (2007) Generating conformer ensembles using a multiobjective genetic algorithm. *J Chem Inf Model* 47:2462–2474. <https://doi.org/10.1021/ci6005646>
47. Forli S, Huey R, Pique ME, et al (2016) Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* 11:905–919. <https://doi.org/10.1038/nprot.2016.051>
48. Gordon JC, Myers JB, Folta T, et al (2005) H⁺⁺: a server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic Acids Res* 33:W368–371. <https://doi.org/10.1093/nar/gki464>
49. Schott-Verdugo S, Gohlke H (2019) PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. *J Chem Inf Model* 59:2522–2528. <https://doi.org/10.1021/acs.jcim.9b00269>
50. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* 21:1049–1074. [https://doi.org/10.1002/1096-987X\(200009\)21:12<1049::AID-JCC3>3.0.CO;2-F](https://doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F)
51. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A., et al., (2009) Gaussian 09, Revision A.02. Wallingford CT,
52. Lindorff-Larsen K, Piana S, Palmo K, et al (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78:1950–1958. <https://doi.org/10.1002/prot.22711>
53. Dickson CJ, Madej BD, Skjevik ÅA, et al (2014) Lipid14: The Amber Lipid Force Field. *J Chem Theory Comput* 10:865–879. <https://doi.org/10.1021/ct4010307>
54. Wang J, Wolf RM, Caldwell JW, et al (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174. <https://doi.org/10.1002/jcc.20035>
55. Jorgensen WL, Chandrasekhar J, Madura JD, et al (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935. <https://doi.org/10.1063/1.445869>
56. Joung IS, Cheatham TE (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B* 112:9020–9041. <https://doi.org/10.1021/jp8001614>
57. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* 23:327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5)

58. Darden T, Perera L, Li L, Pedersen L (1999) New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* 7:R55–R60. [https://doi.org/10.1016/S0969-2126\(99\)80033-1](https://doi.org/10.1016/S0969-2126(99)80033-1)
59. Wang L, Friesner RA, Berne BJ (2011) Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J Phys Chem B* 115:9431–9438. <https://doi.org/10.1021/jp204407d>
60. Van Der Spoel D, Lindahl E, Hess B, et al (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701–1718. <https://doi.org/10.1002/jcc.20291>
61. Tribello GA, Bonomi M, Branduardi D, et al (2014) PLUMED 2: New feathers for an old bird. *Computer Physics Communications* 185:604–613. <https://doi.org/10.1016/j.cpc.2013.09.018>
62. Patriksson A, Spoel D van der (2008) A temperature predictor for parallel tempering simulations. *Phys Chem Chem Phys* 10:2073–2077. <https://doi.org/10.1039/B716554D>
63. Daura X, Gademann K, Jaun B, et al (1999) Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition* 38:236–240. [https://doi.org/10.1002/\(SICI\)1521-3773\(19990115\)38:1/2<236::AID-ANIE236>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M)
64. Kida H, Fukutani Y, Mainland JD, et al (2018) Vapor detection and discrimination with a panel of odorant receptors. *Nat Commun* 9:4556. <https://doi.org/10.1038/s41467-018-06806-w>

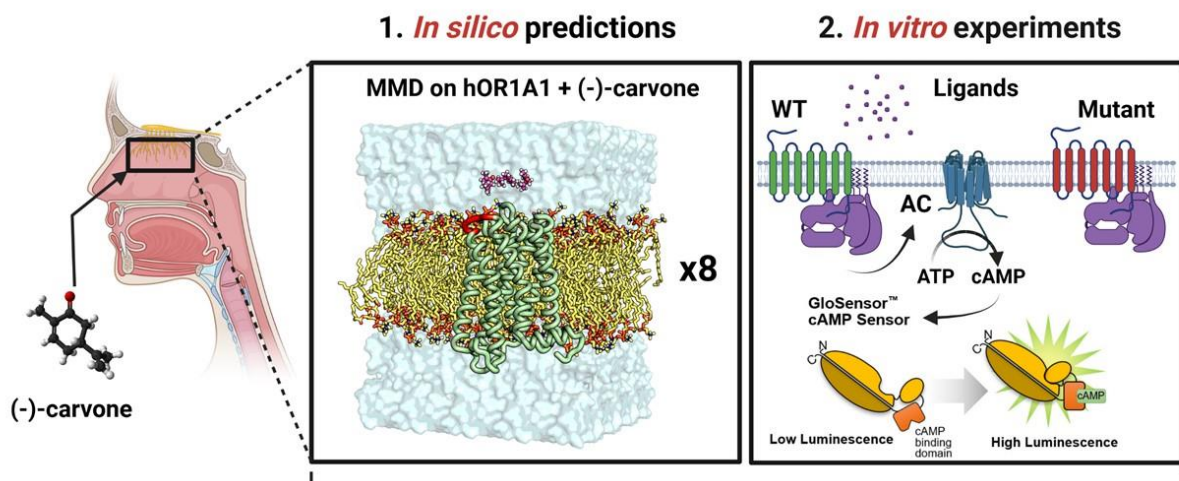
Publication 4

The third extracellular loop of mammalian odorant receptors is involved in ligand binding.

Shim T, Pacalon J, Kim WC, Cong X, Topin J, Golebiowski J, Moon C. 2022, IJMS, under review

Abstract

Mammals recognize chemicals in the air via G protein-coupled odorant receptors (ORs). In addition to their orthosteric binding site, other segments of these receptors modulate ligand recognition. Focusing on human hOR1A1, considered prototypical of class II ORs, we used a combination of molecular modeling, site-directed mutagenesis, and *in vitro* functional assays. We showed that the third extracellular loop of ORs (ECL3) contributes to ligand recognition and receptor activation. Site-directed mutations in ECL3 indeed showed differential effects on the potency and efficacy of both carvones, citronellol, and 2-nonanone.



ECL3 in classII ORs:
1. Form a **PxSxxS** motif.
2. Implicated in **vestibular binding site**
(agonist entry & selection process)

Keywords

Odorant receptors; ECL3; Ligand selectivity; Molecular modeling; Functional assays

Introduction

Mammals rely on their sense of smell to assess the volatile chemical environment. Smell information is decoded by chemical interactions between odorants and G protein-coupled odorant receptors (ORs) expressed in olfactory neurons present in the nasal cavity. While these transmembrane proteins are the cornerstone of chemical recognition by our neurons, perireceptor events can modify the chemical composition of odorants before reaching the Ors [1]. For example, the enzyme carboxyl esterase has been shown to influence odor recognition by altering chemical function [2]. Odorant Binding Proteins also modulate the chemical signal by helping solubilize odorants or playing the role of scavengers [3].

The coding of a perceived odor by the olfactory system nonetheless relies on a combinatorial code where ORs are differentially activated by odorants and where one odorant can activate multiple ORs [4,5]. The subtle interactions between ORs and odorants at the molecular level remain extremely difficult to rationalize. In fact, very subtle modifications in the chemical structure of a molecule can drastically alter its odor [6,7]. Similarly, mutations within an OR gene might strongly alter smell perception [8].

ORs represent more than 3% of the whole proteome and belong to the class A G protein-coupled receptor (GPCR) family of proteins, notably responsible for transmitting signals through the cell membrane. The large number of ORs (~400 in humans and ~1100 in mice, for example) coupled with this combinatorial activation endows mammals with incredible discriminatory power [9,10]. GPCRs are one of the largest and most diverse membrane protein families. They adopt a typical architecture consisting of seven transmembrane helices (TM1 to TM7) linked by intracellular and extracellular loops (ICLs and ECLs, respectively). ECLs, while peripheral within the GPCR tertiary structure, are involved in numerous receptor functions such as ligand or protein recognition and receptor activation [11-14]. In most class A GPCRs, the internal binding site for the ligand is found at ca. 10 Å with respect to the extracellular side of the receptor. In the prototypical beta2-adrenergic receptor, multiple interactions have been described between the ligand and ECL3 during the process of ligand migration from the bulk solvent to the internal binding site [15]. In particular, ECL3 was suggested as a functional region important for ligand specificity. In a sub-class of mammalian ORs, namely the class I ORs, we showed that the extracellular part of the receptor, notably ECL3, was playing the role of a vestibular binding site [16]. The sequence variability at the ECL3 is high between the two

OR classes, which prevents concluding on the role of ECL3 in class II based on results from class I. This function of ECL3 in the class II OR has not been investigated until now.

In this article, we report on the role of ECL3 in class II ORs and show that ECL3 modulates ligand binding independent of the sequence variability between the two classes. We consider hOR1A1 as the prototypical class II OR. We thus conclude that ECL3 is involved in odorant selectivity in all mammalian ORs.

Results

The highly variable ECL3 sequence acts as a vestibular binding pocket for ligands

In class A GPCRs, ECL3 connects the extracellular parts of TM6 and TM7. Up to now, no structure of hOR1A1 has been experimentally solved. A 3D modeled structure was built using a previously established protocol [17,18]. An alternative model was obtained using AlphaFold [19]. Figure 1a depicts the two modeled 3D structures of hOR1A1. The overall structure of the two models is conserved with an rmsd of 3.5 Å on the backbone. In general, the main difference occurs in the unstructured segment of the receptor. In both models, ECL3 was located ca. 4Å toward the extracellular side of the orthosteric binding cavity. In hOR1A1, ECL3 comprised five/six residues (P261/L262 to S266) and showed an unfolded 3D structure. The vicinity of ECL3 and the binding cavity was consistent with a potential role of ECL3 as a vestibular binding site as already observed by us on class I ORs and by others on the beta2-adrenergic receptor and muscarinic M2 receptors [15,16,20].

From a sequence point of view, the conservation analysis and its decomposition into class I and class II ORs revealed that the positions in ECL3 were conserved in class II ORs and formed a P261xSxxS motif (Figure 1b). In human ORs, the proline (here P261) was conserved at 72.3%, S263 at 46.5%, and S266 at 44.1%. In hOR1A1, the sequence reads PLTNYS. They were not only different from those found in class I but also different from any other class A GPCRs.

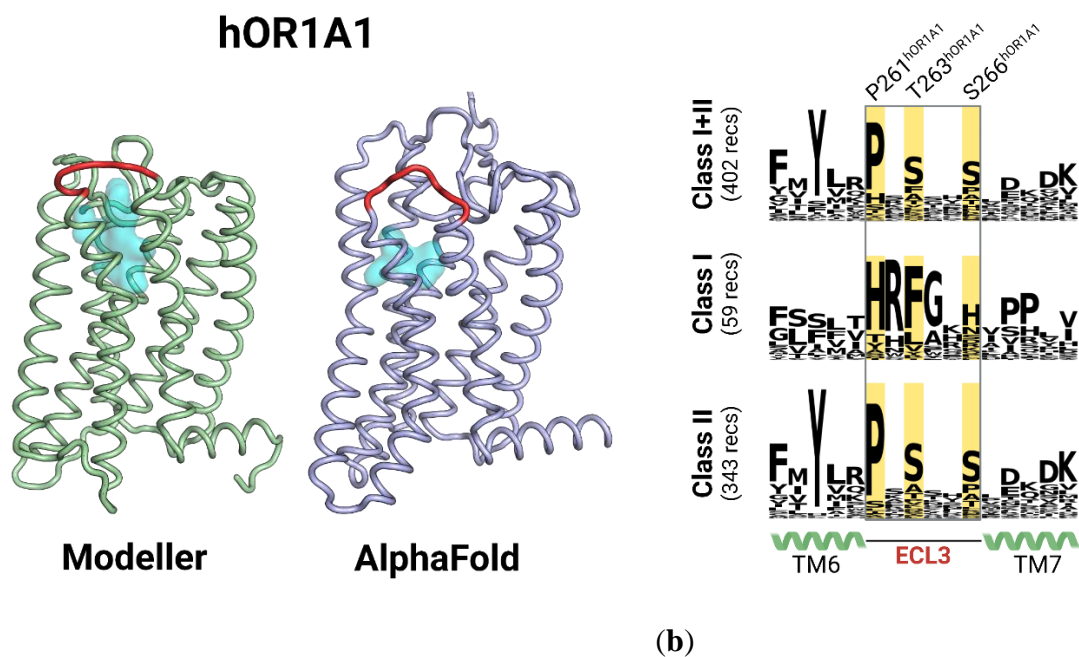


Figure 1: (a) Structure of hOR1A1 from homology modeling (Modeller) compared to that obtained from AlphaFold. In both structures, ECL3 (shown in red) was predicted to be close to the orthosteric binding cavity, shown as a cyan surface. (b) Conservation analysis of ECL3 sequences of both classes of human odorant receptors and the highlight of hOR1A1 specific ECL3 sequence.

The involvement of ECL3 in the binding process was assessed by a molecular dynamics (MD) simulation where (-)-carvone, a strong ligand for hOR1A1[21,22], was initially located within the bulk solvent. When all replicas were aggregated, the binding mechanism could be decomposed into a three-step process. Figure 2 depicts the main events identified during the trajectories and highlights the density of (-)-carvone in the protein with respect to both residue Y251^{6,48} (bottom of the orthosteric cavity) and S266^{ECL3}. Starting from the bulk solvent, the ligand rapidly approached the extracellular segment of the receptor and initiated contact with ECL3 before reaching the orthosteric binding site. The density map confirmed the three regions in which the ligand spent more time. Upon the binding process, the ligand interacted with various residues in ECL3, highlighting this loop involvement in the ligand's entry within the receptor.

In our homology model, P261^{ECL3} is the first residue of ECL3 and acts as a (-)-carvone contact point with the receptor after ca. 100 nanoseconds. The ligand then interacted with the hydrophobic residues Leu262^{ECL3} and diffused towards the orthosteric binding site, engaging H-bond with Ser266^{ECL3} (Figure 2, position 2). Finally, the ligand interacted with Asp269^{7,34} before entering the binding cavity (Figure 2 left, position 3), i.e., at less than 2.5 Å of Y251^{6,48}

(6.48 refers to the Ballesteros-Weinstein notation) [23,24]. To assess the role of amino acids from ECL3 in the receptor recognition process, *in vitro* experiments were further performed.

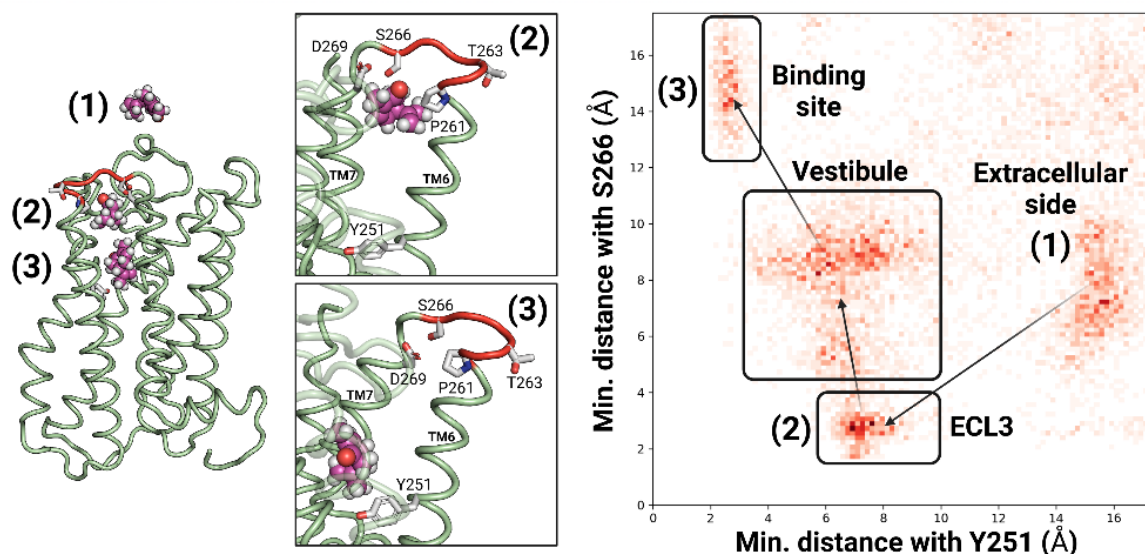


Figure 2: Entry of (-)-carvone inside receptor hOR1A1. The ligand is initially located outside the receptor (1). It then migrates to the cradle of the orthosteric binding cavity (2, 3), as indicated by Y251^{6,48}. During this process, the ligand interacts with several residues from ECL3 (indicated in red). Contour map of (-)-carvone migration as the minimum distance from S266 (taken as the distance from ECL3) and minimum distance from Y251^{6,48} (taken as the distance from the cradle of the cavity). All replicas were considered. The three highlighted basins show the ligand's largest density.

In vitro functional assays highlight the differential interactions with diverse ligands

Based on the results of both the conservation analysis and the MD simulations, several amino acids from ECL3 seem to be involved in the binding process. To investigate the specific chemical functions of amino acids in ECL3 upon ligand binding, we designed mutant ORs with various (small, charged, lipophilic, or aromatic) properties, and assessed their response to four different ligands (Figure 3).

All the mutant ORs investigated in this study were confirmed to be expressed at the membrane surface and showed some basal activities, confirming that they remain functional (Figure 4b-c and SI). However, it appeared that the expression level of the mutant ORs was differentially affected compared to the *wt* OR (Figure 4b-c). The presence of a phenylalanine residue

systematically resulted in a decrease in surface expression for all four mutant ORs. Mutation to an alanine residue at all three hydrophilic positions (T263^{ECL3}, S266^{ECL3} and D269^{7.34}) also decreased surface expression of the receptor, whereas this mutation had no effect on expression for the P261A mutant OR. Surprisingly, the presence of a positive charge at positions P261^{ECL3}, T263^{ECL3}, and S266^{ECL3} did not affect surface expression. The same modification at position D269^{7.34} induced a decrease in surface expression. Overall, we observed that bulky amino acids in ECL3 resulted an increased basal activity (figure S3) independent of surface expression.

The *wt* OR exhibited dose-dependently increased cAMP levels in response to four ligands ((-)-carvone, (+)-carvone, citronellol, and 2-nonanone) known as bona fide agonists *in vitro* (Figure 3 and Figure 4a) [21,23,25]. Mutations at positions P261, T263, S266, and D269 differentially affected the receptor response to agonists (Figure 4a).

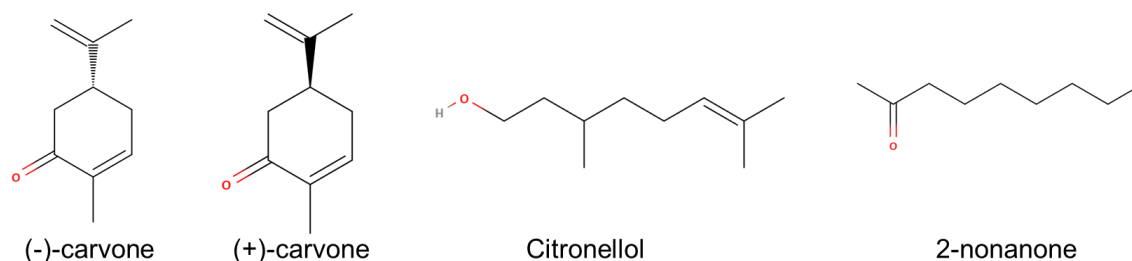


Figure 3: Chemical structure of four agonists of hOR1A1.

MD simulations revealed interactions between the (-)-carvone and P261^{ECL3} early in the binding process. P261X mutant ORs (X=A, R, L, or F) did not show a dose-dependent response for all four ligands: their efficacy and potency were significantly reduced (Figure 4a). Mutations affected the agonist-induced response differently: P261A and P261R remained sensitive to the most potent ligand, (-)-carvone, whereas the receptors' responses were nearly abolished for the other three ligands. Mutations to a more hydrophobic and bulkier residue, such as leucine or phenylalanine (Figure 4a pink and green, respectively), abolished the response to all four ligands. These data show that mutations at position P261 differentially affect the response to ligands and reinforce the hypothesis of the initial interaction between P261 and ligands during the binding process.

T263 and S266 are two small and polar conserved residues in the ECL3 PXSxxS motif, which strongly interacted with (-)-carvone during the MD simulations. Mutations at position 263 had notably been reported to affect OR2AG1 response to agonist [23]. Interestingly, mutations at

these positions had a different impact on the receptor response to the 4 agonists. The mutations into apolar (A), charged (R), or aromatic (F) residues were -like for P261- also associated with differential modification of the receptor response to ligands *in vitro*

The T263F mutation abolished the receptor response to all four ligands (Figure 4a, second row, green curves). The presence of positively charged residues systematically decreased the efficacy of all four ligands (Figure 4a, second row, blue curve). Finally, all four ligands were differentially affected by introducing of an alanine at position 263. The response to 2-nonanone and citronellol remained unchanged compared to *wt* OR, while the efficacy of the two enantiomers of carvone decreased. This observation suggested a specific interaction of residue T263 with both carvone enantiomers.

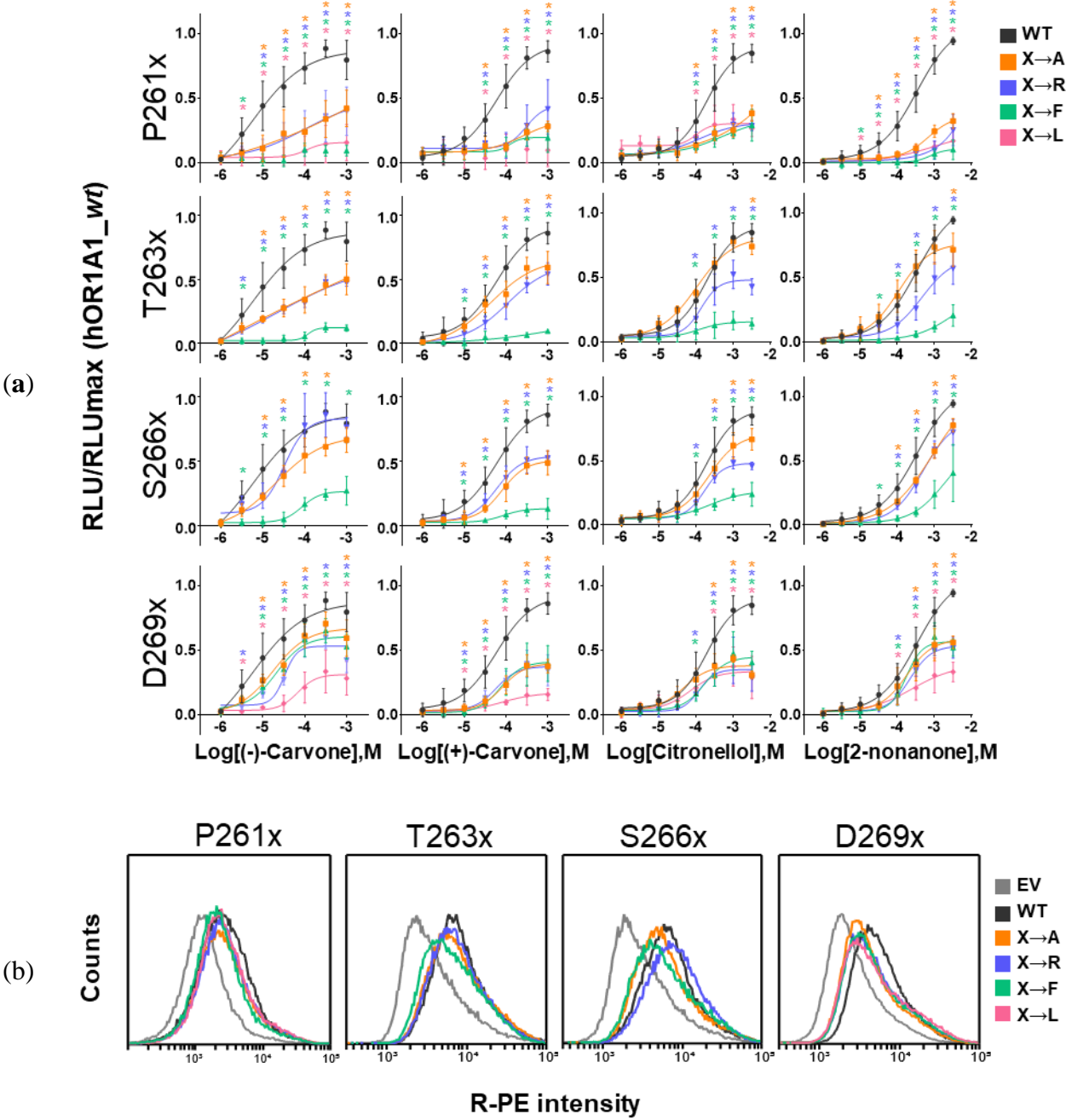
Mutations at position S266 also induced changes in efficacy. Surprisingly, (-)-carvone efficacy for the S266R mutant OR remained unchanged (Figure 4a, 3rd row, black curve vs. blue curve), whereas it was systematically reduced for the other ligands. The S266A and S266F mutant ORs decreased the efficacy of all four ligands. As for the other three residues investigated, introducing a phenylalanine residue almost abrogated the receptor response.

D269^{7,34} was not predicted to belong to ECL3 in both modeler and AlphaFold models. It was located at the upper extremity of TM7. One can observe in Figure 4c that surface expression of the mutant ORs was systematically decreased except for D269L, in line with the role of the extracellular part of TM7 in the activation mechanism [26]. Compared to the mutant ORs at ECL3, the impact of the mutation on receptor response to agonist stimulation was more systematic. The mutations to A, R, or F induced a similar decrease in the receptor response for the four studied ligands. The D269L mutant OR showed a much more pronounced decrease in response to agonist stimulation except for citronellol (Figure 4a, last line pink color). This systematic trend suggests that D269 seems involved in the dynamic of receptor activation rather than in ligand binding recognition.

All these experiments underline the importance of ECL3 residues during binding. It can be observed that the presence of a phenylalanine at the position studied systematically altered the expression level. Furthermore, the response of the phenylalanine mutant ORs to the four ligands was abrogated except for D269F which remained sensitive to agonist stimulation.

In general, the mutations did not alter the EC50. When a dose-response curve was observed, the EC50 was never shifted more than one order of magnitude in concentration. The mutations clearly affect the efficacy of the ligands. Finally, the major takeaway of our study is that, from

a general point of view, efficacies and potencies of ligands are differentially affected by the mutations in ECL3.



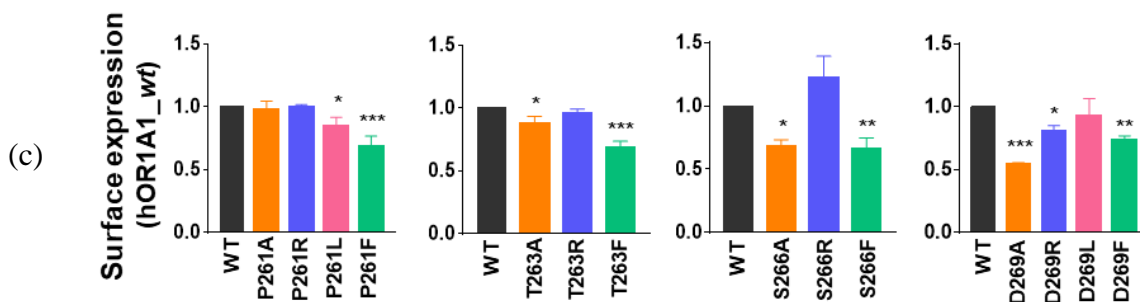


Figure 4: *In vitro* data of hOR1A1 and mutant ORs. (a) *In vitro* dose-response curves of four ligands (-)-carvone, (+)-carvone, citronellol, and 2-nonanone towards *wt* hOR1A1 and mutant ORs at positions P261, T263, S266, and D269. (*) indicates the response value is significantly different compared to *wt* hOR1A1 (one-way ANOVA, followed by a Dunnett test; * $p < 0.05$). (b) Flow cytometry analysis of cell-surface expression of *wt* hOR1A1 and mutant ORs at positions P261, T263, S266, and D269 (c) Normalized graph of cell-surface expression of ECL3 mutant ORs against *wt* hOR1A1 (one-way ANOVA, followed by a Dunnett test; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$).

Discussion

Odorant receptors, as all class A GPCRs, are structured with a seven-transmembrane domain where helices are connected by three extra- and three intra- cellular loops. Among these loops, the short ECL3 connects TM6 to TM7 and shows a large sequence variability. Several works have already evaluated the role of this very poorly conserved ECL3. They notably assessed its essential contribution to the ligand-binding process of the beta2-adrenergic receptor or the muscarinic M2 receptor [15,20]. The active role of ECL3 in ligand binding is consistent with the close vicinity of ECL3 and the orthosteric binding site. It is, for example, illustrated in some peptide-binding class A GPCRs, where the peptide directly interacts with both ECL3 and the orthosteric binding pocket [27,28]. Note that ECL3 structures in various peptide-bound class A GPCR are much more variable than ECL1 or ECL2, suggesting that ECL3 plays a modulation role rather than a ligand recognition role [29]. All in all, the role of ECL3 in class A GPCRs is not only structural but seems to modulate the ligand's recognition in combination with the binding site.

In class I ORs, a sub-family of mammalian ORs, we previously put forward that the ECL3 functions as a vestibule site and contributes to ligand binding [16]. In this work, we hypothesized a similar involvement of ECL3, although the conserved residues are significantly

different between the two sub-families of ORs. Through a joint approach combining molecular modeling, heterologous functional expression, site-directed mutagenesis, and functional assays on hOR1A1 -a prototypical class II OR- we have evaluated the role of ECL3 in ligand binding. In class II ORs, ECL3 is based on the conserved P²⁶¹xS²⁶³xxS²⁶⁶ motif. Various mutant ORs at positions 261, 263, and 266 showed differential responses to four agonists belonging to various chemical families compared to *wt*.

Meanwhile, additional mutant ORs at position D269^{7.34} (at the junction between ECL3 and TM7) showed a conserved modulation regardless of the type of substituted amino acid. Molecular dynamics simulations of a homology model of hOR1A1 in interaction with (-)-carvone sampled structures where the ligand is in regular contact with ECL3 prior to entering the orthosteric binding cavity.

hOR1A1 has been extensively studied for its property of discriminating the two enantiomers of carvone [23]. In a former study combining homology modeling, site-directed mutagenesis, and functional expression studies, the authors identified eleven positions on the transmembrane segments involved in the chiral recognition of carvone. Our results reveal however that the two enantiomers of carvone are similarly affected by mutations in ECL3. The conserved residues in ECL3, although involved in agonist recognition, are apparently not involved in chiral discrimination in hOR1A1.

Finally, our results suggest that ECL3 in hOR1A1 plays a similar role in ligand binding as observed in the β -2-adrenergic receptor, the M2 muscarinic receptor, and class I ORs. This study therefore reconciles all previous studies and demonstrates a similar function of ECL3 on ligand binding in all class A GPCRs, independent of the ECL3 sequence. The large variability in ECL3 sequence among class A GPCR subfamilies is consistent with the diverse chemical space associated with these receptors, which can bind highly polar or highly lipophilic ligands. It could become an interesting and alternative target for allosteric ligand design.

Materials and Methods

In silico experiments

Molecular modeling

hOR1A1 models were created using MODELLER 9.25 [30]. Our previous protocol was described in De March et al. 2015 with four experimental structures as templates (pdb codes: 1U19, 3ODU, 2YDV, and 2LNL) [18]. Cysteine residues 74 with 156, and 146 with 166 were assigned to form two disulfide bonds [31]. The best-generated models were selected based on the DOPE score, and one model was finally chosen based on visual inspection. hOR1A1 AlphaFold model was retrieved from the AlphaFold structure database [32,33].

Modeller and AlphaFold models have a 3.5 Å RMSD (calculated on the backbone of TMs). Both models nicely superimpose concerning TM6 and TM7. Distance between Y251^{6,48} and ECL3's amino acids is comparable between hOR1A1 Modeller and AlphaFold model (respectively 15.7/17.4 Å, 19.0/20.2 Å, 19.0/19.4 Å and 14.7/14.3 Å for residues P261, T263, S266 and D269 for Modeller/AlphaFold). While ECL3 is made up of almost identical residues, its structure is predicted with low confidence by AlphaFold (pLDDT between 50 and 70). The model built by homology was kept for the molecular dynamics simulations.

Molecular dynamics preparation

As the N-terminal of GPCR is not resolved in the template we used, the N-terminal part of the hOR1A1 Modeller model was discarded until residue E24. Propka3 was used to predict protonation states of the protein at a target pH 6.5 [34]. The extremities of the model were capped accordingly. hOR1A1 orientation in its membrane was determined using the OPM server [35]. Three (-)-carvone molecules were added in different orientations on the extracellular side at 5-10 Å of the top of the receptor. The system was embedded into a POPC-only model membrane using PACKMOL-memgen [36]. The simulation box was completed using TIP3P water molecules and neutralized using K⁺ and Cl⁻ ions with a final concentration of 0.15 M. The total system comprises 49536 atoms in a 7.10⁵ Å³ periodic box.

Molecular Dynamics protocol

Molecular dynamics simulations were performed with the sander and pmemd.cuda modules of AMBER18 [37], with the ff14SB force field for the proteins and the lipid14 forcefield for the membrane. (-)-carvone parameters were generated by calculating partial atomic charges with the HF/6-31G* basis set using Gaussian 09 [38]. The obtained electrostatic potential was fitted by the RESP program [39]. The other parameters were taken from the General Amber Force Field 2 (gaff2). Bonds involving hydrogen atoms were constrained using the SHAKE algorithm and long-range electrostatic interactions were handled using Particle Mesh Ewald. The cut-off for non-bonded interaction was set to 10 Å. Each system was first minimized with the AMBER sander module, with 5000 steps of steepest descent and then 5000 steps of the conjugate gradient with a 50 kcal·mol⁻¹·Å² harmonic potential restraint on the protein. A second minimization of the same length without restraint was applied. The systems were then thermalized from 100 to 310 K for 10000 steps (restraining the protein and ligands with a 200 kcal·mol⁻¹·Å² harmonic potential). Each system underwent 50000 steps of equilibration in the NPT ensemble and 1 bar (restraining the protein and ligands with a 15 kcal·mol⁻¹·Å² harmonic potential) before the production phase. During the equilibration and production phase, the temperature was kept constant in the system at 310 K using a Langevin thermostat with a collision frequency of 5 ps⁻¹. To increase sampling, all 3 (-)-carvone molecules were constrained in a sphere of 30-40 Å radius, centered on the center of mass of the Thr89 hOR1A1 (with a potential of 10 kcal·mol⁻¹·Å²).

The system stability was evaluated from the root mean square deviation (RMSD) evolution computed on the receptor backbone. For the 8 replicas, the receptors underwent small fluctuations (RMSD < 4 Å), showing that they remained correctly folded during microsecond simulations (Figure S1). Binding events occurred in Rep1 (stable) and Rep3 (partial entry) (Figure S2). Cpptraj and Pytraj v2.02.dev0 were used for distance and RMSD analysis [40].

In vitro experiments

Cell culture

HEK293T cell line was obtained from ATCC (#CRL-3216, ATCC, Manassas, VA, USA) and cultured in Dulbecco's modified Eagle's medium (DMEM; #10-017-CV, Corning, NY, USA)

plus 10% fetal bovine serum (#16000-044, Thermo Fisher Scientific, Waltham, MA, USA) and penicillin/streptomycin (#15140122, Thermo Fisher Scientific, Waltham, MA, USA) at 37°C with 5% CO₂.

DNA constructs, site-directed mutagenesis, and gene transfection

The full genomic DNA sequence of hOR1A1 was obtained from HEK293T cells using cloning primers (5'-GCA CGC GTA TGA GGG AAA ATA ACC AGT C-3' and 5'-GCG CGG CCG CTT ACG AGG AGA TTC TCT T-3'). PCR product was subcloned to LUCY-FLAG-Rho tagged pCI mammalian expression vector using MluI and NotI restriction enzyme. The pCI vector, which is the Rho sequence-tagged at N-terminal, RTP1s-pCI, G_{olf}-pCI, and Ric8b-pCI was a kind gift from Dr. Matsunami (Duke U., USA). A cleavable leucine-rich 17 amino acid signal peptide (LUCY; MRPQILLLLALLTLGLA) and FLAG sequences were tagged in front of the Rho sequence to promote or detect the functional expression on the plasma membrane, respectively. [41] pHIV-EGFP was a gift from Bryan Welm & Zena Werb (Addgene plasmid #21373). pGloSensorTM-22F cAMP Plasmid was purchased (#E2301, Promega, Madison, WI, USA). Site-directed mutant ORs were generated through the QuikChange PCR protocol [42] using the mutagenic oligonucleotide primers, designed individually according to the desired mutation. The sequence of plasmids was confirmed through BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, Waltham, MA, USA). Furthermore, JetPRIME (#114-75, PolyPlus-transfection, Illkirch-Graffenstaden, France) was used for gene transfection. For FACS experiments, the OR or mutant OR RTP1s, Ric8b, G_{olf}, and pHIV-EGFP plasmid were transfected, and for cAMP luminescence assay, the pGloSensorTM-22F cAMP plasmid was transfected instead of the pHIV-EGFP plasmid.

cAMP luminescence assay

The cAMP luminescence assay was performed using the GloSensorTM cAMP assay (Promega, Madison, WI, USA), followed by the manufacturer's protocol. In brief, transfected HEK293T cells on the 96-well white plate were exchanged its media to CO₂ independent media (#18045088, Thermo Fisher Scientific, Waltham, MA, USA) with GloSensorTM cAMP Reagent

(#E1291, Promega, Madison, WI, USA) before exposed to solvent or odorants. Each odorant was diluted into CO₂ independent media and used to treat each well. The endpoint of the luminescence level was measured with a SpectraMax L Microplate Reader (Molecular Devices, San Jose, CA, USA). Data were analyzed through Microsoft Excel and GraphPad Prism. To compare OR responses plate to plate, empty pCI vector and wild-type OR were always included as a control. To read the basal activity of each OR, at least six wells' values in the absence of odorants were averaged. The experiments were repeated twice to measure an odorant-induced OR activity, and each condition was triplicated. The measured luminescence value was further corrected by subtracting that for the lowest response value to each odorant of that receptor. The basal activity and odorant-induced responses were normalized to that of wild-type OR.

Functional expression

The functional expression of OR on the plasma membrane was measured through Fluorescent Activated Cell Sorting (FACS) as previously described [43]. To summarize, transfected HEK293T cells on the six-well plates were gently detached from the plate using Cellstripper (#25-056-CI, Thermo Fisher Scientific, Waltham, MA, USA). Incubate the cell with Rho4D2 antibody (#ab98887, Abcam, Cambridge, MA, USA) for an hour with gentle rotation. And after washing twice with washing & staining solution (2% fetal bovine serum (FBS) and 15M NaN₃ in 500 ml PBS), attaching R-PE-conjugated secondary antibody (#715-116-151, Jackson ImmunoResearch Inc., West Grove, PA, USA) for 30 min with gentle rotation. Lastly, the cells were washed twice and labeled the dead cells with 7-AAD (#SML1633, Merck, Darmstadt, Germany). The fluorescence of immunolabeled cells was detected and analyzed using BD AccuriTM C6 Plus Flow Cytometer (BD Biosciences, San Jose, CA, USA). After removing 7-AAD signal-positive dead cells, the intensity of the R-PE signal among the GFP-positive cells was measured and plotted.

Author Contributions

Conceptualization, J.T., J.G., C.M.; resources and funding acquisition: J.G. and C.M.; formal analysis: T.S., J.P., X.C., J.T., J.G., and C.M.; writing: J.T., J.G., and C.M. All authors have read and agreed to the published version of the manuscript.

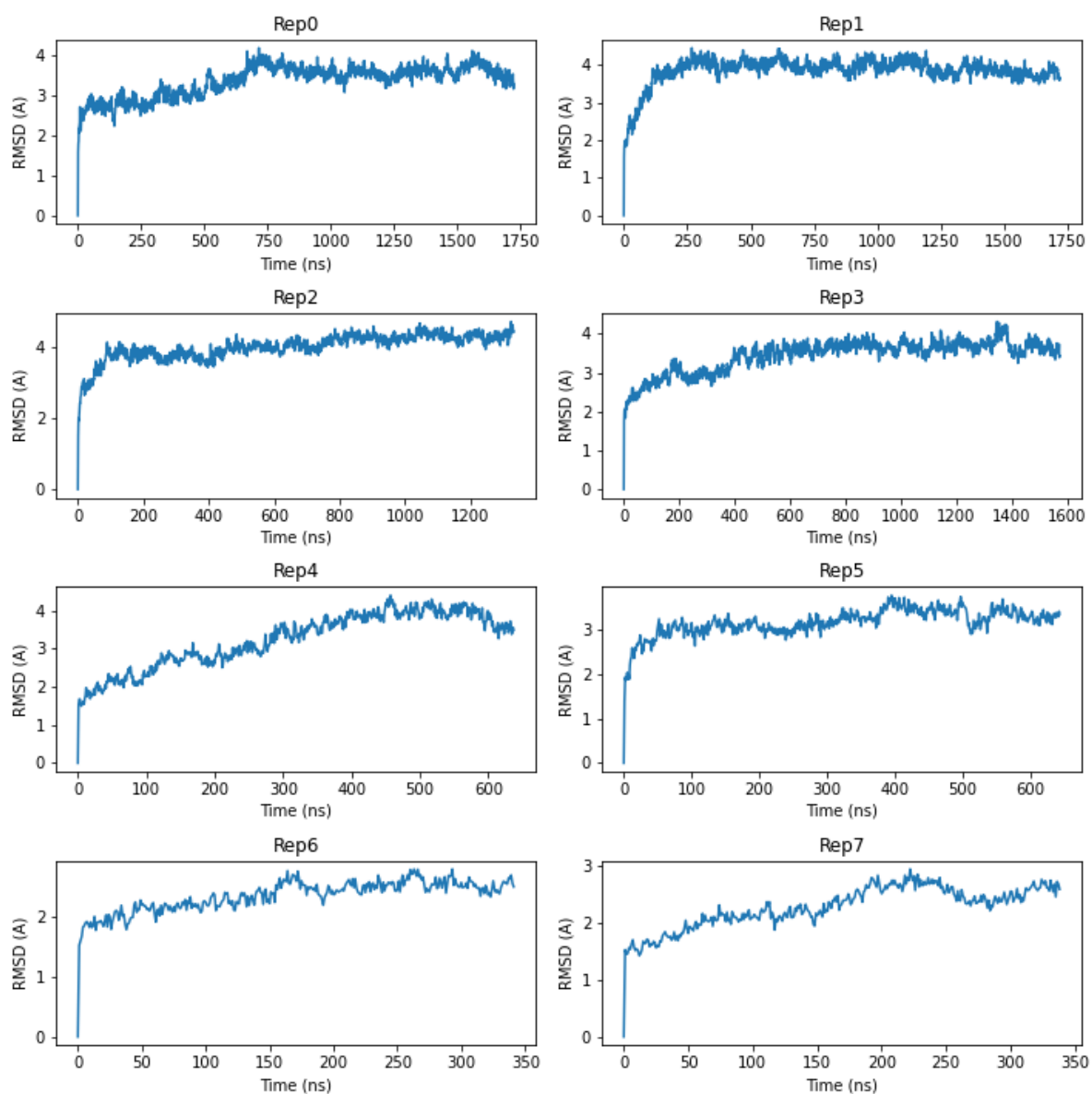
Funding

This work was supported by the Fondation Roudnitska under the aegis of Fondation de France to J.P and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03040516) to C.M. Calculations were carried on Azzurra HPC center (Université Côte D'Azur).

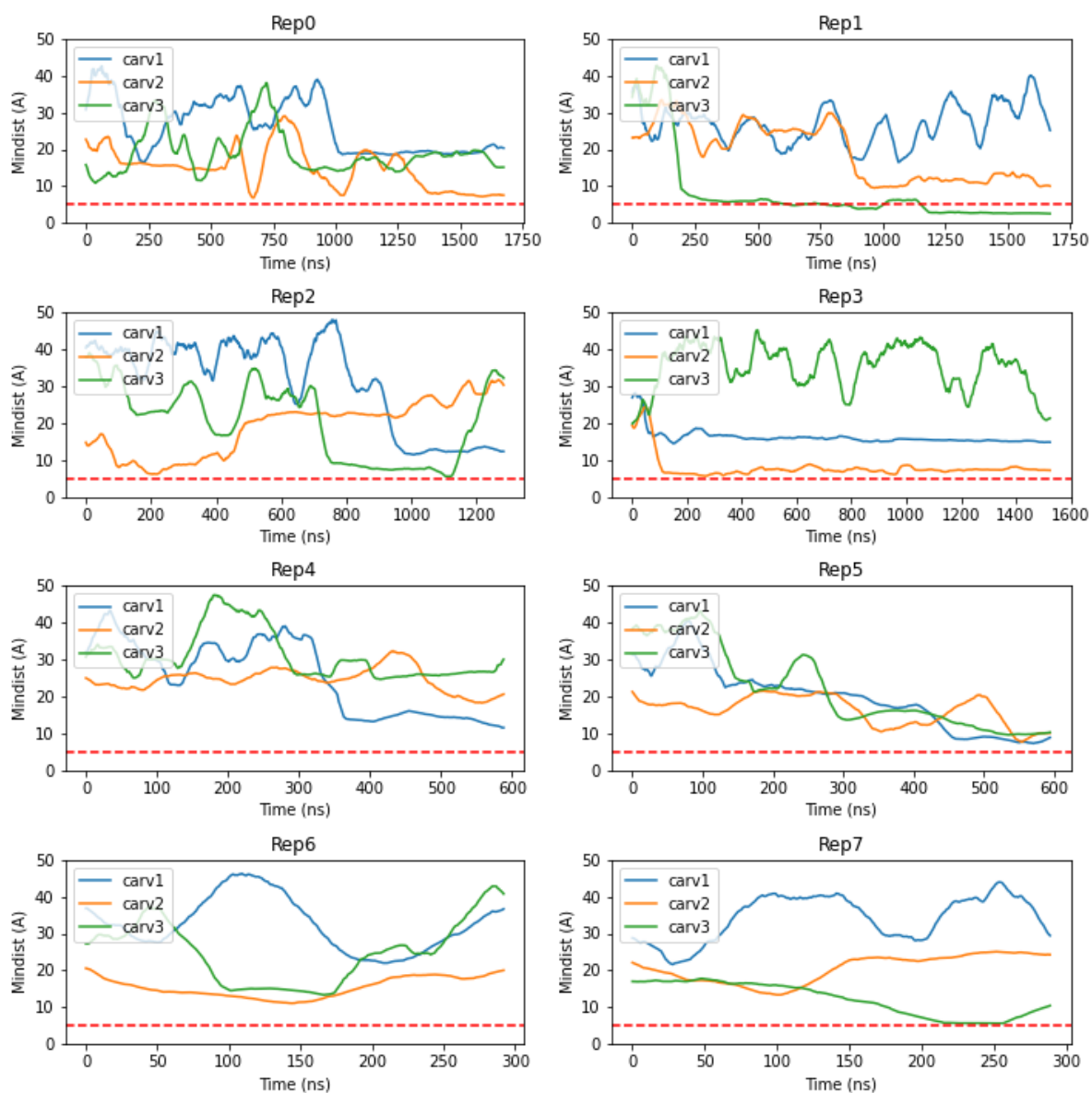
Conflicts of Interest

The authors declare no conflict of interest.

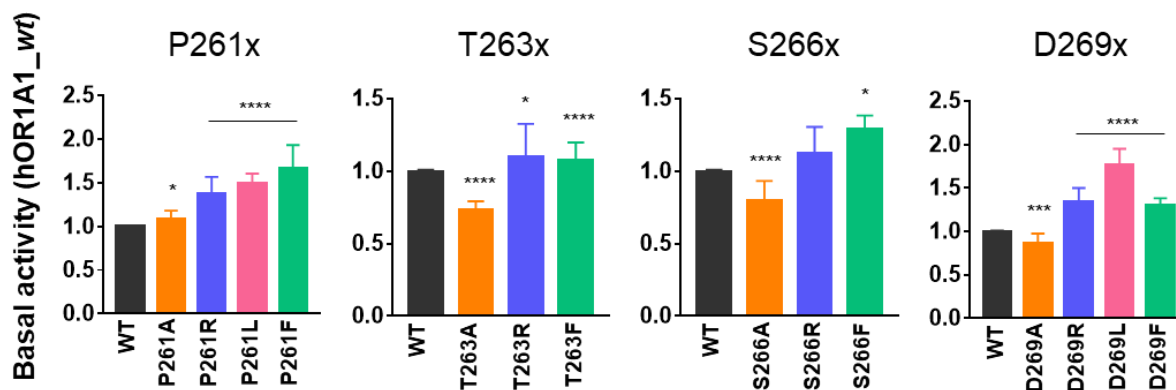
Supplementary material



Supplementary figure S1: Time evolution of the Root Mean Square Deviation (RMSD) of the 6 replicas of hOR1A1 studied with respect to time. The RMSD is computed for the receptor backbone with respect to the initial model structure.



Supplementary figure S2: Evolution of the minimum distance between the center of mass of the three (-)-carvone molecules (in blue, yellow and green) and the center of mass of residue Y251 (the cradle of the binding cavity). The dashed red line at 5 Å is the threshold to consider an interaction between a molecule and the Y251.



Supplementary figure S3: Basal activity graph of wt hOR1A1 and mutants at positions P261, T263, S266, and D269 (two-tailed unpaired t-test; * $p < 0.05$, *** $p < 0.001$, and **** $p < 0.0001$).

References

1. Debat, H.; Eloit, C.; Blon, F.; Sarazin, B.; Henry, C.; Huet, J.C.; Trotier, D.; Pernollet, J.C. Identification of human olfactory cleft mucus proteins using proteomic analysis. *J Proteome Res* **2007**, *6*, 1985-1996, doi:10.1021/pr0606575.
2. Nagashima, A.; Touhara, K. Enzymatic conversion of odorants in nasal mucus affects olfactory glomerular activation patterns and odor perception. *J Neurosci* **2010**, *30*, 16391-16398, doi:10.1523/JNEUROSCI.2527-10.2010.
3. Pelosi, P.; Knoll, W. Odorant-binding proteins of mammals. *Biol Rev Camb Philos Soc* **2022**, *97*, 20-44, doi:10.1111/brv.12787.
4. Buck, L.; Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **1991**, *65*, 175-187, doi:10.1016/0092-8674(91)90418-x.
5. Malnic, B.; Godfrey, P.A.; Buck, L.B. The human olfactory receptor gene family. *Proceedings of the National Academy of Sciences* **2004**, *101*, 2584-2589.
6. Muratore, A.; Clinet, J.C.; Dunach, E. Synthesis of new exo- and endo-3,8-dihydro-beta-santalols and other norbornyl-derived alcohols. *Chem Biodivers* **2010**, *7*, 623-638, doi:10.1002/cbdv.200900119.
7. Kraft, P.; Frater, G. Enantioselectivity of the musk odor sensation. *Chirality* **2001**, *13*, 388-394, doi:10.1002/chir.1050.

8. Trimmer, C.; Keller, A.; Murphy, N.R.; Snyder, L.L.; Willer, J.R.; Nagai, M.H.; Katsanis, N.; Vosshall, L.B.; Matsunami, H.; Mainland, J.D. Genetic variation across the human olfactory receptor repertoire alters odor perception. *Proc Natl Acad Sci U S A* **2019**, *116*, 9475-9480, doi:10.1073/pnas.1804106115.
9. Gerkin, R.C.; Castro, J.B. The number of olfactory stimuli that humans can discriminate is still unknown. *Elife* **2015**, *4*, e08127, doi:10.7554/eLife.08127.
10. Bushdid, C.; Magnasco, M.O.; Vosshall, L.B.; Keller, A. Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **2014**, *343*, 1370-1372, doi:10.1126/science.1249168.
11. Lawson, Z.; Wheatley, M. The third extracellular loop of G-protein-coupled receptors: more than just a linker between two important transmembrane helices. *Biochem Soc Trans* **2004**, *32*, 1048-1050, doi:10.1042/BST0321048.
12. Wheatley, M.; Wootten, D.; Conner, M.T.; Simms, J.; Kendrick, R.; Logan, R.T.; Poyner, D.R.; Barwell, J. Lifting the lid on GPCRs: the role of extracellular loops. *Br J Pharmacol* **2012**, *165*, 1688-1703, doi:10.1111/j.1476-5381.2011.01629.x.
13. Peeters, M.C.; van Westen, G.J.; Li, Q.; AP, I.J. Importance of the extracellular loops in G protein-coupled receptors for ligand recognition and receptor activation. *Trends Pharmacol Sci* **2011**, *32*, 35-42, doi:10.1016/j.tips.2010.10.001.
14. Wheatley, M.; Simms, J.; Hawtin, S.R.; Wesley, V.J.; Wootten, D.; Conner, M.; Lawson, Z.; Conner, A.C.; Baker, A.; Cashmore, Y.; et al. Extracellular loops and ligand binding to a subfamily of Family A G-protein-coupled receptors. *Biochem Soc Trans* **2007**, *35*, 717-720, doi:10.1042/BST0350717.
15. Dror, R.O.; Arlow, D.H.; Maragakis, P.; Mildorf, T.J.; Pan, A.C.; Xu, H.; Borhani, D.W.; Shaw, D.E. Activation mechanism of the beta2-adrenergic receptor. *Proc Natl Acad Sci U S A* **2011**, *108*, 18684-18689, doi:10.1073/pnas.1110499108.
16. Bushdid, C.; de March, C.A.; Topin, J.; Do, M.; Matsunami, H.; Golebiowski, J. Mammalian class I odorant receptors exhibit a conserved vestibular-binding pocket. *Cell Mol Life Sci* **2019**, *76*, 995-1004, doi:10.1007/s00018-018-2996-4.
17. Charlier, L.; Topin, J.; Ronin, C.; Kim, S.K.; Goddard, W.A., 3rd; Efremov, R.; Golebiowski, J. How broadly tuned olfactory receptors equally recognize their agonists. Human OR1G1 as a test case. *Cell Mol Life Sci* **2012**, *69*, 4205-4213, doi:10.1007/s00018-012-1116-0.

18. 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* **2015**, *24*, 1543-1548, doi:10.1002/pro.2717.
19. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583-589, doi:10.1038/s41586-021-03819-2.
20. Jakubik, J.; Randakova, A.; Zimcik, P.; El-Fakahany, E.E.; Dolezal, V. Binding of N-methylscopolamine to the extracellular domain of muscarinic acetylcholine receptors. *Sci Rep* **2017**, *7*, 40381, doi:10.1038/srep40381.
21. 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* **2017**, *42*, 181-193, doi:10.1093/chemse/bjw117.
22. Geithe, C.; Krautwurst, D. Chirality Matters - Enantioselective Orthologous Odorant Receptors for Related Terpenoid Structures. *Acs Sym Ser* **2015**, *1212*, 161-181.
23. Geithe, C.; Protze, J.; Kreuchwig, F.; Krause, G.; Krautwurst, D. Structural determinants of a conserved enantiomer-selective carvone binding pocket in the human odorant receptor OR1A1. *Cell Mol Life Sci* **2017**, *74*, 4209-4229, doi:10.1007/s00018-017-2576-z.
24. de March, C.A.; Yu, Y.; Ni, M.J.; Adipietro, K.A.; Matsunami, H.; Ma, M.; Golebiowski, J. Conserved Residues Control Activation of Mammalian G Protein-Coupled Odorant Receptors. *J Am Chem Soc* **2015**, *137*, 8611-8616, doi:10.1021/jacs.5b04659.
25. Schmiedeberg, K.; Shirokova, E.; Weber, H.P.; Schilling, B.; Meyerhof, W.; Krautwurst, D. Structural determinants of odorant recognition by the human olfactory receptors OR1A1 and OR1A2. *J Struct Biol* **2007**, *159*, 400-412, doi:10.1016/j.jsb.2007.04.013.
26. de March, C.A.; Topin, J.; Bruguera, E.; Novikov, G.; Ikegami, K.; Matsunami, H.; Golebiowski, J. Odorant Receptor 7D4 Activation Dynamics. *Angew Chem Int Ed Engl* **2018**, *57*, 4554-4558, doi:10.1002/anie.201713065.
27. Robertson, M.J.; Meyerowitz, J.G.; Panova, O.; Borrelli, K.; Skiniotis, G. Plasticity in ligand recognition at somatostatin receptors. *Nat Struct Mol Biol* **2022**, *29*, 210-+, doi:10.1038/s41594-022-00727-5.
28. Asada, H.; Horita, S.; Hirata, K.; Shiroishi, M.; Shiimura, Y.; Iwanari, H.; Hamakubo, T.; Shimamura, T.; Nomura, N.; Kusano-Arai, O.; et al. Crystal structure of the human

angiotensin II type 2 receptor bound to an angiotensin II analog. *Nat Struct Mol Biol* **2018**, *25*, 570-576, doi:10.1038/s41594-018-0079-8.

29. Vu, O.; Bender, B.J.; Pankewitz, L.; Huster, D.; Beck-Sickinger, A.G.; Meiler, J. The Structural Basis of Peptide Binding at Class A G Protein-Coupled Receptors. *Molecules* **2021**, *27*, doi:10.3390/molecules27010210.

30. Webb, B.; Sali, A. Protein structure modeling with MODELLER. *Methods Mol Biol* **2014**, *1137*, 1-15, doi:10.1007/978-1-4939-0366-5_1.

31. Cook, B.L.; Steuerwald, D.; Kaiser, L.; Graveland-Bikker, J.; Vanberghem, M.; Berke, A.P.; Herlihy, K.; Pick, H.; Vogel, H.; Zhang, S. Large-scale production and study of a synthetic G protein-coupled receptor: human olfactory receptor 17-4. *Proc Natl Acad Sci U S A* **2009**, *106*, 11925-11930, doi:10.1073/pnas.0811089106.

32. Shen, M.Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* **2006**, *15*, 2507-2524, doi:10.1110/ps.062416606.

33. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **2022**, *50*, D439-D444, doi:10.1093/nar/gkab1061.

34. Olsson, M.H.; Sondergaard, C.R.; Rostkowski, M.; Jensen, J.H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* **2011**, *7*, 525-537, doi:10.1021/ct100578z.

35. 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* **2012**, *40*, D370-376, doi:10.1093/nar/gkr703.

36. Schott-Verdugo, S.; Gohlke, H. PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. *J Chem Inf Model* **2019**, *59*, 2522-2528, doi:10.1021/acs.jcim.9b00269.

37. Lee, T.S.; Cerutti, D.S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T.J.; Roitberg, A.; Case, D.A.; Walker, R.C.; York, D.M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J Chem Inf Model* **2018**, *58*, 2043-2050, doi:10.1021/acs.jcim.8b00462.

38. Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.;

Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision E.01. **2009**.

39. Wang, J.M.; Cieplak, P.; Kollman, P.A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* **2000**, *21*, 1049-1074, doi:10.1002/1096-987x(200009)21:12<1049::Aid-Jcc3>3.0.Co;2-F.

40. Roe, D.R.; Cheatham, T.E., 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013**, *9*, 3084-3095, doi:10.1021/ct400341p.

41. Shepard, B.D.; Natarajan, N.; Protzko, R.J.; Acres, O.W.; Pluznick, J.L. A cleavable N-terminal signal peptide promotes widespread olfactory receptor surface expression in HEK293T cells. *PLoS One* **2013**, *8*, e68758, doi:10.1371/journal.pone.0068758.

42. Hogrefe, H.H.; Cline, J.; Youngblood, G.L.; Allen, R.M. Creating randomized amino acid libraries with the QuikChange Multi Site-Directed Mutagenesis Kit. *Biotechniques* **2002**, *33*, 1158-1160, 1162, 1164-1155, doi:10.2144/02335pf01.

43. Zhuang, H.; Matsunami, H. Evaluating cell-surface expression and measuring activation of mammalian odorant receptors in heterologous cells. *Nat Protoc* **2008**, *3*, 1402-1413, doi:10.1038/nprot.2008.120.

Conclusions

The first step of olfactory perception begins at the level of membrane receptors located on the surface of olfactory neurons. Olfactory receptors (ORs) are metabotropic in vertebrates, ionotropic in insects. It is at this level that chemical information (odorants) is transformed by a subtle signaling cascade into electrical information (action potential): the signal transduction. ORs are extremely varied, and each subtype of OR recognize one or a particular spectrum of ligands. The encounter between the odorous substances and this set of receptors creates a combinatorial code that the brain will use to discriminate the molecules and finally create an olfactory percept and the behavioral response that follows.

The lack of information is the main lock that slows down the understanding of the mechanisms of olfaction. In 2022, there are still 46.5% human ORs that are orphans, i.e. without known ligands. It is crucial to accelerate the deorphanization of ORs. There is also no experimental vertebrate receptor structure currently available, nor an experimental structure of the Orco/OR complex in insects. In this work, we combined several molecular modeling approaches with *in vitro* functional testing techniques to better understand the sequence-structure-function relationships of insect and vertebrate ORs.

In chapter 1, the combined approach of homology modeling, MD and two patch voltage clamps on *Xenopus laevis* oocytes allowed us to predict the diffusion pathway of a ligand to its binding site on the insect Orco. Sequence conservation analysis on residues identified as important reveals that the diffusion pathway is highly conserved in insects. VUAA1 follows a stepwise desolvation process through 2 vestibules to its active site. An analysis of the physicochemical properties of the cavities identifies polarity as having a profound influence on the ability of VUAA1 to activate the channel. Furthermore, a comparison with the experimental structure of MhOR5 reveals that the binding site is conserved between Orco and the ORs, supporting the hypothesis of a divergent evolution of Orco from ancient ORs functioning as homotetramers.

In chapter 2, we predicted the impact of a polymorphism on TAAR5 activation by trimethylamine using a combined approach of model building using AlphaFold2 (AF2), molecular dynamics simulations, and site-directed mutagenesis. We demonstrated that although the mutant is still able to be expressed in the membrane, the mutation abolishes the response of the receptor to various amines, explaining the altered perception of fish odor. Changes in the allosteric interaction network are identified between TM2 and TM7, making receptor activation impossible. It is interesting to note that the mutation at the same position in TAAR1 has the same outcome.

In chapter 3, we focused on two important parts for the function of vertebrate ORs: In the first article (3a), extracellular loop 2 (ECL2) that is extremely diverse across class A GPCRs and have been identified as important for ligand binding and receptor activation. The ECL2 of ORs is extremely long among the class A family of GPCRs. Through a homology modeling, molecular dynamics, site-directed mutagenesis, we observe that ECL2 features modulate the recognition spectrum of mOR256-3 and mOR256-8, acting as a hydrophobic and flexible lid on the active site. Thus, the models suggest that the broader recognition spectrum of mOR256-3 would be explained by a much larger cavity than that of mOR253-8, divided into 2 connected parts. A virtual screening allowed us to identify 6 new agonists and one antagonist of mOR256-3. In the second paper (3b), we considered the function of ECL3 in class 2 ORs, using homology modeling, coupled with molecular dynamics and site-directed mutagenesis. ECL3 forms a vestibule and interacts with ligands in their passage to the active site in class I ORs and some GPCRs like the beta 2 adrenergic receptor. Surprisingly, this role seems to be conserved in class II ORs despite a high sequence diversity and their specific PxSxxS motif.

The work presented in this thesis is an example of the strength of using both computational and experimental methods to better understand the molecular basis of olfactory perception. I had the chance to live several revolutions during this PhD (experimental structure of *AbakOrco*, then of *MhOR5*, arrival of AlphaFold, recent announcement of an experimental structure of a mammalian consensus OR). These advances will certainly lead to a better understanding of the structure-function relationships of ORs, whether metabotropic or ionotropic, as we have begun to do since the discovery of olfactory genes. Faced with the complexity of the combinatorial code, enhanced sampling techniques seem a good avenue to apply the methods presented here on a larger scale.

Discussion

The quality of the models directly dictates the quality of the predictions obtained by the molecular modeling techniques. Several points are to be considered and could be improved.

Sequence identity and template choice

The model of the DmelOrco homotetramer was generated from the experimental structure of the *AbakOrco* [1]. With 62% sequence identity, the alignment with the template is trivial and the quality of the model can be considered excellent in the transmembrane regions. hTAAR5, which has all the characteristics of the aminergic A class, is also a good candidate for homology modeling with the available templates at 36% sequence identity. This is not the case for the mouse mOR256-3/8 and human hOR1A1 homology models, which show less than 20% sequence identity with the best available model. *In vitro* testing is mandatory to validate the sequence alignment, as mentioned before [2] and AF2 may be a better option in this tricky case for homology modeling.

N terminal part

The N-terminal part of GPCRs is often unresolved in experimental structures due to its high flexibility. It is common not to consider it when modeling by homology because they are disordered and usually highly variable in sequence and length. However, variability does not necessarily mean uselessness [3]. TAAR family possess two conserved cysteine in the N terminal region (C17 and C24 for TAAR5), which allows the receptor to probably form 3 disulfide bridges in the extracellular part of the receptor, unlike ORs and most GPCRs which can only form 2 [4]. Our AF2 model of TAAR1 and TAAR5 has taken this part of the receptor into account, but further experimental testing is needed to confidently determine the correct combination of disulfide bridges and guide the modeling.

Limits of AlphaFold2?

AF2 has been received by the community as a paradigm shift in molecular modeling. Indeed, several studies already attest to the quality of the models generated for various proteins [5, 6]. Moreover, AF2 is versatile and has surpassed or equaled the state of the art in various problems, such as the prediction of disordered protein structures [7] or protein-protein docking [8].

However, several obstacles remain to be overcome before considering the problem of protein folding as solved. Indeed, the prediction of the structural impact of mutations on the stability or function of a protein is still not possible [9]. Moreover, even if AF2 accurately predicts the native structure of a protein, it is far from precisely informing its folding path or different microstates [10]. AF2 also only provides a structure that is biased towards the active or inactive form depending on the class in the case of GPCRs. To address this need, Heo *et Feig* (2022) developed a protocol to model GPCRs with different states by refining the information provided to AF, notably at the level of templates and MSA [11].

As seen previously, loop modeling is a problem in itself when building a model. Even if AF2 keeps its accuracy for loops of less than 10 amino acids, it loses in prediction quality when the length exceeds 20 residues. Moreover, alpha helices and beta sheets are slightly over predicted [12]. This raises a concern about the modeling of insect and mammalian ORs by AF2. Even in the case where an experimental structure is available, some parts are not solved. This is the case of the ICL2 in the *AbakOrco* structure [1]. When modeling the *S. littoralis* Orco (*SlitOrco*) monomer with "monomeric" AF2, the ICL2 is a disordered loop. However, when using the multimeric version of AF2, and modelling both *SlitOrco* together with its *SlitOR24*, the ECL2 is modeled as 2 helices (figure 1a). On the mammalian OR side, AF2 forms a small alpha helix at ECL2, this area being predominantly disordered in homology modeling (figure 1b). An experimental structure of a "consensus" receptor for OR was recently presented at the European Chemoreception Research Organisation (ECRO) conference in 2022. The ECL2 was folded as a short α -helix as for the AF2 model. However, the consensus OR receptor was specifically designed to increase its stability and may not be representative of a true OR, so the question of whether this small helix exists or whether AF2 overpredicts helices remains open.

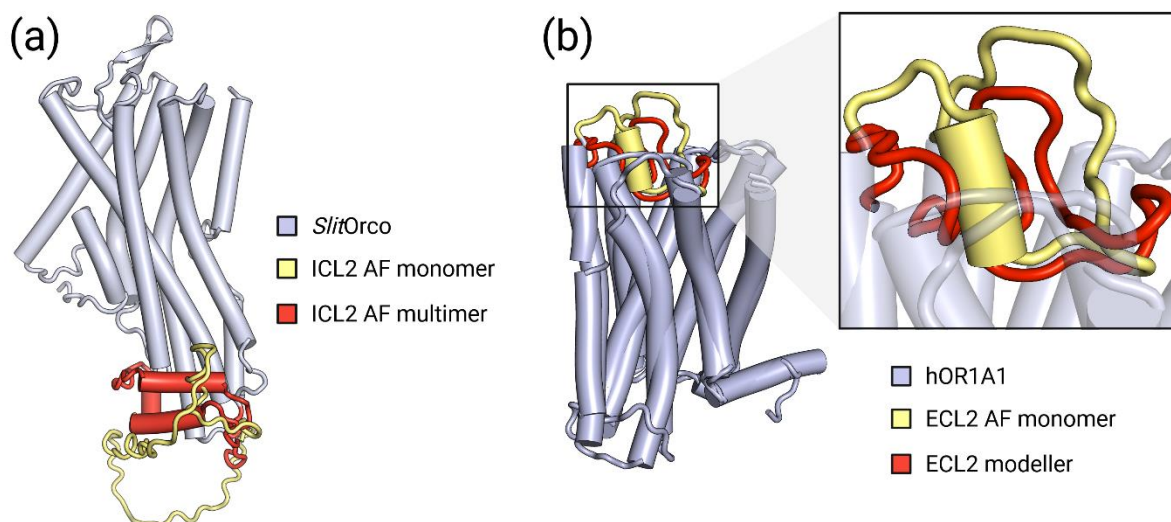


Figure 1: (a) Comparison between AF2 monomer (SlitOrco) and AF2 multimer models (modeled with 2 SlitOrco and 2 SlitOR24 not shown), focus on SlitOrco (blue). ICL2 is shown in red and yellow for the SlitOrco model generated with AF2 multimer and monomer respectively. The multimer model has an ECL2 with 2 alpha helices, the AF2 monomer model an unstructured loop. (b) Comparison between the model of hOR1A1 obtained by AF2 (yellow ECL2) or by modeller (red ECL2). The ECL2 from AF2 forms a small alpha helix, that from modeller an unstructured loop.

Perspectives

Reclassification of mammal ORs

The so-called Glusman classification of the OR family of vertebrates is based on a divergence model of evolution created by phylogenetic analysis of sequences [13]. This classification is useful, but there is a disparity between the recognition spectrum of the ORs and the current classification (figure 2a and b). This means that some receptors classified in the same family will recognize different ligands, and conversely some receptors distant in terms of sequence will recognize similar ligands. To try to improve this classification, we propose a new approach to classify ORs, based on their binding pockets properties retrieve from structural information. Hierarchical clustering analysis (HCA) classification using cavity descriptors (polarity, hydrophobicity, volume) improves the match between ligand chemical space and receptor space (Figure 2c).

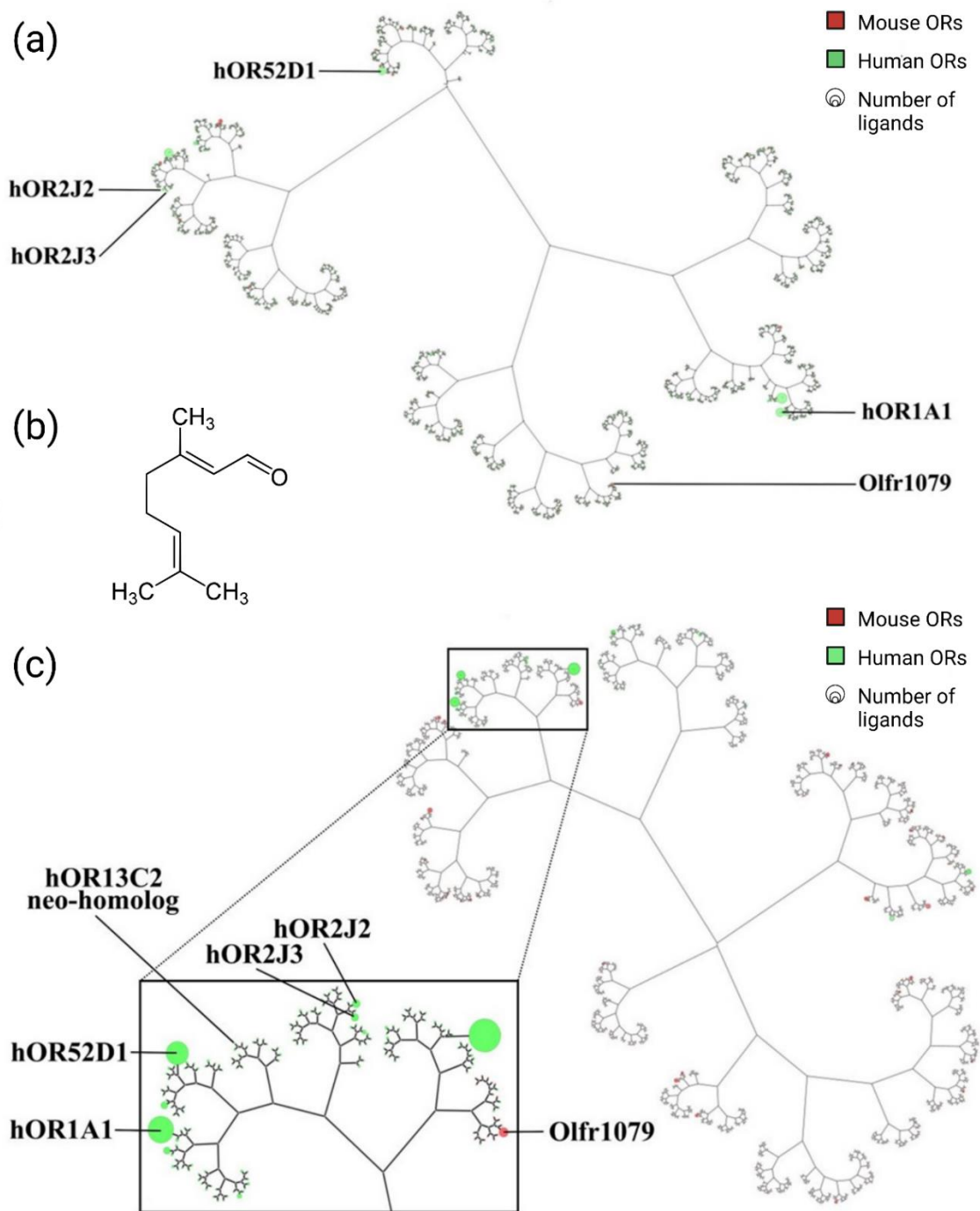


Figure 2: (a) Phylogenetic tree of ~1100 mice and ~400 humans ORs based on sequence identity [13]. The mouse and human ORs are represented by red and green circles, respectively. The size of the circles represents the number of known odorants per receptor (from orphan to 66 ligands). Each named receptor is a known citral binder. (b) Structure of citral. (c) Version of the structural phylogenetic tree based on the pocket descriptors for each receptor (volume, hydrophobicity, polarity). Citral-binding receptors are located on the same branch of the pocket-based structural phylogenetic tree, which improves the correspondence between pocket and chemical spectrum compared to the Glusmann classification.

The development of a score function assessing the overlap between ligand and pocket space will guide the choice of pocket descriptors to improve the actual classification. Considering the assumption that 2 receptors with similar cavities can have similar chemical spaces, experimentally testing the predicted "neo-homologs" (i.e., an orphan receptor with a binding pocket similar to a deorphanized receptor) will allow us to evaluate the quality of the new classification. All the methodologies needed for this project is already developed and the work can be done in a close future.

Structure-emotion relationships

Odorant molecules are known to provoke various affects when we sniff them. It has been proven for example that the smell of lavender leads to a relaxing and soothing effect (figure 3, [14, 15]), well known in aromatherapy. On the contrary, some molecules are strongly aversive and can stress us [16]. Can we predict an emotion from the structure of a molecule? The aim of this ongoing research project is to identify whether there are implicit links between the structure of odorants and their effect on our emotional state. Such links may be identified through the use of machine learning.

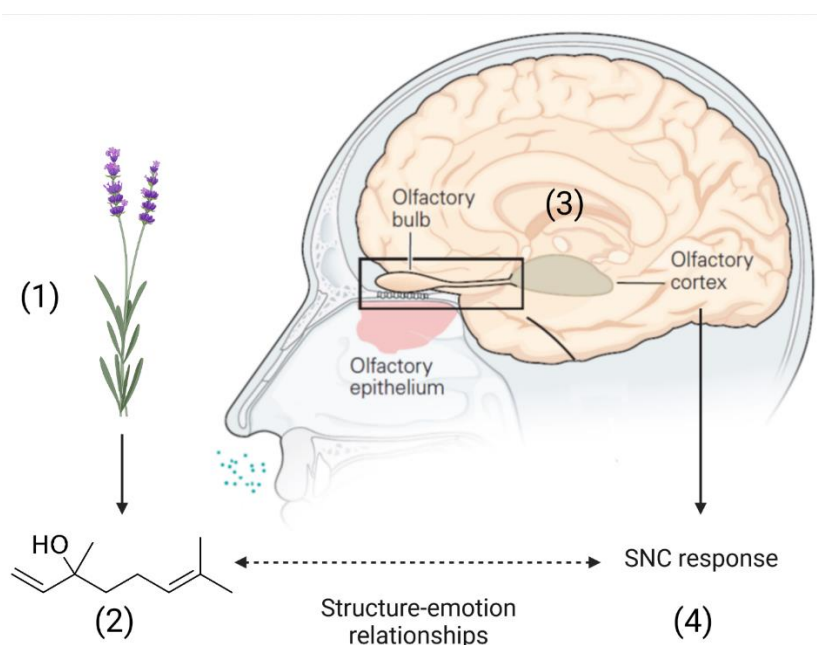


Figure 3: Lavender (1) emits odorant molecules like linalool (2). The olfactory cue is processed by the olfactory system (olfactory epithelium, olfactory bulb, and olfactory cortex) to create a central nervous system response: relaxation. Adapted from "Principles of neural science 6th edition" [17].

We have built a first database containing the psychophysiological reactions (skin temperature, respiratory rate, skin conductance and heart rate) of volunteers stimulated by different odorants [18]. This database of 66 molecules is progressively enriched, and we are currently testing different statistical models on it. The molecules were divided into 3 roughly equal groups (relaxing, neutral, and stimulating). Several classification models were tested (figure 4) but for the moment are not satisfying. This illustrates one limit of ML approaches: training such models sometimes requires a large dataset, especially for problems with numerous variables. Here the dataset is relatively small and several factors (age, sex, culture, experience or internal state) may influence the psychophysiological response of a subject to an odorant [18, 19]. Such long-term project then requires more experimental recordings and goes beyond the current PhD thesis but may be achieved in a close future.

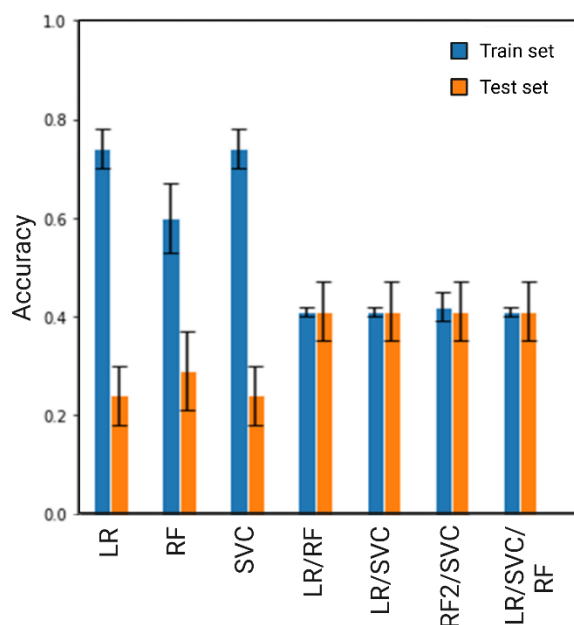


Figure 4: Accuracy for different 3 class classifications models. Error bars gathered by 5 fold cross validation. LR, RF and SVC for logistic regression, random forest and standard vector classification respectively. Performances for train set and test set in blue or orange respectively.

References

1. Butterwick JA, Del Marmol J, Kim KH, et al (2018) Cryo-EM structure of the insect olfactory receptor Orco. *Nature* 560:447–452. <https://doi.org/10.1038/s41586-018-0420-8>
2. Yu Y, Ma Z, Pacalon J, et al (2022) Extracellular loop 2 of G protein–coupled olfactory receptors is critical for odorant recognition. *Journal of Biological Chemistry* 298:. <https://doi.org/10.1016/j.jbc.2022.102331>
3. Lindner D, Walther C, Tennemann A, Beck-Sickinger AG (2009) Functional role of the extracellular N-terminal domain of neuropeptide Y subfamily receptors in membrane integration and agonist-stimulated internalization. *Cell Signal* 21:61–68. <https://doi.org/10.1016/j.cellsig.2008.09.007>
4. Naranjo AN, Chevalier A, Cousins GD, et al (2015) Conserved disulfide bond is not essential for the adenosine A2A receptor: Extracellular cysteines influence receptor distribution within the cell and ligand-binding recognition. *Biochim Biophys Acta* 1848:603–614. <https://doi.org/10.1016/j.bbamem.2014.11.010>
5. van Breugel M, Rosa e Silva I, Andreeva A (2022) Structural validation and assessment of AlphaFold2 predictions for centrosomal and centriolar proteins and their complexes. *Commun Biol* 5:1–10. <https://doi.org/10.1038/s42003-022-03269-0>
6. Skuhersky MA, Tao F, Qing R, et al (2021) Comparing Native Crystal Structures and AlphaFold2 Predicted Water-Soluble G Protein-Coupled Receptor QTY Variants. *Life (Basel)* 11:1285. <https://doi.org/10.3390/life11121285>
7. Akdel M, Pires DEV, Pardo EP, et al (2021) A structural biology community assessment of AlphaFold 2 applications. 2021.09.26.461876
8. Evans R, O'Neill M, Pritzel A, et al (2022) Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034. <https://doi.org/10.1101/2021.10.04.463034>
9. Pak MA, Markhieva KA, Novikova MS, et al (2021) Using AlphaFold to predict the impact of single mutations on protein stability and function. 2021.09.19.460937
10. Outeiral C, Nissley DA, Deane CM (2022) Current structure predictors are not learning the physics of protein folding. *Bioinformatics* 38:1881–1887. <https://doi.org/10.1093/bioinformatics/btab881>
11. Heo L, Feig M Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins: Structure, Function, and Bioinformatics* n/a: <https://doi.org/10.1002/prot.26382>
12. Stevens AO, He Y (2022) Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction. *Biomolecules* 12:985. <https://doi.org/10.3390/biom12070985>
13. Glusman G, Bahar A, Sharon D, et al (2000) The olfactory receptor gene superfamily: data mining, classification, and nomenclature. *Mammalian Genome: Official Journal of the International Mammalian Genome Society* 11:1016–1023

14. Linck VM, da Silva AL, Figueiró M, et al (2010) Effects of inhaled Linalool in anxiety, social interaction and aggressive behavior in mice. *Phytomedicine* 17:679–683. <https://doi.org/10.1016/j.phymed.2009.10.002>
15. Karan NB (2019) Influence of lavender oil inhalation on vital signs and anxiety: A randomized clinical trial. *Physiology & Behavior* 211:112676. <https://doi.org/10.1016/j.physbeh.2019.112676>
16. Hermann C, Ziegler S, Birbaumer N, Flor H (2000) Pavlovian aversive and appetitive odor conditioning in humans: subjective, peripheral, and electrocortical changes. *Exp Brain Res* 132:203–215. <https://doi.org/10.1007/s002210000343>
17. Kandel ER, Koester JD, Mack SH, Siegelbaum SA (2021) *Principles of Neural Science, Sixth Edition, 6th edition*. McGraw Hill / Medical
18. Bontempi C, Jacquot L, Brand G (2021) Sex Differences in Odor Hedonic Perception: An Overview. *Front Neurosci* 15:764520. <https://doi.org/10.3389/fnins.2021.764520>
19. Oleszkiewicz A, Alizadeh R, Altundag A, et al (2020) Global study of variability in olfactory sensitivity. *Behav Neurosci* 134:394–406. <https://doi.org/10.1037/bne0000378>

Annexes

Publication 5

Functional molecular switches of mammalian G protein-coupled bitter-taste receptors

Jérémie Topin, Cédric Bouysset, Jody Pacalon, Yiseul Kim, Mee-Ra Rhyu, Sébastien Fiorucci & Jérôme Golebiowski.

Cell Mol Life Sci . 2021 Dec;78(23):7605-7615. doi: 10.1007/s00018-021-03968-7. Epub 2021 Oct 23.

Abstract

Bitter taste receptors (TAS2Rs) are a poorly understood subgroup of G protein-coupled receptors (GPCRs). The experimental structure of these receptors has yet to be determined, and key-residues controlling their function remain mostly unknown. We designed an integrative approach to improve comparative modeling of TAS2Rs. Using current knowledge on class A GPCRs and existing experimental data in the literature as constraints, we pinpointed conserved motifs to entirely re-align the amino-acid sequences of TAS2Rs. We constructed accurate homology models of human TAS2Rs. As a test case, we examined the accuracy of the TAS2R16 model with site-directed mutagenesis and *in vitro* functional assays. This combination of *in silico* and *in vitro* results clarifies sequence-function relationships and proposes functional molecular switches that encode agonist sensing and downstream signaling mechanisms within mammalian TAS2Rs sequences.

Introduction

Bitterness is one of the basic taste modalities detected by the gustatory system. It is generally considered to be a warning against the intake of noxious compounds [1] and, as such, is often associated with disgust and food avoidance [2]. At the molecular level, this perception is initiated by the activation of bitter taste receptors. In humans, 25 genes functionally express these so-called type 2 taste receptors (TAS2Rs), which provide the capacity to detect a wide array of bitter chemicals [3]. Further, TAS2Rs are also ectopically expressed in non-chemosensory tissues, making them important emerging pharmacological targets [4–6]. TAS2Rs are G protein-coupled receptors [7] (GPCRs) classified as distantly related to class A GPCRs. They were previously classified with class F GPCRs [8] and more recently as a separate sixth class evolved from class A [9, 10]. The sequence similarity between TAS2Rs and class A GPCRs is in the range of 14–29% [11]. Structure-based sequence alignment has placed TAS2Rs in the class A family, which contains the olfactory chemosensory receptors sub-family [12]. TAS2Rs have been recently labelled as class T in the GPCR database (GPCRdb) (Fig.1a) [13]. Structurally, GPCRs are made up of seven transmembrane (TM) helices named TM1 to TM7 that form a bundle across the cell membrane. How GPCRs achieve specific robust signaling and how these functions are encoded in their sequences are pending fundamental questions.

GPCR activation relies on so-called molecular switches, which allosterically connect the ligand binding pocket to the intra- cellular G protein coupling site to trigger downstream signaling [14]. In class A GPCRs (including olfactory receptors, ORs), these molecular switches consist of conserved sequence motifs (Fig.1c). The “toggle/transmission switch” CWxP TM6 (or FYGx TM6 in ORs) senses agonist binding. The other motifs, which propagate the signal, include the “hydrophobic connector” PIF TM3-5-6, the NPxxY TM7, the “ionic lock” DRY TM3, and a hydrophobic barrier between the last two [15–18]. To date, experimental structures have not been determined for any TAS2Rs, but the following hallmark motifs have been defined based on sequence conservation: NGFI TM1, LAxSR TM2, KIANFS TM3, LLG TM4, PF TM5, HxKALKT TM6, YFL TM6, and PxxHSFIL TM7 [7]. These conserved motifs are highly dissimilar between TAS2Rs and class A GPCRs (Fig. 1b, d, Table1 and TableS1), leading to various sequence alignments (TableS2). The main differences occur in TM3, TM4, TM6, and TM7 [11, 19–30], making it difficult to infer TAS2R functional molecular switches. These alternative ways to align the sequences remain a central issue in understanding the complex allosteric TAS2R machinery. The present study aims to identify the molecular switches that control TAS2R functions. We present an integrative protocol that advances comparative modeling of TAS2Rs. Case studies of site-directed mutagenesis followed by invitro functional assays on human TAS2R16 then evaluated the roles of the predicted molecular switches in TAS2Rs.

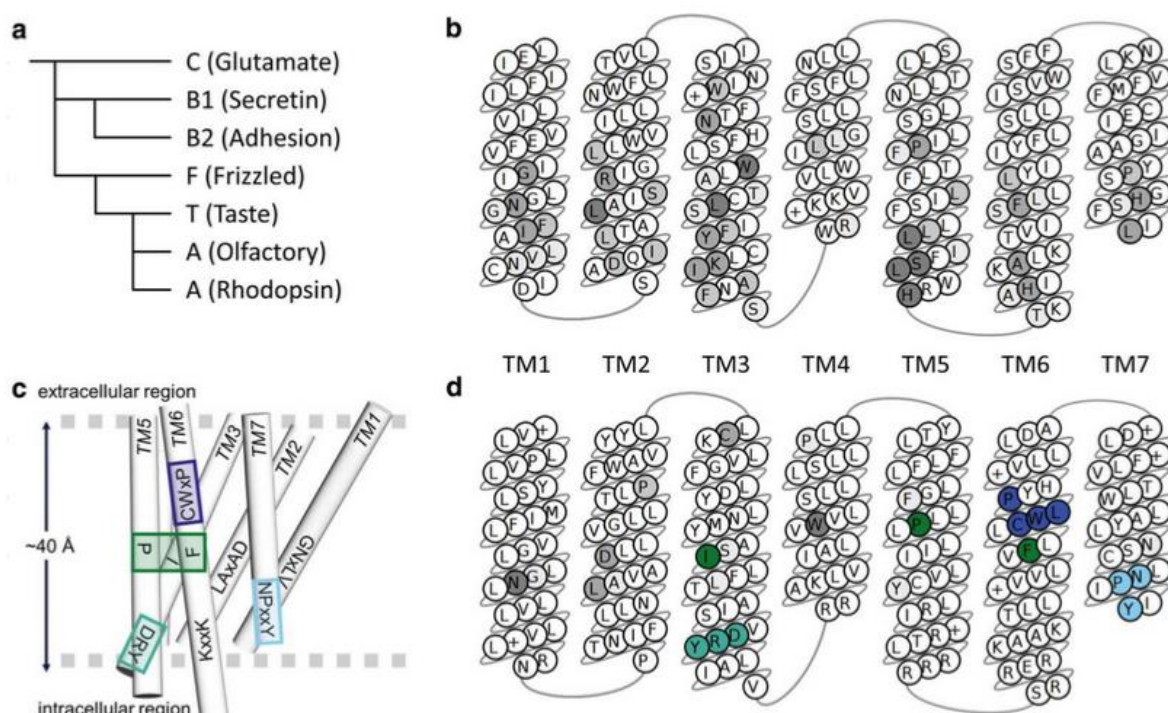


Fig. 1: a Schematic phylogenetic tree of GPCR classes according to Cvicek *et al.* [12]. b Snake plot representation of transmembrane segments (TM) of mammalian TAS2Rs consensus sequences, colored in grey scale according to sequence conservation. c Non-olfactory class A GPCR sequence hallmarks (transmission switch in blue, hydrophobic connector in green, ionic lock in sea green, hydrophobic barrier in light blue). d Snake plot representation of non-olfactory class A GPCR consensus sequences.

Table 1: Key residues and consensus motifs

TM	class A GPCR	OR	TAS2R
1	GN ^{1.50} _x LV	GN ^{1.50} LLI	N ^{1.50} GFI
2	LAXAD ^{2.50}	LSFXD ^{2.50}	LAXSR ^{2.50}
3	L ^{3.43}	L ^{3.43}	L ^{3.43}
	DR ^{3.50} Y	MAYDR ^{3.50} YVAIC	K ^{3.50} IANFS
4	W ^{4.50}	W ^{4.50}	L ^{4.50} LG
5	P ^{5.50}	P ^{5.50} F	P ^{5.50} F
	Y ^{5.58}	Y ^{5.58}	F ^{5.58}
6	K ^{6.32} _{xx} K	RxK ^{6.32} AFSTC	HxK ^{6.32} ALKT
	CW ^{6.48} LP	FY ^{6.48} G	YF ^{6.48} L
7	SxxNP ^{7.50} _{xx} Y	PxxNP ^{7.50} _x IY	PxxHS ^{7.50} FIL

Superscripts refer to the Ballesteros–Weinstein numbering scheme.

Methods

Sequence alignment

Automatic multiple sequence alignment (MSA) of TAS2Rs was performed with class A and class F templates (labelled *ClustalO* and *classF*, respectively) using ClustalO [31] with default settings in the Jalview interface (v2.11.0) [32]. These MSAs were not modified. Another MSA, labelled *Chemosim*, was completed using class A templates, 339 class II ORs and TAS2Rs. In particular, the inclusion of OR sequences is of major importance for the alignment of TM3, TM6 and TM7. TAS2Rs are missing the specific DRY^{TM3}, CWLP^{TM6} and NPxxY^{TM7} class A motifs. Including OR sequences allows to overcome this lack of sequence similarity between TAS2Rs and class A GPCR because ORs show residues which can be more easily compared with the two families. The *Chemosim* alignment was then manually refined using constraints from functional assays in the literature (as described in the results section). We specifically

focused on the 339 class II ORs because they contain relevant motifs for TAS2Rs alignment and because TM sequence conservation is higher than in a mixture of class I and class II human ORs. TM segments were predicted by the PPM webserver [33]. The final *Chemosim* MSA is provided as a supplementary information file (TAS2R-OR-templates.pir).

Template selection for comparative modeling of bitter taste receptors

Class A GPCR templates were selected by submitting each of the 25 human TAS2Rs UniprotKB accession numbers to the Swiss-Model modeling server [34]. From the proposed templates for human TAS2Rs, 46 with at least 10% sequence identity were kept. Templates were then grouped by protein name and sorted by resolution and average sequence identity with TAS2Rs. The highest resolution template from each group was retained, resulting in 19 templates. Finally, six GPCR class A templates were selected to maximize structural diversity. As TAS2Rs have been suggested to be part of the same family as the frizzled receptors [35], three class F GPCR templates were also considered: the human FZD4 receptor [36] and two structures of the human SMO receptor [37]. The PDB code for the six class A templates were as follows: rhodopsin (6FUF) [38], β 1-adrenergic (4BVN) [39], β 2-adrenergic receptor (5JQH) [40], angiotensin II type 1 (4YAY) [41], chemokine receptor CXCR4 (3ODU) [42], serotonin receptor 5-HT_{2C} (6BQG) [43].

Integrative structural modeling of TAS2R

Using the protocols described above (*Chemosim*, *ClustalO*, and *classF*), we built a large number of 3D models. For each alignment and each template, we generated 1000 homology models using Modeller v9.21 [44] with a maximum of 300 conjugate gradient minimization steps and refinement by molecular dynamics with simulated annealing (“md_level” = slow). The remaining parameters were set to default from the “automodel” class. The BitterDB and GPCRdb webserver (labelled BittedDB and GPCRdb in the analysis) provided additional 3D models of each TAS2R. The GOMoDo [45] webserver was also used to automatically generate models of TAS2Rs based only on the sequence (labelled *Gomodo* in the analysis). Default

options were used, excepting the number of models which was set to the maximum (99 models). All the models generated were evaluated and ranked using a meta-score defined as the average of the pocket and helicity score (Fig. 2). This score provides a unique descriptor that accounts for both GPCR structural requirements and TAS2R experimental constraints. Evaluation of the pocket score: the pocket score, ranging from 0 (worst) to 1 (best), depicts how the models agree with site-directed mutagenesis experiments. To identify residues oriented toward the binding pocket, the following protocol was implemented in Python: (i) For each of the 25 human TAS2Rs, a reference 3D model was selected from the *Chemosim* models. All reference models were then structurally aligned to the TAS2R16 reference. (ii) A unique grid of points broadly covering the binding site of class A GPCRs was generated and aligned to the coordinates of the TAS2R16 reference. (iii) Each TAS2R model was aligned to its reference based on the alpha carbons of the TM residues. (iv) Residues whose sidechain center of mass (SCM) was within 8.0 angstroms of any grid point, and whose angle between the SCM, the alpha carbon, and any grid point was lower or equal to 30 degrees, were considered as oriented towards the pocket. Only residues annotated as involved in ligand binding were kept (see supplementary file TAS2R-msa_annotated.xlsx). (v) The pocket score was calculated as the fraction of residues oriented towards the pocket for each TM, averaged across all TMs. 3D structure alignment was performed with MDAnalysis v1.0.0 [46], and distance and angle calculations were performed with scipy v1.5.0 [47] and numpy v1.19.0 [48].

Evaluation of TM helicity score: the helicity score, ranging from 0 to 1 as the pocket score, illustrates how the models agrees with GPCR structural requirements. The Ramachandran number [49] (R) was used to check the structural quality of the TM domains of each model produced. R , which is based on the ϕ and ψ dihedral angles, can be seen as a short numerical form of the Ramachandran plot. First, we analyzed the helicity of 358 class A GPCR X-ray structures to set the experimental range and found an average value of 0.35. Thus, a residue was considered in an alpha-helix conformation if its R value fell between 0.32 and 0.38. To discard misshapen 3D models having severe kinks in the middle of TM domains, we introduced a function based on R . We defined the function $f(r) = \text{count}(|r_i - R_{\text{ref}}| \leq \sigma)$, where r is a moving subset of six consecutive R values that are shifted forward until all R values for a given TM helix have been sampled; $R_{\text{ref}} = 0.35$ is the average R value based on X-ray structures; and $\sigma = 0.07$ is a parameter that was optimized to exclude misfolded TM proteins while keeping X-ray structures. If at any point the result of $f(r)$ was lower than 4 for any TM residue, the model was discarded. A helicity score (H) was then calculated as the fraction of TM residues satisfying

the condition: $H = \text{count}(0.32 \leq R_i \leq 0.38) / \text{length}(R)$. Among all considered X-ray structures, the minimum H value obtained was 0.789. This threshold was used to filter out irrelevant models.

Assessing meta-score accuracy: the meta-score was defined as the average of the pocket and helicity scores. The relevance of the meta-score was assessed by building various homology models of class A and class F GPCR structures from a class A template. The RMSD between the experimental structure of each receptor and the best model according to the meta-score or the scores available in Modeller or the QMEANBrane [50] webserver was then calculated. As shown in Fig. S2, the meta-score performed as well as other metrics when ranking GPCR models and outperformed them when ranking GPCR models based on distantly related GPCR templates.

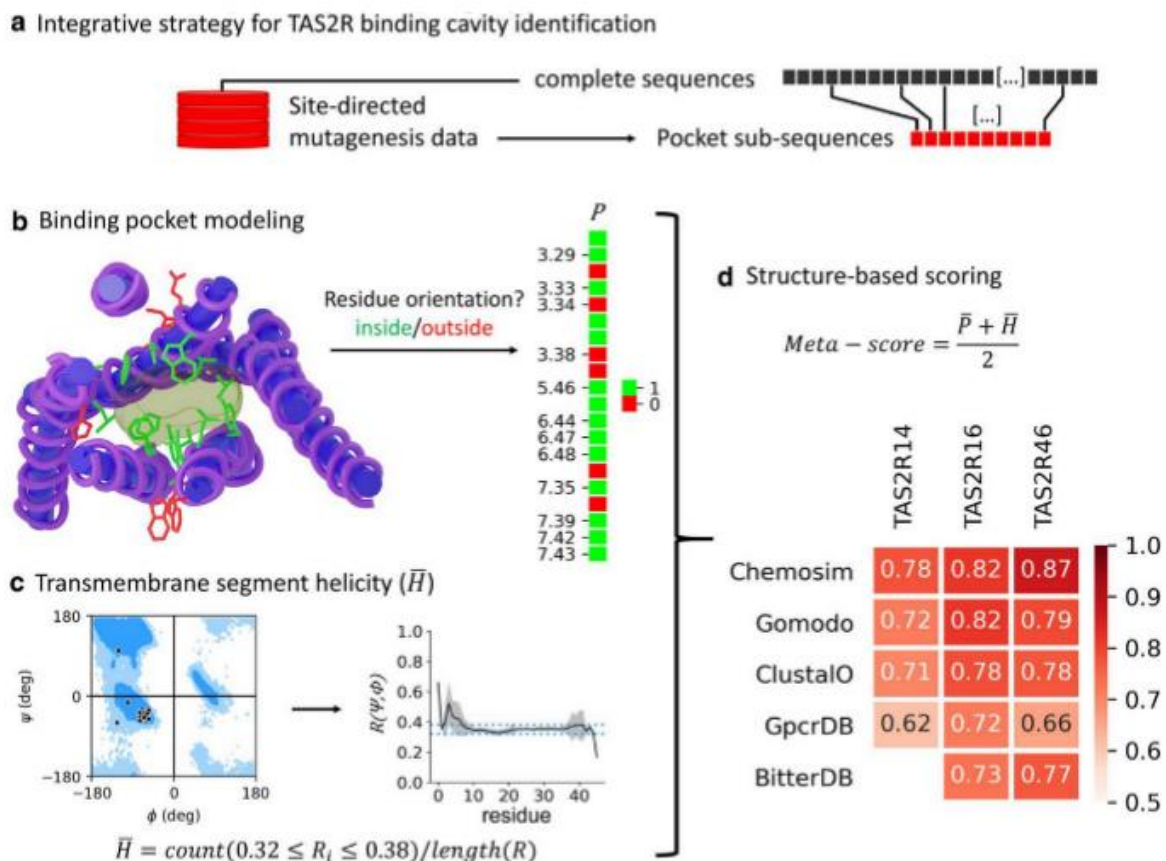


Fig. 2: **a** An integrative approach to identify the TAS2R binding pocket that is used as a constraint in comparative modeling with the Chemosim protocol. **b** A pocket fingerprint was extracted based on the positions of binding residues in the 3D model. The light brown surface represents the binding pocket. **c** The helicity of the TM segment was analyzed and **d** combined with the pocket fingerprint to calculate a structure-based normalized meta-score. The meta-scores of the best 3D models of TAS2R14, 16 and 46 structures generated by the different comparative modeling protocols are shown in panel d.

Cell culture and transfection

Plasmids encoding TAS2R16 and G16αgust44 were constructed as previously described [51]. G16αgust44 and TAS2R16 were cloned into a CMV promoter-based vector and expressed constitutively. Point mutations on the TAS2R16 clone were obtained from a commercial service (Macro- gen Inc., Seoul, Republic of Korea), which also performed DNA sequencings of the mutant genes. The TAS2R16 and G16αgust44 expression plasmids were co-transfected (4:1) into HEK293T cells using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA). Cellular responses were measured 18–24 h after transfection. Cells were cultured at 37 °C in a humidified atmosphere of 5% CO₂. The culture medium was Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal bovine serum, 100 IU/ml penicillin G, 100 µg/ml streptomycin, 2 mM L-glutamine, and 1 mM sodium pyruvate (Invitrogen).

Quantitative measurement of intracellular Ca²⁺ in bitter taste receptors upon stimulation with salicin

The compound-induced changes in cytosolic Ca²⁺ concentrations were measured using a FlexStation III microplate reader (Molecular Devices, Sunnyvale, CA, USA). Cells transfected with TAS2R16 were seeded onto 96-well blackwall CellBind surface plates (Corning, NY, USA). After 18–24 h seeding, the cells were washed with assay buffer (130 mM NaCl, 10 mM glucose, 5 mM KCl, 2 mM CaCl₂, 1.2 mM MgCl₂, and 100 mM HEPES; pH 7.4) and incubated in the dark, first at 37 °C for 30 min, and then at 27 °C for 15 min in assay buffer consisting of Calcium-4 (FLIPR Calcium 4 Assay Kit, Molecular Devices). After the samples were treated, the cell fluorescence intensity (excitation, 486 nm; emission, 525 nm) was measured. The results were plotted with $\Delta F/F^\circ$ on the y-axis, where ΔF is the change in Calcium-4 fluorescence intensity at each time point, and F° is the initial fluorescence intensity. The responses from at least three separate experiments ($n = 3$) with the same stimulus were averaged.

Results and discussion

Matching conserved motifs between Class A GPCRs and TAS2Rs

Since their discovery in 2000, TAS2Rs have been extensively studied. An important variety of methodologies have been employed to understand structure–function relationship of TAS2R. The sequence identity between TAS2R and class A GPCR is low (below 30%) and the identification of conserved motifs in TM segments remains challenging. [28, 52] Most of the studies then rely on functional experiments and point mutations combined with molecular modeling of the receptors. [53] They focus on the identification of residues involved in ligand recognition and in the activation mechanism of receptors TAS2R1, [21] 4, [54–58] 7, [30, 59] 9, [60] 10, [61] 14, [58, 62, 63] 16, [19, 64–66] 20, [58] 31, [20] 38, [67, 68] 43 [20] and 46 [20, 24]. These studies have led to differences in the published alignments, specifically in the alignment of TM3, TM4, TM6, and TM7 (Table S2). Hence, the prediction of TAS2Rs tertiary structure based on sequence similarity, together with the identification of molecular switches is still an open issue.

Previously, we have shown that refining the sequence alignment of ORs with non-olfactory class A GPCRs by including site-directed mutagenesis produces relevant three dimensional models of chemosensory receptors. These models have been supported by experimental data [16, 18, 69, 70]. We thus apply a similar integrative strategy to TAS2Rs. To overcome the lack of sequence similarity between TAS2Rs and GPCRs with known structures in TM3, TM6 and TM7, we inserted 339 human class II OR sequences in the alignment. A similar systematic bioinformatic analysis has successfully identified known residues involved in ligand binding. [71]

Subsequent manual data curation involved integration of site-directed mutagenesis data from the literature for 136 amino-acids positions, i.e., 45% of the entire TAS2Rs sequence (see ESI TAS2R-msa_annotated.xlsx). Our alignment (Fig. S1) highlights the key residues and consensus motifs in all human TAS2Rs, which correspond to the functional molecular switches in ORs and non-olfactory class A GPCRs (Fig. 1b, d). They are detailed above and summarized in Table 1.

TM1 and 2 contain highly conserved residues facilitating their alignment and resulting in a consensus in the alignment. [52] In TM1, the NGFI^{TM1-TAS2R} motif corresponds to GNLLI^{TM1-OR} in OR and GNxLV^{TM1-classA} in non-olfactory GPCR templates (see Fig. S1). In TM2, R^{2.50-TAS2R} in the LAxSR^{TM2-TAS2R} motif aligns with D^{2.50-OR/classA}, which in class A GPCRs constitutes a sodium ion binding site that stabilizes inactive receptor conformations [72]. Position 2.50 in TAS2Rs is positively charged and unlikely to be involved in sodium binding. Moreover, it has been shown that position 1.50 and 2.50 are involved in downstream signaling by stabilizing the structure of TAS2Rs [21].

The sequence alignment of TM4 was not straightforward, as it lacks the canonical W 4.50-OR/classA. No consensus has emerged from previously published alignments (Table S2), and we chose to align the highly conserved leucine L 4.50 of the LLG TM4-TAS2R motif with the most conserved W^{4.50-OR/classA}. TM3, 5, 6, and 7 contained functional molecular switches which have been identified in class A GPCR experimental structures [14]. While the alignment of TM5 and 7 is now commonly accepted, the alignment of TM3 and 6 is much more complicated as suggested by the different published alignments (Table S2).

In TM3, K 3.50 in the KIANFS^{TM3-TAS2R} motif matches R^{3.50} of the DRY^{TM3-classA} and MAYDRYVAIC TM3-OR motifs. The DRY motif constitutes the ionic lock in ORs and non-olfactory class A GPCRs. This also aligns the highly conserved L^{3.43}, with a leucine found at position 3.43 in both non-olfactory class A GPCRs and OR (Table 1). In TM5, similarly to previously published alignments, the conserved P^{5.50} of the PF^{TM5-TAS2R} motif corresponds to the PF^{TM5-OR} and P^{TM5-classA} motifs/residue involved in the so-called “hydrophobic connector” (P^{5.50}I^{3.40}F^{6.44} in class A GPCRs). Another conserved aromatic residue that is found in 52% of TAS2Rs, F^{5.58}, consistently aligns with the conserved Y^{5.58} known to be important for GPCR activation [18, 73].

In TM6, the HxKALKT^{TM6-TAS2R} motif matches both a comparable motif in non-olfactory class A GPCRs and the typical OR motif RxKAFST^{TM6-OR}. The “toggle/transmission switch” (CW^{6.48}LP classA and FY^{6.48}G^{OR}) responsible for downstream signaling aligns with the YF^{6.48}L motif in TAS2Rs. The inclusion of ORs sequences in the alignment helps the identification of this motif involving two consecutive aromatic residues in TAS2Rs, as it is the case in OR sequences. The alignment of this YF^{6.48}L motif is consistent with site-directed mutagenesis results, suggesting a role of agonist-sensing, as for class A GPCRs [16, 74].

The extracellular part of TM7 is well-documented to belong to the ligand binding pocket in TAS2Rs and other GPCRs [20, 24, 74]. This is consistent with its high sequence variability (see

Fig. S1). TM7 intracellular residues show higher conservation, as they are involved in GPCR signaling [16, 74]. These conserved motifs, however, show little similarity between TAS2Rs and other GPCRs. Here, the comparison with ORs is highly instructive: from the $P^{7.46}xLNP^{7.50}xIY^{TM7-OR}$ motif found in ORs, $P^{7.46}$ is shared with TAS2Rs, and $NP^{7.50}xxY$ is found in other class A GPCRs. $P^{7.46}$ and $P^{7.50}$ are conserved in 76% and 28% of human TAS2Rs, respectively. The $PxxHSFIL^{TM7-TAS2R}$ motif is consequently aligned with $PxLNPxIY^{TM7-OR}$, which itself matches the highly conserved $xxxNPxxY^{TM7-classA}$ motif [20].

Predicted tertiary structure of TAS2Rs

We tested various protocols (based on different alignments described in methods) and structural templates to build accurate 3D homology models of TAS2Rs. Among the TAS2Rs, receptors TAS2R14, 16, and 46 were selected to evaluate the approach, as previous works on these receptors involving site-directed mutagenesis provide data to determine the residues within their binding pocket. According to our meta-score, the best models of these three receptors were obtained using the Chemosim approach and a single template, either the β 2-adrenoceptor (PDB 5JQH) or the β 1-adrenoceptor (PDB 4BVN) structure (Fig. 2 and S3). The performance of each protocol is compared in Fig. S3 and S4. *Gomodo* and *ClustalO* approaches led to comparable models, with slight improvement over *BitterDB* and, in most cases, substantial improvement over *GPCRdb*. The use of class F templates systematically led to models with misfolded helices (Fig. S4).

These models and analysis were then extrapolated to the full human TAS2Rs repertoire. Even if limited experimental data are available, we were able to define a consensus TAS2R cavity based on the positions identified simultaneously in TAS2R14, 16 and 46. We also extended the definition of a specific TAS2R cavity to residues identified by site-directed mutagenesis. The best models for the entire TAS2R family were obtained using GPCR templates in their closed conformation (Fig. S6), with the exception of TAS2R38, for which the open-conformation 5-HT_{2C} receptor (PDB 6BQG) was best. On average, the templates 5JQH, 4BVN all of which correspond to adrenergic receptors, performed best. In this study, we found no relationship between the performance of the protocols and the percentage sequence identity of the templates used to build the models. At 10–15%, the sequence identity between TAS2Rs and class A templates is too low to be a discriminating criterion. The best *Chemosim* model obtained for each human TAS2R is provided as a PDB file in the supplementary information.

Projecting TAS2Rs sequence conservation onto the 3D structure showed that the models retain the structural characteristics of the GPCR (Fig. S5). The most conserved residues were located in the intracellular region of the receptor that binds the G protein, while the greatest variability was found in the extracellular ligand-binding pocket. Analysis of the binding cavity (Fig. S7) revealed high diversity within the hTAS2Rs family. The pocket volume ranged up to 400 Å³ and 700 Å³ for hTAS2R13 and hTAS2R39, respectively, corresponding to the structural features of a GPCR [75]. Although no obvious structure–function relationship was revealed by the analysis of the cavity volume, the hydrophobicity partially correlated with the receptor range of response. The binding cavities of TAS2Rs with broad ligand spectrums tended to be more hydrophobic than those of narrow-spectrum receptors (Fig. S7), consistent with previous studies showing a correlation between hydrophobicity and GPCR promiscuity [52, 76].

Based on our model, we selected three positions, 90^{3.34}, 91^{3.35}, and 185^{5.47}, from the binding site and three positions, 42^{ICL1}, 43^{ICL1} and 100^{3.44}, predicted to be far from the binding pocket which served as negative controls. Their functional role was evaluated by site-directed mutagenesis followed by *in vitro* functional assays with salicin (Fig. 3 and Table S3). All mutants showed a specific, dose-dependent response to salicin *in vitro* (Fig. 3), confirming that they are expressed and functional at the cell surface.

The TAS2R16 I90A/S^{3.34}, L91A/S^{3.35}, and L185H^{5.47} mutants showed a reduced response to salicin *in vitro*, consistent with their orientation toward the interior of the receptor bundle (Fig. 3 and Table S4). Positions 3.35 and 5.47 have been previously reported to directly interact with ligands [26, 30, 62].

The L42^{ICL1}A/S, M43^{ICL1}A, and T100^{3.44}A mutations served as negative controls (Table S3) and did not statistically affect salicin potency *in vitro* (Fig. 3 and Table S4). Only mutation of position 43 to a serine induced a weak decrease of salicin-dependent response in TAS2R16 compared to WT.

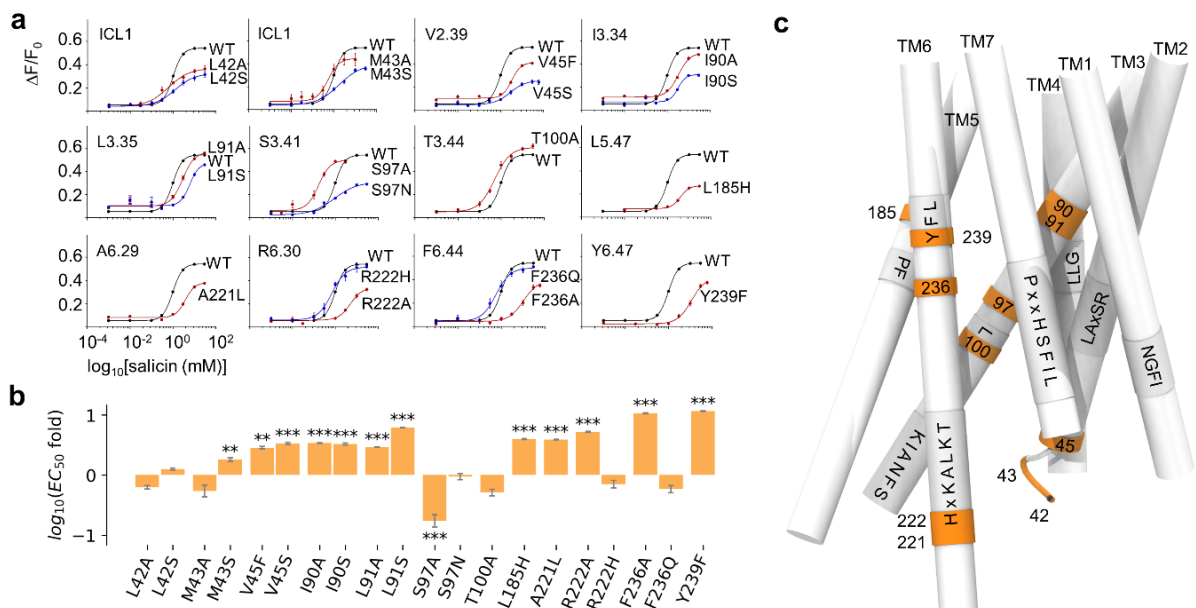


Fig. 3: **a** *In vitro* functional assays of wild-type (WT) TAS2R16 and single-point mutants stimulated by salicin. **b** EC_{50} fold (compared to WT) expressed as $\log(\text{EC}_{50}(\text{MUT})/\text{EC}_{50}(\text{WT}))$ for the twenty TAS2R16 mutants considered in this study. Positive values indicate a reduced response to salicin in the mutated receptor compared to the WT. *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$ versus the WT group (one-way ANOVA followed by Dunnett's test). **c** Representative structure of TAS2R16 highlighting the location of the mutated residues. The TM domains are presented as sticks. The positions of mutated residues are colored in orange, and the molecular switches revealed by the sequence alignment are indicated on the structure.

Evaluating the function role of molecular switches

To evaluate the functional role of the predicted molecular switches, twelve residue positions on TAS2R16 were subjected to site-directed mutagenesis followed by *in vitro* functional assays with salicin (Fig. 3 and Table S2). The residues mostly belonged to TM3 and TM6, which, in GPCRs, are well-known to be involved in agonist sensing and activation [14].

Using our model as a basis, we investigated residues found in the ligand binding pocket (90^{3.35}, 91^{3.36}, and 185^{5.47}) and at or around the predicted molecular switches (45^{2.39}, 97^{3.41}, 221^{6.29}, 222^{6.30}, 236^{6.44}, and 239^{6.47}). Residues 42^{ICL1}, 43^{ICL1}, and 100^{3.44} were predicted to be far from the molecular switches. All mutants showed a specific, dose-dependent response to salicin (Fig. 3), confirming that they are expressed and functional at the cell surface.

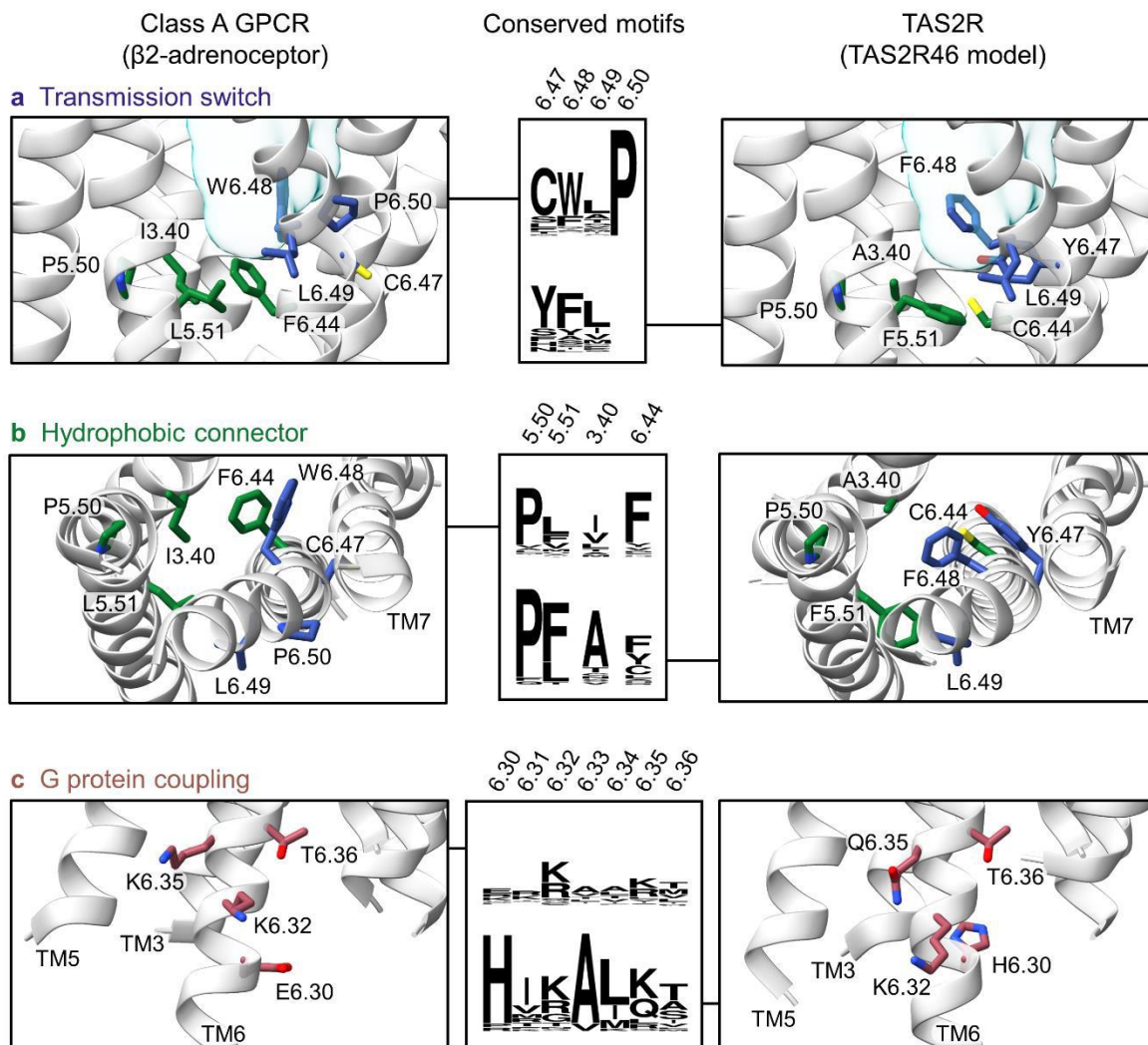


Fig. 4: Sequence logos and molecular details of conserved motifs involved in the activation mechanism of class A GPCRs and TAS2Rs, i.e. **a** the transmission switch (colored in blue), **b** the hydrophobic connector (in green), and **c** the G protein-coupling region (in red). The binding pocket is depicted as a pale blue surface. The structure of the β 2-adrenoceptor is taken from PDB code 5JQH.

In GPCRs, the residue 6.48 is defined as the toggle switch and is well known to be involved in agonist sensing and activation [14]. Position 239^{6.47} is conserved as Y (64%) and F (8%) in human TAS2Rs (Fig. 4a). In mammals, an aromatic residue (F, Y or H) is also found in 85% of the sequences. Conservation of an aromatic residue also occurs in ORs [16]. The Y239F^{6.47} mutation decreased the potency of salicin by a factor of 11, confirming its importance in receptor activation (Fig. 3). Position Y239^{6.47} corresponded to Y239 and Y241 in TAS2R10 and TAS2R46, respectively. For both of these receptors, the tyrosine to phenylalanine mutation is reported to lead to a significant reduction in ligand responsiveness [20, 61]. Born *et al.* also

observed a complete loss of response to agonists with the Y239A^{6.47} TAS2R10 construction [61]. Further, we found that the introduction of an alanine at this position eliminated any response to salicin (data not shown).

Adjacent to Y239^{6.47}, residue F240^{6.48} is conserved as aromatic in 72% of human TAS2Rs and in 67% of mammalian TAS2Rs. As the toggle-switch residue, its nature and function in agonist sensing is similar in ORs (conserved as F^{6.48}) [16] and non-olfactory GPCRs (conserved as W^{6.48}) [14]. F240^{6.48} has previously been reported to affect TAS2R16 agonist response. Sakurai et al. showed that mutation of F240^{6.48} to a leucine residue in TAS2R16 drastically alters the function of the receptor, while mutation to aromatic residues (Y and W) leads to moderate changes in the EC50 [19]. Further, the potencies of various other agonists were affected in the same manner in vitro, highlighting the critical role this residue plays in signaling initiation, as it is the case for numerous class A GPCRs [14–16].

Altogether, these observations suggest the functional equivalence of the Y^{6.47}FLx motif in TAS2Rs with the F^{6.47}YGx in ORs [16] and the C^{6.47}WLP [14] in non-olfactory class A GPCRs [9]. This motif is particularly important as it forms part of the cradle of the binding pocket and senses the presence of agonists [74].

The hydrophobic connector molecular switch involved in class A GPCRs activation [15] was conserved as P^{5.50}I^{3.40}F^{6.44} [14, 15, 17]. Similarly to other TAS2Rs, a P^{5.50}A^{3.40}F^{6.44} motif (Fig. 4b) was located at the core of TAS2R16, close to the cradle of the binding pocket. In class A GPCRs, this motif, together with NPxxY^{TM7}, holds a central role in receptor signaling, ligand-independent constitutive activation, and β -arrestin signaling in the β 2-adrenoceptor [17]. It is plausible that this motif has similar functions in TAS2Rs [77], as suggested by the modulated response to salicin we found in our mutants (Fig. 3). F236^{6.44}, conserved in 75% of mammalian TAS2Rs as Y/F (Fig. 4b), is predicted to be part of the hydrophobic connector molecular switch. The F236A^{6.44} TAS2R16 mutant consistently showed a significantly weaker response to salicin, while no difference in response was found for the F236Q^{6.44} mutant. In a previous study, Thomas et al. found that a F236Y^{6.44} mutation prevented agonist-dependent signaling. [66] In TAS2R14, an alanine residue occupies position 6.44, and mutation to a leucine leads to a decrease in receptor sensitivity to numerous ligands. [62]

Adjacent to position 3.40, S97^{3.41} does not belong to the binding pocket and points toward the membrane. In accordance with a previous report showing its importance for TAS2R16 trafficking [26], the S97A^{3.41} mutation altered receptor response (gain of function).

Our model predicted that V45^{2.39} is part of a hydrophobic cluster in the intracellular part of TM2 and is conserved as a hydrophobic residue in 72% of TAS2Rs. This hydrophobic area occurs near the highly conserved L229^{7.53} (96% and 93% in humans and mammals, respectively) and the HSFIL^{TM7} motifs and likely forms part of the hydrophobic barrier that prevents flooding of the intracellular region. Mutating V45^{2.39} into a hydrophilic residue (S) strongly altered salicin activation both in this work and in the literature [26]; substitution with a bulkier hydrophobic residue (F) was better tolerated.

In TM6, position 6.29 and adjacent residues have been documented to control G protein selectivity in class A GPCRs [78]. A221^{6.29} and H222^{6.30} are conserved in 60% and 92% of human TAS2Rs, respectively, and in 70% and 94% of mammalian TAS2Rs (Fig. 4c). Position 222^{6.30} is an arginine in TAS2R16. Salicin induced reduced responses in the A221L^{6.29} and R222A^{6.30} mutants, whereas the response of the R222H^{6.30} mutant was not statistically different from the WT. In TAS2R4, the H233A^{6.30} mutation inhibited the response to quinine [55]. Altogether, these findings suggest the need for a positive charge at position 6.30 and that this region could be involved in G protein-coupling.

Conclusions

This study brings attention to potential key residues and consensus functional motifs of bitter taste receptors (TAS2Rs) using a combination of bioinformatics, molecular modeling, and in vitro assays. In particular, we performed sequence alignment of human TAS2Rs with olfactory and non-olfactory class A GPCRs, including residue conservation and experimental data as constraints. We propose a consensus alignment of TAS2Rs which recapitulates key results from previous studies. The consensus sequence motifs match well-known ones in class A GPCRs.

Using site-directed mutagenesis, we then evaluated the functional roles of these motifs in TAS2R16 as a case study. In addition to the residues lining the binding pocket, we identified plausible candidate for the “toggle/transmission switch” (the YF^{6.48}L motif in TM6) and the “hydrophobic connector” (P^{5.50}A^{3.40}F^{6.44}) for agonist sensing. Other molecular switches were proposed in the intracellular regions of TM6 and TM7. In class A GPCR, these residues have been shown to be involved in G protein selectivity suggesting that this conserved position could have the same function in TAS2Rs. These molecular switches extend to the whole mammalian TAS2R repertoire (see supplementary files). The approach, templates, and 3D model provided

in this study serve as a foundation for rational design of specific TAS2Rs agonists and antagonists, and for decoding sequence-structure–function relationships in these receptors.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s00018-021-03968-7>.

Acknowledgements

The authors thank Dr. Xiaojing Cong for fruitful discussion and critical review of the manuscript. This work was funded by the French Ministry of Higher Education and Research [PhD Fellowship to CB], by the National Research Foundation of Korea (NRF) [grant number NRF2020R1A2C2004661], by GIRACT (Geneva, Switzerland) [9th European PhD in Flavor Research Bursaries for first year students to CB], and the Gen Foundation (Registered UK Charity No. 1071026), a charitable trust that primarily funds research in natural sciences, particularly food sciences/technology [grant to CB and JT]. We also benefited from funding by the French government through the UCAJEDI “Investments in the Future” project, managed by the ANR [grant No. ANR-15-IDEX-01 to SF and JG]. Computation for the work described in this paper was supported by the Université Côte d’Azur’s Center for High-Performance Computing.

Author contributions

JT[†], CB[†], and **JP** performed numerical modeling. YK and MR conducted functional assays. JT, CB, and SF performed data curation. JT, CB, **JP**, YK, and MR conducted formal analyses. JT, SF, and JG supervised and managed the study and wrote the paper. MR and JG provided resources for this study.

Code and data availability

The scripts used to generate and analyze the models as well as PDB files of TAS2Rs 3D models with the highest meta-score have been deposited on GitHub. (https://github.com/chemosim-lab/TAS2R_data).

Declarations

Conflict of interest: The authors declare no competing financial interest.

References

1. Lindemann B (2001) Receptors and transduction in taste. *Nature* 413(6852):219–225
2. Meyerhof W (2005) Elucidation of mammalian bitter taste. In: *Reviews of physiology, biochemistry and pharmacology*. Springer, Berlin, Heidelberg, pp 37–72
3. Mueller KL, Hoon MA, Erlenbach I, Chandrashekar J, Zuker CS, Ryba NJ (2005) The receptors and coding logic for bitter taste. *Nature* 434(7030):225–229
4. Lee S-J, Depoortere I, Hatt H (2019) Therapeutic potential of ectopic olfactory and taste receptors. *Nat Rev Drug Discovery* 18(2):116–138
5. Foster S, Blank K, See Hoe L, Behrens M, Meyerhof W, Peart J, Thomas W (2014) Novel bitter taste receptor ligands elicit G protein-dependent negative inotropic effects in mouse heart (LB572). *FASEB J*. https://doi.org/10.1096/fasebj.28.1_supplement.lb572
6. Malki A, Fiedler J, Fricke K, Ballweg I, Pfaffl MW, Krautwurst D (2015) Class I odorant receptors, TAS1R and TAS2R taste receptors, are markers for subpopulations of circulating leukocytes. *J Leukoc Biol* 97(3):533–545
7. Adler E, Hoon MA, Mueller KL, Chandrashekar J, Ryba NJ, Zuker CS (2000) A novel family of mammalian taste receptors. *Cell* 100(6):693–702
8. Fredriksson R, Lagerström MC, Lundin L-G, Schiöth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* 63(6):1256–1272

9. Nordström KJ, Sällman Almén M, Edstam MM, Fredriksson R, Schiöth HB (2011) Independent HHsearch, Needleman–Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families. *Mol Biol Evol* 28(9):2471–2480
10. Krishnan A, Almén MS, Fredriksson R, Schiöth HB (2012) The origin of GPCRs: identification of mammalian like Rhodopsin, Adhesion, Glutamate and Frizzled GPCRs in fungi. *PLoS ONE* 7(1):9817
11. Di Pizio A, Levit A, Slutzki M, Behrens M, Karaman R, Niv MY (2016) Comparing Class A GPCRs to bitter taste receptors: structural motifs, ligand interactions and agonist-to-antagonist ratios. *Methods Cell Biol* 132:401–427
12. Cvicek V, Goddard WA III, Abrol R (2016) Structure-based sequence alignment of the transmembrane domains of all human GPCRs: phylogenetic, structural and functional implications. *PLoS Comput Biol* 12(3):e1004805
13. Munk C, Isberg V, Mordalski S, Harpsøe K, Rataj K, Hauser A, Kolb P, Bojarski A, Vriend G, Gloriam D (2016) GPCRdb: the G protein-coupled receptor database—an introduction. *Br J Pharmacol* 173(14):2195–2207
14. Deupi X, Standfuss J (2011) Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr Opin Struct Biol* 21(4):541–551
15. Zhou Q, Yang D, Wu M, Guo Y, Guo W, Zhong L, Cai X, Dai A, Jang W, Shakhnovich EI (2019) Common activation mechanism of class A GPCRs. *Elife*. <https://doi.org/10.7554/eLife.50279>
16. de March CA, Yu Y, Ni MJ, Adipietro KA, Matsunami H, Ma M, Golebiowski J (2015) Conserved residues control activation of mammalian G protein-coupled odorant receptors. *J Am Chem Soc* 137(26):8611–8616
17. Schönege A-M, Gallion J, Picard L-P, Wilkins AD, Le Gouill C, Audet M, Stallaert W, Lohse MJ, Kimmel M, Lichtarge O (2017) Evolutionary action and structural basis of the allosteric switch controlling β 2 AR functional selectivity. *Nat Commun* 8(1):1–12
18. de March CA, Topin J, Bruguera E, Novikov G, Ikegami K, Matsunami H, Golebiowski J (2018) Odorant receptor 7D4 activation dynamics. *Angew Chem* 130(17):4644–4648
19. Sakurai T, Misaka T, Ishiguro M, Masuda K, Sugawara T, Ito K, Kobayashi T, Matsuo S, Ishimaru Y, Asakura T (2010) Characterization of the β -d-glucopyranoside binding site of the human bitter taste receptor hTAS2R16. *J Biol Chem* 285(36):28373–28378

20. Brockhoff A, Behrens M, Niv MY, Meyerhof W (2010) Structural requirements of bitter taste receptor activation. *Proc Natl Acad Sci* 107(24):11110–11115
21. Singh N, Pydi SP, Upadhyaya J, Chelikani P (2011) Structural basis of activation of bitter taste receptor T2R1 and comparison with Class A G-protein-coupled receptors (GPCRs). *J Biol Chem* 286(41):36032–36041
22. Karaman R, Nowak S, Di Pizio A, Kitaneh H, Abu-Jaish A, Meyerhof W, Niv MY, Behrens M (2016) Probing the binding pocket of the broadly tuned human bitter taste receptor TAS2R14 by chemical modification of cognate agonists. *Chem Biol Drug Des* 88(1):66–75
23. Fierro F, Giorgetti A, Carloni P, Meyerhof W, Alfonso-Prieto M (2019) Dual binding mode of “bitter sugars” to their human bitter taste receptor target. *Sci Rep* 9(1):1–16
24. Sandal M, Behrens M, Brockhoff A, Musiani F, Giorgetti A, Carloni P, Meyerhof W (2015) Evidence for a transient additional ligand binding site in the TAS2R46 bitter taste receptor. *J Chem Theory Comput* 11(9):4439–4449
25. Di Pizio A, Kruetzfeldt L-M, Cheled-Shoval S, Meyerhof W, Behrens M, Niv MY (2017) Ligand binding modes from low resolution GPCR models and mutagenesis: chicken bitter taste receptor as a test-case. *Sci Rep* 7(1):1–11
26. Thomas A, Sulli C, Davidson E, Berdough E, Phillips M, Puffer BA, Paes C, Doranz BJ, Rucker JB (2017) The Bitter Taste Receptor TAS2R16 Achieves High Specificity and Accommodates Diverse Glycoside Ligands by using a Two-faced Binding Pocket. *Sci Rep* 7(1):7753
27. Biarnés X, Marchiori A, Giorgetti A, Lanzara C, Gasparini P, Carloni P, Born S, Brockhoff A, Behrens M, W, (2010) Meyerhof, Insights into the binding of Phenyltiocarbamide (PTC) agonist to its target human TAS2R38 bitter receptor. *PLoS ONE* 5(8):e12394
28. Prasad Pydi S, Upadhyaya J, Singh N, Pal Bhullar R, Chelikani P (2012) Recent advances in structure and function studies on human bitter taste receptors. *Curr Prot Pept Sci* 13(6):501–508
29. Slack JP, Brockhoff A, Batram C, Menzel S, Sonnabend C, Born S, Galindo MM, Kohl S, Thalmann S, Ostopovici-Halip L (2010) Modulation of bitter taste perception by a small molecule hTAS2R antagonist. *Curr Biol* 20(12):1104–1109
30. Wang Y, Zajac AL, Lei W, Christensen CM, Margolskee RF, Bouysset C, Golebiowski J, Zhao H, Fiorucci S, Jiang P (2019) Metal ions activate the human taste receptor TAS2R7. *Chem Senses* 44(5):339–347

31. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7(1):539
32. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191
33. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res* 40(D1):D370–D376
34. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46(W1):W296–W303
35. Fredriksson R, Schiöth HB (2005) The repertoire of G-protein– coupled receptors in fully sequenced genomes. *Mol Pharmacol* 67(5):1414–1425
36. Yang S, Wu Y, Xu T-H, de Waal PW, He Y, Pu M, Chen Y, DeBruine ZJ, Zhang B, Zaidi SA (2018) Crystal structure of the Frizzled 4 receptor in a ligand-free state. *Nature* 560(7720):666–670
37. Zhang X, Zhao F, Wu Y, Yang J, Han GW, Zhao S, Ishchenko A, Ye L, Lin X, Ding K (2017) Crystal structure of a multi-domain human smoothed receptor in complex with a super stabilizing ligand. *Nat Commun* 8(1):1–10
38. Tsai C-J, Pamula F, Nehmé R, Mühle J, Weinert T, Flock T, Nogly P, Edwards PC, Carpenter B, Gruhl T (2018) Crystal structure of rhodopsin in complex with a mini-Go sheds light on the principles of G protein selectivity. *Sci Adv* 4(9):eaat7052
39. Miller-Gallacher JL, Nehme R, Warne T, Edwards PC, Schertler GF, Leslie AG, Tate CG (2014) The 2.1 Å resolution structure of cyanopindolol-bound β 1-adrenoceptor identifies an intramembrane Na⁺ ion that stabilises the ligand-free receptor. *PLoS ONE* 9(3):e92727
40. Staus DP, Strachan RT, Manglik A, Pani B, Kahsai AW, Kim TH, Wingler LM, Ahn S, Chatterjee A, Masoudi A (2016) Allosteric nanobodies reveal the dynamic range and diverse mechanisms of G-protein-coupled receptor activation. *Nature* 535(7612):448–452
41. Zhang H, Unal H, Gati C, Han GW, Liu W, Zatsepin NA, James D, Wang D, Nelson G, Weierstall U (2015) Structure of the angiotensin receptor revealed by serial femtosecond crystallography. *Cell* 161(4):833–844

42. Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330(6007):1066–1071
43. Peng Y, McCorvy JD, Harpsøe K, Lansu K, Yuan S, Popov P, Qu L, Pu M, Che T, Nikolajsen LF (2018) 5-HT_{2C} receptor structures reveal the structural basis of GPCR polypharmacology. *Cell* 172(4):719-730.e14
44. Webb B, Sali A (2016) Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinform* 54(1):5.6.1-5.6.37
45. Sandal M, Duy TP, Cona M, Zung H, Carloni P, Musiani F, Giorgetti A (2013) GOMoDo: a GPCRs online modeling and docking webserver. *PLoS ONE* 8(9):e74092
46. Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, Domanski J, Dotson DL, Buchoux S, Kenney IM (2019) MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. Los Alamos National Lab(LANL), Los Alamos
47. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Method* 17(3):261–272
48. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ (2020) Array programming with NumPy. *Nature* 585(7825):357–362
49. Mannige RV, Kundu J, Whitelam S (2016) The Ramachandran number: an order parameter for protein geometry. *PLoS ONE* 11(8):e0160023
50. Studer G, Biasini M, Schwede T (2014) Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics* 30(17):i505–i511
51. Ueda T, Ugawa S, Yamamura H, Imaizumi Y, Shimada S (2003) Functional interaction between T2R taste receptors and G-protein α subunits expressed in taste receptor cells. *J Neurosci* 23(19):7376–7380
52. Di Pizio A, Levit A, Slutzki M, et al (2016) Comparing Class A GPCRs to bitter taste receptors. In: *Methods in Cell Biology*. Elsevier Ltd, pp 401–427.
53. Behrens M, Ziegler F (2020) Structure-function analyses of human bitter taste receptors—where do we stand? *Molecules* 25(19):4423

54. Pydi SP, Bhullar RP, Chelikani P (2012) Constitutively active mutant gives novel insights into the mechanism of bitter taste receptor activation. *J Neurochem* 122(3):537–544
55. Pydi SP, Singh N, Upadhyaya J, Bhullar RP, Chelikani P (2014) The third intracellular loop plays a critical role in bitter taste receptor activation. *Biochimica et Biophysica Acta (BBA) Biomembr* 1838(1):231–236
56. Pydi SP, Sobotkiewicz T, Billakanti R, Bhullar RP, Loewen MC, Chelikani P (2014) Amino acid derivatives as bitter taste receptor (T2R) blockers. *J Biol Chem* 289(36):25054–25066
57. Upadhyaya J, Singh N, Bhullar RP, Chelikani P (2015) The structure–function role of C-terminus in human bitter taste receptor T2R4 signaling. *Biochimica et Biophysica Acta (BBA) Biomembr* 1848(7):1502–1508
58. Jaggupilli A, Singh N, De Jesus VC, Gounni MS, Dhanaraj P, Chelikani P (2019) Chemosensory bitter taste receptors (T2Rs) are activated by multiple antibiotics. *FASEB J* 33(1):501–517
59. Liu K, Jaggupilli A, Premnath D, Chelikani P (2018) Plasticity of the ligand binding pocket in the bitter taste receptor T2R7. *Biochimica et Biophysica Acta (BBA) Biomembr* 1860(5):991–999
60. Dotson CD, Zhang L, Xu H, Shin Y-K, Vignes S, Ott SH, Elson AE, Choi HJ, Shaw H, Egan JM (2008) Bitter taste receptors influence glucose homeostasis. *PLoS ONE* 3(12):e3974
61. Born S, Levit A, Niv MY, Meyerhof W, Behrens M (2013) The human bitter taste receptor TAS2R10 is tailored to accommodate numerous diverse ligands. *J Neurosci* 33(1):201–213
62. Nowak S, Di Pizio A, Levit A, Niv MY, Meyerhof W, Behrens M (2018) Reengineering the ligand sensitivity of the broadly tuned human bitter taste receptor TAS2R14. *Biochimica et Biophysica Acta (BBA) Gen Subj* 1862(10):2162–2173
63. Shaik FA, Jaggupilli A, Chelikani P (2019) Highly conserved intracellular H208 residue influences agonist selectivity in bitter taste receptor T2R14. *Biochimica et Biophysica Acta (BBA) Biomembr* 1861(12):183057
64. Soranzo N, Bufe B, Sabeti PC, Wilson JF, Weale ME, Marguerie R, Meyerhof W, Goldstein DB (2005) Positive selection on a high-sensitivity allele of the human bitter-taste receptor TAS2R16. *Curr Biol* 15(14):1257–1265
65. Greene TA, Alarcon S, Thomas A, Berdougou E, Doranz BJ, Breslin PA, Rucker JB (2011) Probenecid inhibits the human bitter taste receptor TAS2R16 and suppresses bitter perception of salicin. *PLoS ONE* 6(5):e20123

66. Thomas A, Sulli C, Davidson E, Berdougou E, Phillips M, Puffer BA, Paes C, Doranz BJ, Rucker JB (2017) The bitter taste receptor TAS2R16 achieves high specificity and accommodates diverse glycoside ligands by using a two-faced binding pocket. *Sci Rep* 7(1):1–15
67. Biarnés X, Marchiori A, Giorgetti A, Lanzara C, Gasparini P, Carloni P, Born S, Brockhoff A, Behrens M, Meyerhof W (2010) Insights into the binding of Phenyltiocarbamide (PTC) agonist to its target human TAS2R38 bitter receptor. *PLoS ONE* 5(8):e12394
68. Marchiori A, Capece L, Giorgetti A, Gasparini P, Behrens M, Carloni P, Meyerhof W (2013) Coarse-grained/molecular mechanics of the TAS2R38 bitter taste receptor: experimentally-validated detailed structural prediction of agonist binding. *PLoS ONE* 8(5):e64675
69. Charlier L, Topin J, Ronin C, Kim S-K, Goddard WA, Efremov R, Golebiowski J (2012) How broadly tuned olfactory receptors equally recognize their agonists. Human OR1G1 as a test case. *Cell Mol Life Sci* 69(24):4205–4213
70. Bushdid C, Claire A, Topin J, Do M, Matsunami H, Golebiowski J (2019) Mammalian class I odorant receptors exhibit a conserved vestibular-binding pocket. *Cell Mol Life Sci* 76(5):995–1004
71. Fierro F, Suku E, Alfonso-Prieto M, Giorgetti A, Cichon S, Carloni P (2017) Agonist binding to chemosensory receptors: a systematic bioinformatics analysis. *Front Mol Biosci* 4:63
72. Katritch V, Fenalti G, Abola EE, Roth BL, Cherezov V, Stevens RC (2014) Allosteric sodium in class A GPCR signaling. *Trends Biochem Sci* 39(5):233–244
73. Miao Y, Nichols SE, Gasper PM, Metzger VT, McCammon JA (2013) Activation and dynamic network of the M2 muscarinic receptor. *Proc Natl Acad Sci* 110(27):10982–10987
74. Venkatakrishnan A, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM (2013) Molecular signatures of G-protein-coupled receptors. *Nature* 494(7436):185–194
75. Dalton JA, Lans I, Giraldo J (2015) Quantifying conformational changes in GPCRs: glimpse of a common functional mechanism. *BMC Bioinformatics* 16(1):1–15
76. Levit A, Beuming T, Krilov G, Sherman W, Niv MY (2014) Predicting GPCR promiscuity using binding site features. *J Chem Inf Model* 54(1):184–194
77. Kim D, Cho S, Castaño MA, Panettieri RA, Woo JA, Liggett SB (2019) Biased TAS2R bronchodilators inhibit airway smooth muscle growth by downregulating phosphorylated extracellular signal-regulated kinase 1/2. *Am J Respir Cell Mol Biol* 60(5):532–540

78. Flock T, Hauser AS, Lund N, Gloriam DE, Balaji S, Babu MM (2017) Selectivity determinants of GPCR–G-protein binding. Nature 545(7654):317–322

Supplementary Information

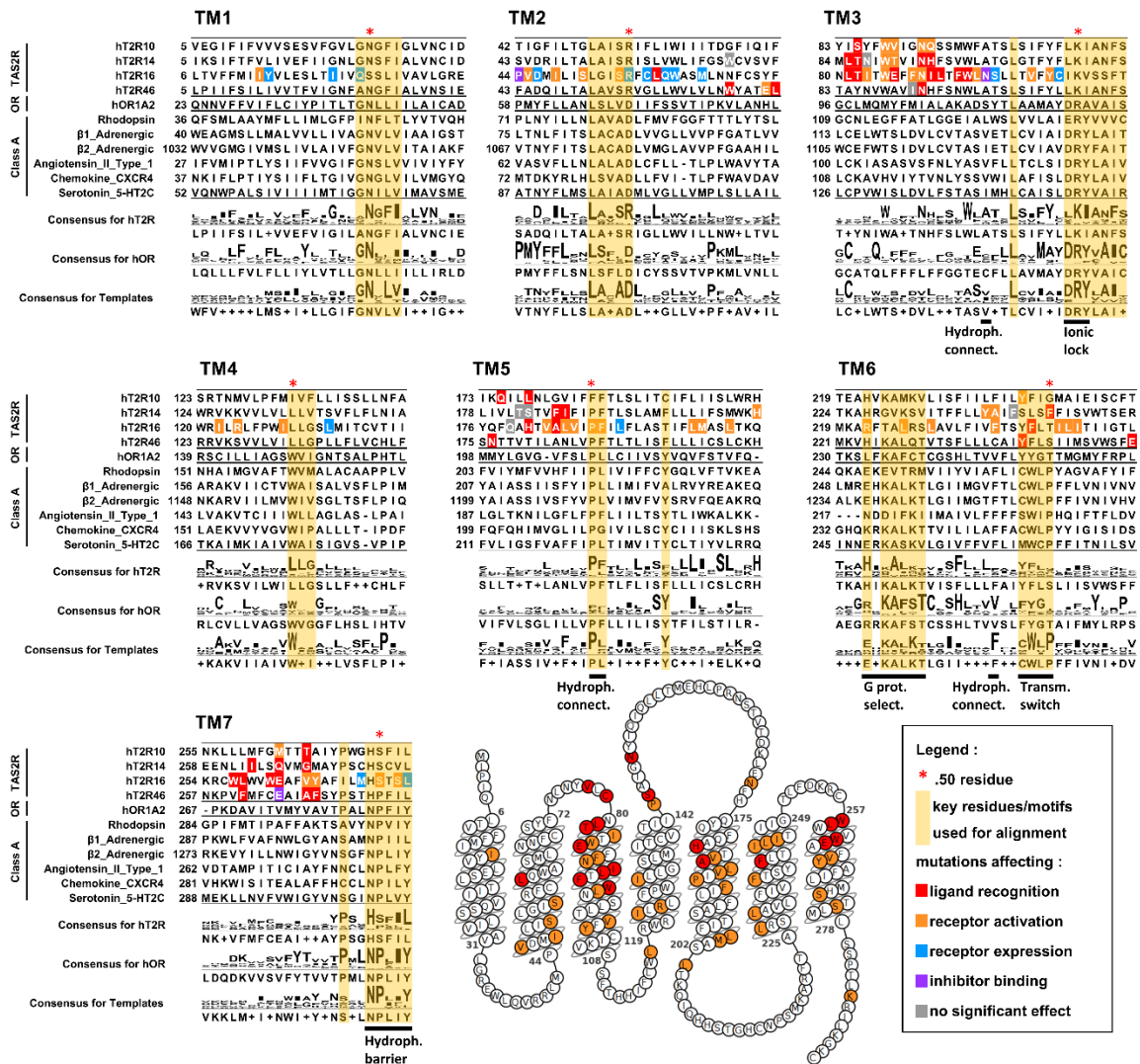


Figure S1: Alignment and results from site directed mutagenesis.

Selected human type 2 taste receptors (TAS2R), human Olfactory Receptors (OR) and non-OR class A G protein-coupled receptors (GPCRs). In the sequence alignment, residues are colored by their roles as reported in the literature (see legend). The aligned motifs are highlighted in yellow. Consensus sequences for TAS2Rs, ORs and Templates contained 25 human TAS2Rs, 339 human class II ORs, and 6 class A GPCRs, respectively. Functional molecular switches (transmission switch, hydrophobic connector, ionic lock, and hydrophobic barrier) and residues involved in G-protein coupling are indicated under the alignment.

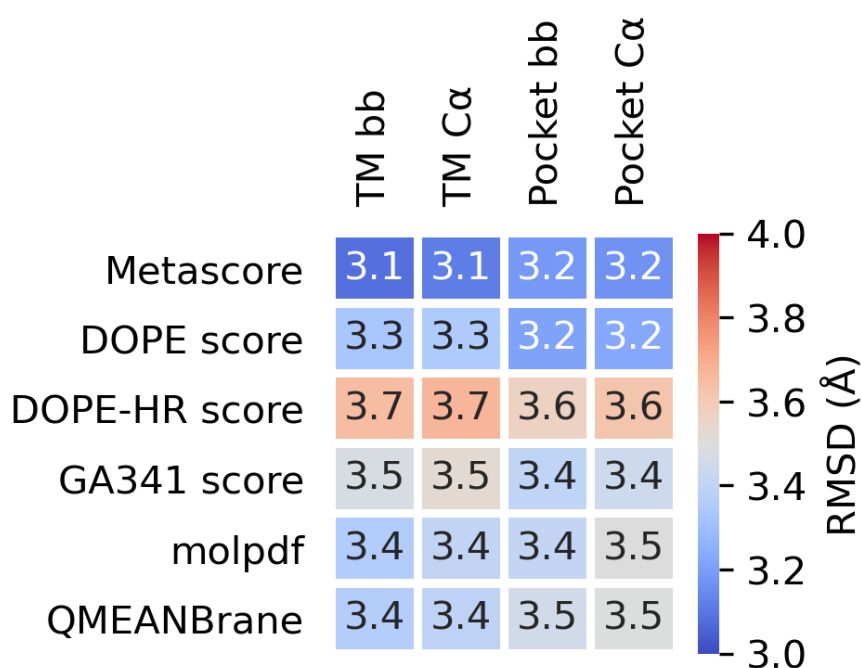


Figure S2: RMSD of class F models built using a class A template

The human smoothed receptor (class F) models were built by homology modeling with a class A template (β 2-adrenoceptor, PDB 5JQH [1]). The sequence alignment was taken from the GPCRdb [2] and manually refined with UCSF Chimera's structure-based sequence alignment tool (v1.14) [3] based on the 5JQH template and a structure of the smoothed receptor (PDB 4JKV [4]). The same Modeller [5] protocol detailed in the manuscript was used to generate 1000 models of the smoothed receptor. The models were structurally aligned to the 4JKV reference based on the trans membrane (TM) domains and ranked by their meta-scores. Finally, the RMSDs between the reference and each best model were calculated based on the TM domain backbone (TM bb), the TM alpha carbons (TM Ca), the pocket residue backbone (Pocket bb), and the pocket residue alpha carbons (Pocket Ca). The pocket residues were identified by visual inspection of four class F X-ray structures in complex with a ligand (PDB codes 6O3C [6], 4JKV [4], 4QIM [7], and 4N4W [7]).

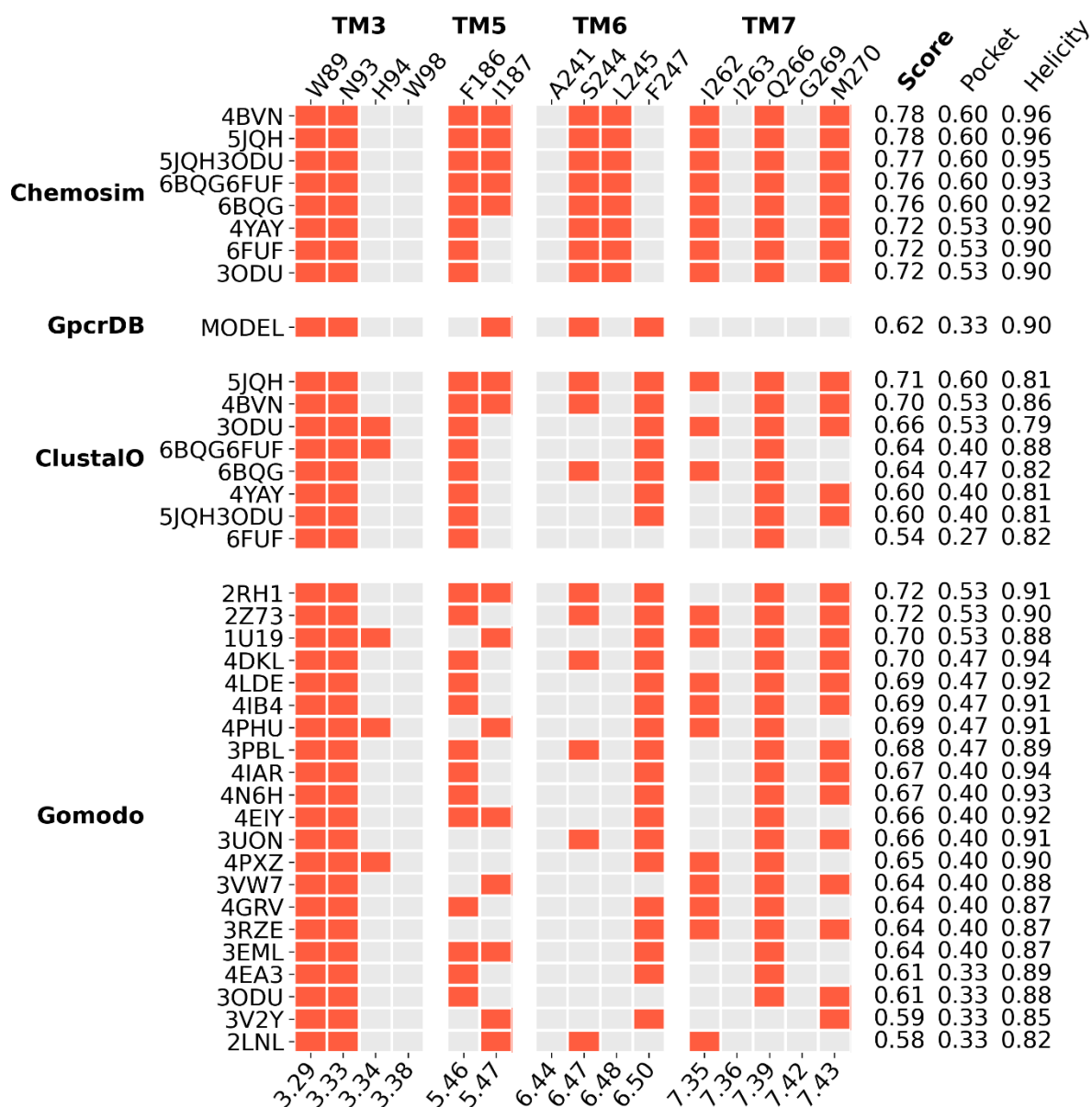


Figure S3.a: Detailed analysis of TAS2R14 binding pocket residues

Meta-scores of top models for each protocol and template. Best models following the Gomodo [8] and ClustalO [9] protocols were selected based on their DOPE score [10]. For BitterDB [11], the only available model did not satisfy our structure quality criteria. The x-axis labels correspond to the Ballesteros-Weinstein numbering of each residue [12]. The left y-axis provides the PDB code of each template except for GPCRdb, where the model was retrieved directly from their website. The right y-axis shows the meta-score, pocket score, and helicity score for each selected model.

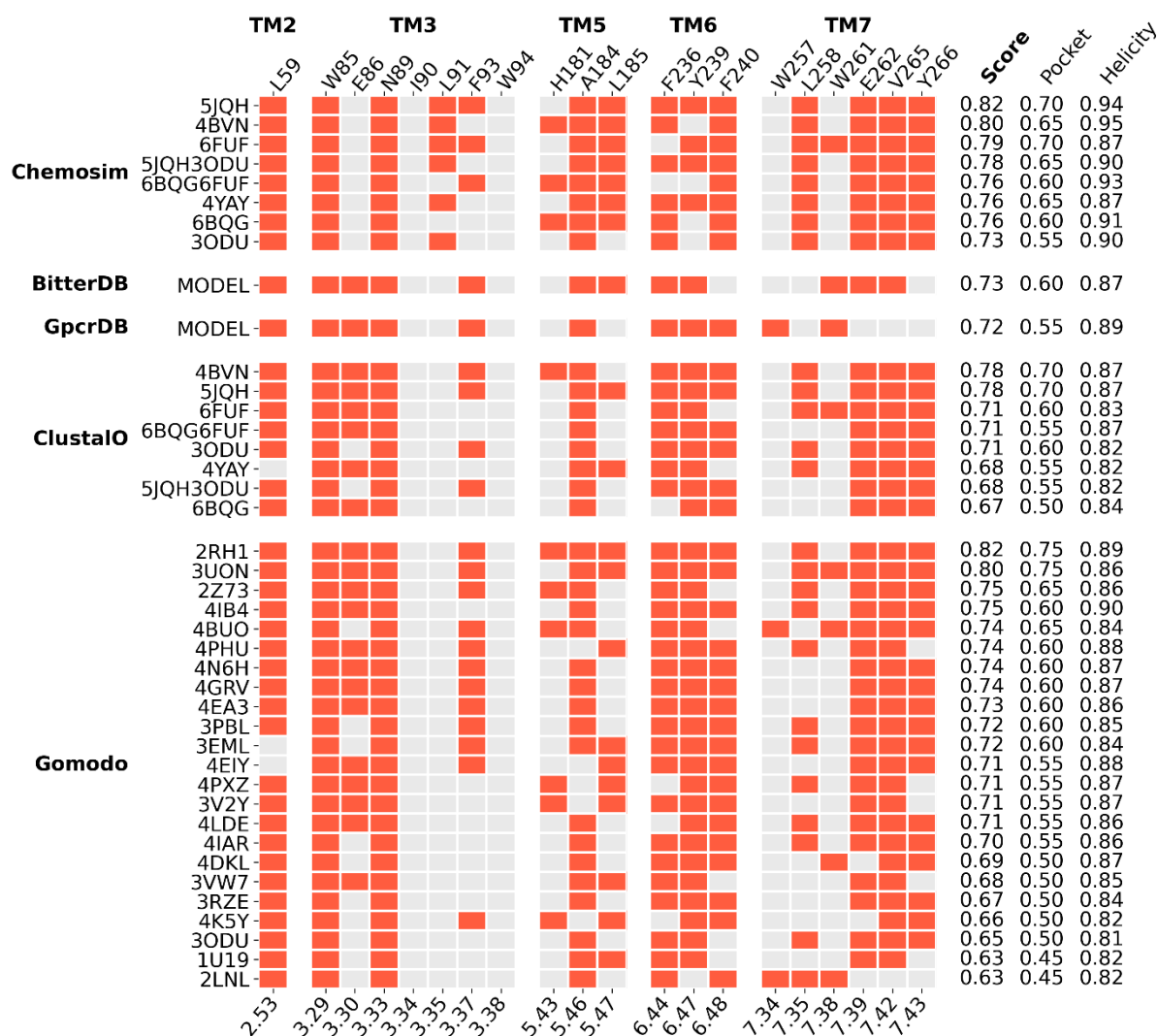


Figure S3.b: Detailed analysis of TAS2R16 binding pocket residues

Meta-scores of top models for each protocol and template. Best models following the Gomodo and ClustalO protocols were selected based on their DOPE score. The x-axis labels correspond to the Ballesteros-Weinstein numbering of each residue. The left y-axis provides the PDB code of each template except for BitterDB and GPCRdb, where the model was retrieved directly from their website. The right y-axis shows the meta-score, pocket score, and helicity score for each selected model.

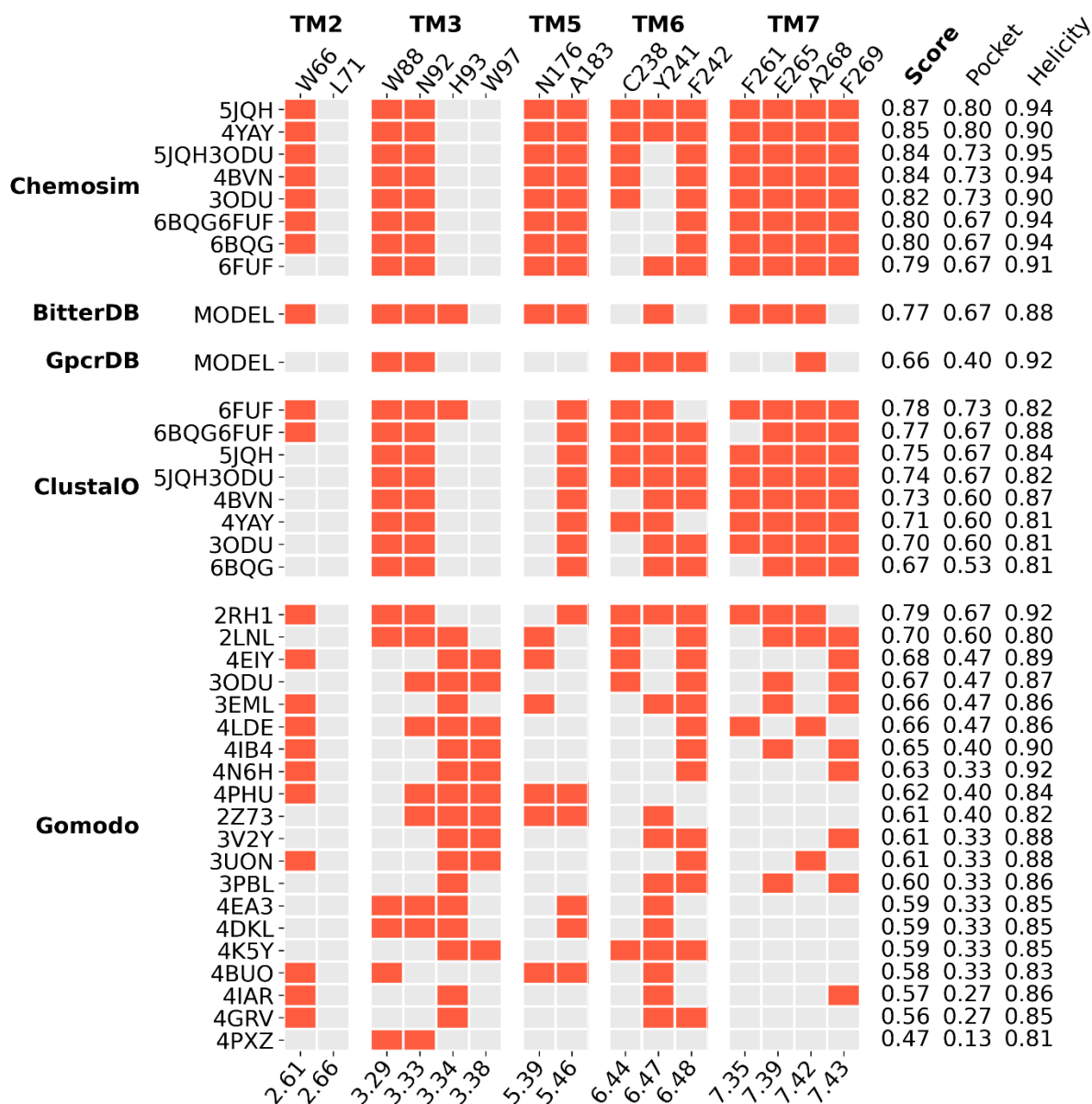


Figure S3.c: Detailed analysis of TAS2R46 binding pocket residues

Meta-scores of top models for each protocol and template. Best models following the Gomodo and ClustalO protocols were selected based on their DOPE score. The x-axis labels correspond to the Ballesteros-Weinstein numbering of each residue. The left y-axis provides the PDB code of each template except for BitterDB and GPCRdb, where the model was retrieved directly from their website. The right y-axis shows the meta-score, pocket score, and helicity score for each selected model.



Figure S4.a: Analysis of TAS2R14 transmembrane helicity

Ramachandran number (R) plot of each residue, numbered by their Ballesteros-Weinstein (BW) position, for the models produced by the best template for each protocol. Standard deviation is represented by the shaded area, and the green zone corresponds to R values typically found in alpha helices of crystallographic GPCR structures (0.32 to 0.38).

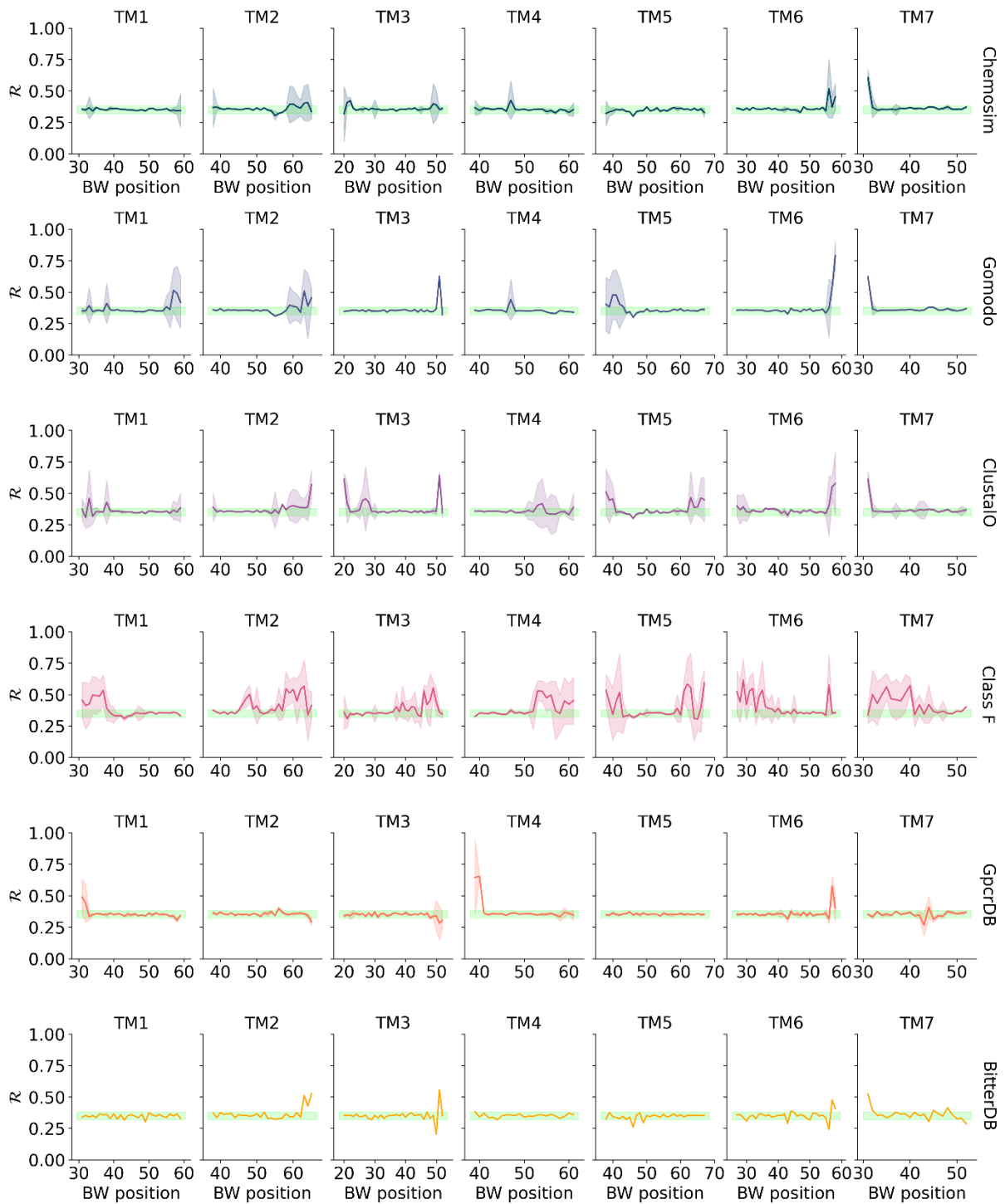


Figure S4.b: Analysis of TAS2R16 transmembrane helicity

See figure caption S4.a.



Figure S4.c: Analysis of TAS2R46 transmembrane helicity

See figure caption S4.a.

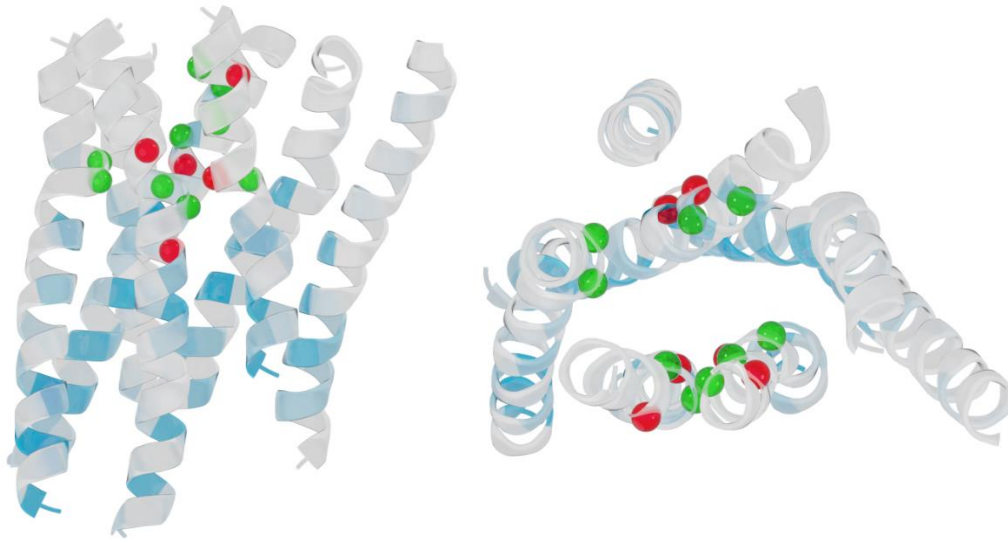


Figure S5.a: Structure of the TAS2R14 model with the highest meta-score

Structure of the best *Chemosim* model obtained from the present study. The residues defining the binding pocket are shown as spheres if their side chains are oriented outward (red) or inward (green) from the pocket and follow from the results shown in Fig S3. Positions of the highly conserved residues in the human TAS2R family are indicated by a color scale, from 50% or less conservation (white) to 100% (blue).

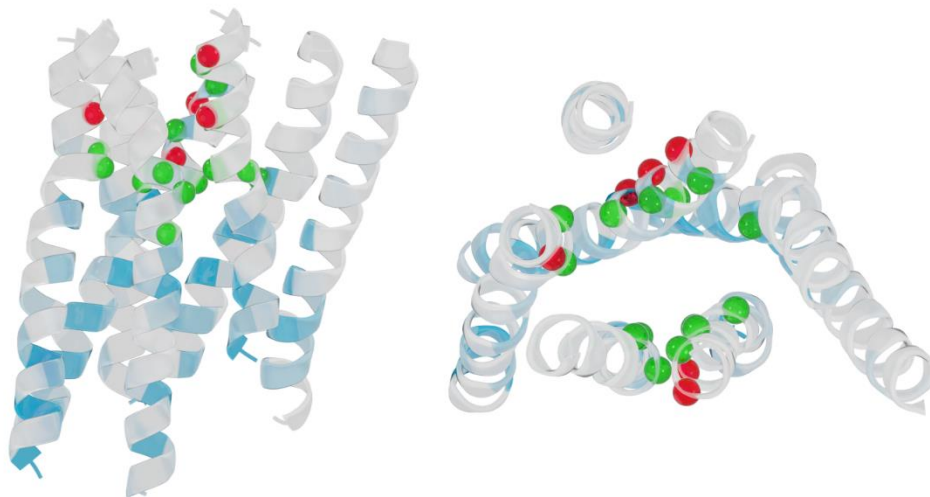


Figure S5.b: Structure of the TAS2R16 model with the highest meta score. See figure caption S5.a.

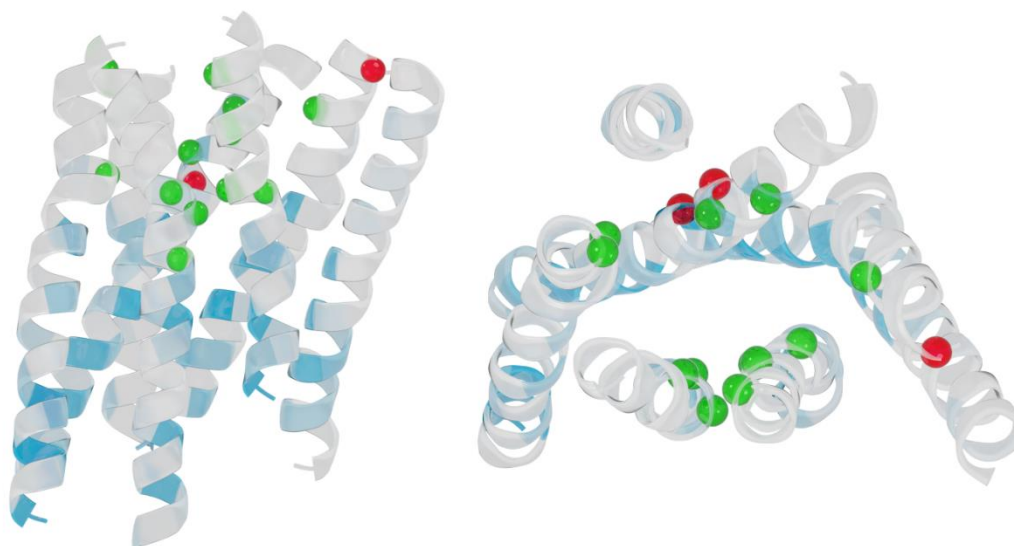


Figure S5.c: Structure of the TAS2R46 model with the highest meta score

See figure caption S5.a.

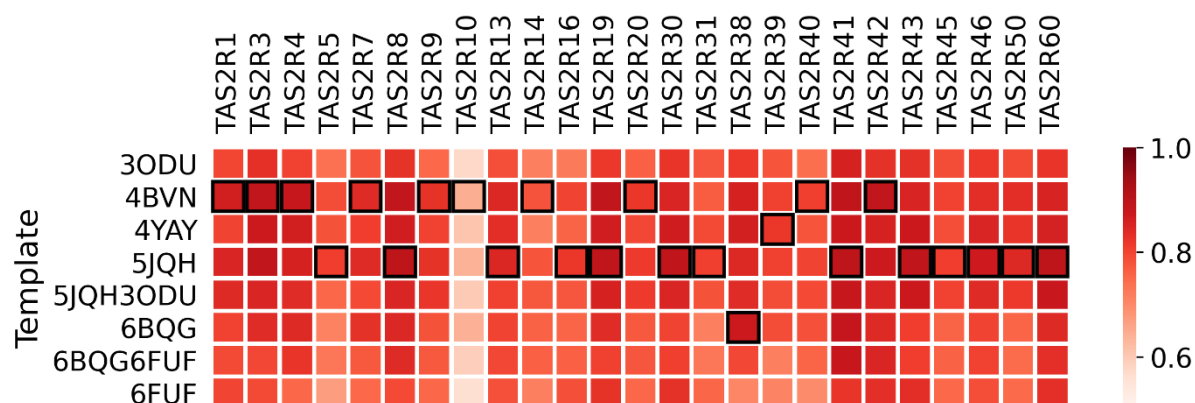


Figure S6: Selection of TAS2R models according to various class A templates

TAS2R models were built following the *Chemosim* protocol. The best models are shown with black boxes and were selected according to the highest meta-score. For all receptors, a consensus TAS2R cavity was used for the detection of residues oriented in the binding pocket. This consensus cavity was composed of residues 3.29, 3.33, 3.34, 3.38, 5.46, 6.44, 6.47, 6.48, 7.35, 7.39, 7.42, and 7.43 and was completed by receptor-specific cavity residues highlighted in the annotated TAS2Rs MSA that is provided in the supplementary files (TAS2R-msa-annotated.xlsx).

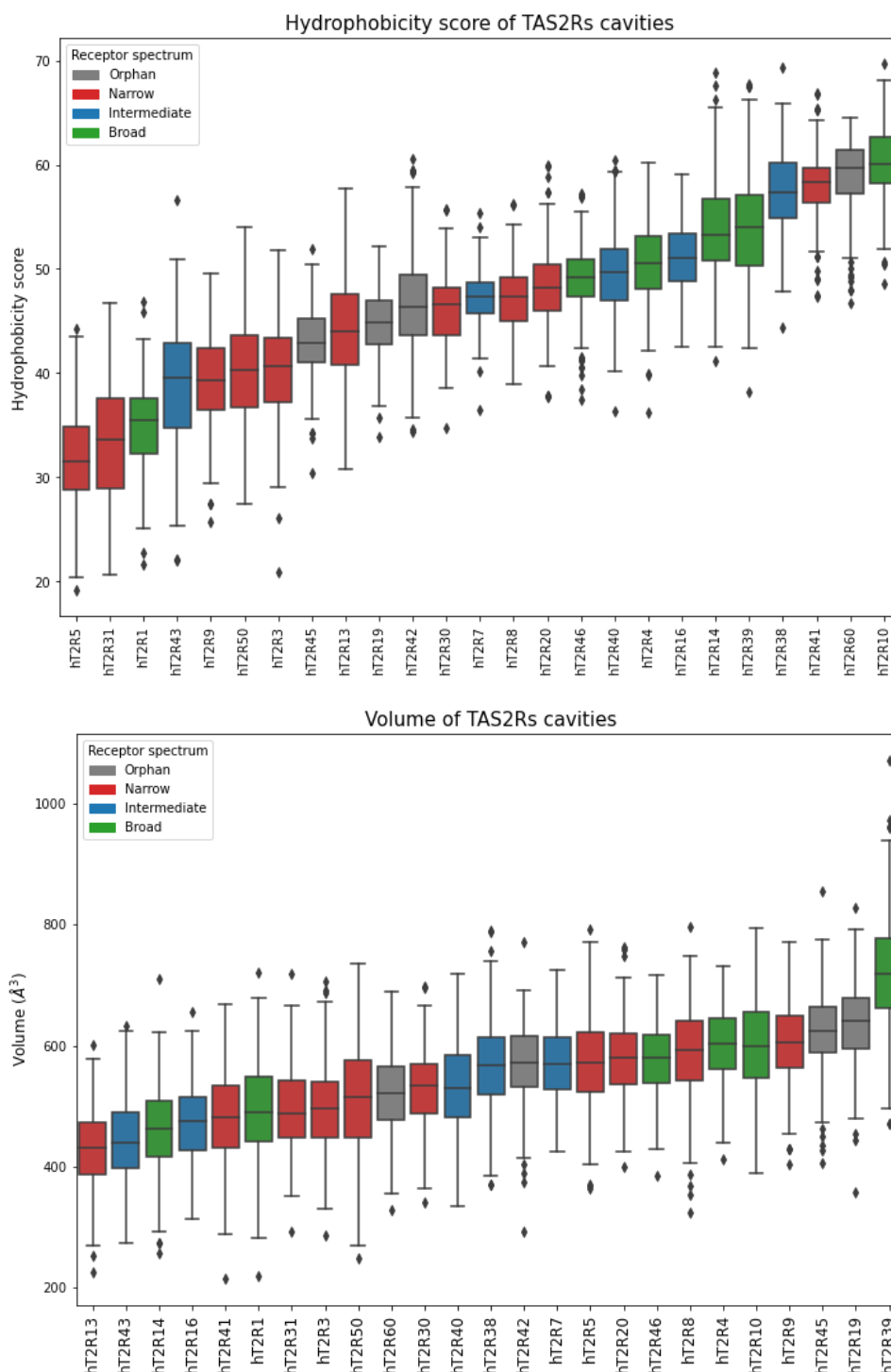


Figure S7: Structural analysis of TAS2R binding pocket

Box-plot of hydrophobicity and volume of TAS2Rs binding pocket. The box extends from the lower to upper quartile values of the data, with a line at the median and outliers plotted in diamonds. The top 250 models for each TAS2R produced by the *Chemosim* protocol and selected templates as shown in Figure S6 were analyzed by MDpocket [13] and colored according to the receptive range (broad, intermediate/specific, narrow, and orphan receptors in green, blue, red, and grey, respectively). A positive hydrophobicity score means that the cavity is mainly hydrophobic.

Table S1: Summary of the most conserved TAS2R amino acids. The most conserved TAS2R residues (above 80% sequence identity) and those involved in TAS2R hallmarks (in yellow/bold) used for multiple sequence alignment with OR and class A templates.

	ClassA motif	OR motif	TAS2R Motif	TAS2R Consensus	Conservation	BW numbering	TAS2R14	TAS2R16	TAS2R46
TM1	GNxLV	GNLLI	NGFI	G	88%	1.46	G20	I21	G20
				N	92%	1.50	N24	S25	N24
				G	72%	1.51	S25	S26	G25
				F	84%	1.52	F26	L27	F26
				I	92%	1.53	I27	I28	I27
ICL1			W	80%		W35	W36	W35	
TM2	LAXAD	LSxxD	LAXSR	D	84%	2.40	D45	D46	D45
				I	84%	2.42	I47	I48	I47
				L	80%	2.43	L48	L49	L48
				L	100%	2.46	L51	L52	L51
				A	64%	2.47	A52	G53	A52
				S	84%	2.48	S54	S55	S54
				R	96%	2.49	R55	R56	R55
L	92%	2.53	L58	L59	L58				
TM3	DRY	MAYDRYVAIC	KIANFS	W	84%	3.29	W89	W85	W88
				N	84%	3.33	N93	N89	N92
				W	100%	3.38	W98	W94	W97
				L	96%	3.43	L103	L99	L102
				F	80%	3.46	F106	F102	F105
				Y	92%	3.47	Y107	Y103	Y106
				K	92%	3.50	K110	K106	K109
				I	88%	3.51	I111	V107	I110
				A	76%	3.52	A112	S108	A111
				N	64%	3.53	N113	S109	N112
				F	84%	3.54	F114	F110	F113
S	64%	3.55	S115	T111	S114				
ICL2			F	88%		F119	F115	F118	
			L	88%		L122	L118	L121	
TM4	W	W	LLG	K	84%	4.39	K123	R119	K122
				L	88%	4.50	L134	L130	L133
				L	80%	4.51	L135	L131	L134
				G	72%	4.52	V136	G132	G135
ECL2				N	100%		N162	N163	N161
				T	96%		T164	T165	T163
TM5	P	PF	PF	P	92%	5.50	P190	P188	P187
				F	72%	5.51	F191	F189	F188
	Y	Y	F	L	80%	5.55	L195	L193	L192
				F	52%	5.58	F198	T196	F195
				L	100%	5.61	L201	L199	L198
				S	100%	5.64	L204	S202	L201
L	96%	5.65	M205	L203	L202				
ICL3				H	96%		H208	Q206	H205
				G	84%		I218	G213	I215
				D	84%		D221	N216	D218
				P	80%		A222	P217	P219
	TM6	KxxK	RxKAFSTC	HxKALKT	H	92%	6.30	H227	R222
K/R					60%	6.32	G229	T224	K226
A					92%	6.33	V230	A225	A227
L					64%	6.34	K231	L226	L228
K/Q					88%	6.35	S232	R227	Q229
T/S		60%	6.36	V233	S228	T230			
F		96%	6.40	F237	L232	F234			
L		80%	6.43	Y240	V235	L237			
CWLP		FYG	YFL	Y	64%	6.47	S244	Y239	Y241
				F	60%	6.48	L245	F240	F242
	L/I/V			76%	6.49	S246	L241	L243	
	P			76%	7.46	P273	I269	P272	
	H			96%	7.49	H276	H272	H275	
TM7	NPxxY	PxxNPxIY	PxxHSFIL	S	68%	7.50	S277	S273	P276
				F	60%	7.51	C278	T274	F277
				I	76%	7.52	V279	S275	I278
				L	96%	7.53	L280	L276	L279
				I	92%	7.54	I281	M277	I280
				N	80%		N284	S280	N283
L	96%		L287	L283	L286				

Table S2: Mutations tested *in vitro* to assess the 3D model

Mutations	TAS2R motifs	Location/role
I90A/S ^{3.35} , L91A/S ^{3.36} , L185A ^{5.47}	n.a.	inside pocket
T100A ^{3.44}	Negative control	outside pocket
S97A/N ^{3.41}	n.a.	receptor surface/ receptor trafficking
F236A/Q ^{6.44}	P ^{5.50} A ^{3.40} F ^{6.44}	pocket cradle/ hydrophobic connector, agonist sensing
Y239F ^{6.47}	YF ^{6.48} L	pocket cradle/ transmission switch, agonist sensing
V45S/F ^{2.39}	Next to PxxHS ^{7.50} FIL	intracellular part/ hydrophobic barrier
L42A/S ^{ICL1} , M43A/S ^{ICL1}	Negative control	intracellular part
A221L ^{6.29} , R222A/H ^{6.30}	HxK ^{6.32} ALKT	G protein binding site/ G protein selectivity

Table S3: Salicin-induced *in vitro* response in wild-type and mutant TAS2R16

	Mutations	EC ₅₀ [†] (mM)	Maximal Response (ΔF/F ₀)
WT		0.98 ± 0.01	0.55
I90	I90A	3.34 ± 0.03 ^{***}	0.50
	I90S	3.20 ± 0.11 ^{***}	0.31
L91	L91A	2.85 ± 0.03 ^{***}	0.57
	L91S	6.05 ± 0.03 ^{***}	0.47
L42	L42A	0.61 ± 0.04	0.37
	L42S	1.23 ± 0.05	0.33
M43	M43A	0.53 ± 0.12	0.45
	M43S	1.77 ± 0.13 ^{**}	0.40
V45	V45S	3.30 ± 0.12 ^{***}	0.41
	V45F	2.79 ± 0.12 ^{**}	0.26
S97	S97A	0.17 ± 0.04 ^{***}	0.50
	S97N	0.92 ± 0.11	0.31
T100	T100A	0.50 ± 0.06	0.61
L185	L185H	3.87 ± 0.05 ^{***}	0.27
A221	A221L	3.78 ± 0.04 ^{***}	0.38
R222	R222A	5.10 ± 0.08 ^{***}	0.34
	R222H	0.69 ± 0.10	0.52
F236	F236A	10.38 ± 0.11 ^{***}	0.39
	F236Q	0.57 ± 0.08	0.50
Y239	Y239F	11.30 ± 0.08 ^{***}	0.42

[†] Values are means ± SEM; Statistical significance is indicated by *** $P < 0.001$, ** $P < 0.01$, and * $P < 0.05$ vs. the WT group (one-way ANOVA followed by Dunnett's test)

Other supplementary information files

The MSA of human TAS2Rs and a selection of ORs and class A templates (**TAS2R-OR-templates.pir**); the MSA of reviewed mammalian TAS2R sequences obtained from Uniprot (**mammalian-TAS2R.pir**); and an annotated MSA of human TAS2Rs (**TAS2R-msa-annotated.xlsx**).

References

- [1] D.P. Staus, R.T. Strachan, A. Manglik, B. Pani, A.W. Kahsai, T.H. Kim, L.M. Wingler, S. Ahn, A. Chatterjee, A. Masoudi, Allosteric nanobodies reveal the dynamic range and diverse mechanisms of G-protein-coupled receptor activation, *Nature*, 535 (2016) 448-452.
- [2] A.J. Kooistra, S. Mordalski, G. Pándy-Szekeres, M. Esguerra, A. Mamyrbekov, C. Munk, G.M. Keserű, D.E. Gloriam, GPCRdb in 2021: integrating GPCR sequence, structure and function, *Nucleic Acids Research*, 49 (2021) D335-D343.
- [3] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, *Journal of computational chemistry*, 25 (2004) 1605-1612.
- [4] C. Wang, H. Wu, V. Katritch, G.W. Han, X.-P. Huang, W. Liu, F.Y. Siu, B.L. Roth, V. Cherezov, R.C. Stevens, Structure of the human smoothed receptor bound to an antitumour agent, *Nature*, 497 (2013) 338-343.
- [5] B. Webb, A. Sali, Comparative Protein Structure Modeling Using MODELLER, *Current Protocols in Bioinformatics*, 54 (2016) 5.6.1-5.6.37.
- [6] I. Deshpande, J. Liang, D. Hedeem, K.J. Roberts, Y. Zhang, B. Ha, N.R. Latorraca, B. Faust, R.O. Dror, P.A. Beachy, Smoothened stimulation by membrane sterols drives Hedgehog pathway activity, *Nature*, 571 (2019) 284-288.
- [7] C. Wang, H. Wu, T. Evron, E. Vardy, G.W. Han, X.-P. Huang, S.J. Hufeisen, T.J. Mangano, D.J. Urban, V. Katritch, Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs, *Nature communications*, 5 (2014) 1-11.
- [8] M. Sandal, T.P. Duy, M. Cona, H. Zung, P. Carloni, F. Musiani, A. Giorgetti, GOMoDo: a GPCRs online modeling and docking webserver, *PloS one*, 8 (2013) e74092.
- [9] F. Sievers, D.G. Higgins, Clustal Omega for making accurate alignments of many protein sequences, *Protein Science*, 27 (2018) 135-145.
- [10] M.y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures, *Protein science*, 15 (2006) 2507-2524.
- [11] A. Wiener, M. Shudler, A. Levit, M.Y. Niv, BitterDB: a database of bitter compounds, *Nucleic acids research*, 40 (2012) D413-D419.
- [12] J.A. Ballesteros, H. Weinstein, [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors, in: *Methods in neurosciences*, vol. 25, Elsevier, 1995, pp. 366-428.
- [13] P. Schmidtke, A. Bidon-Chanal, F.J. Luque, X. Barril, MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories, *Bioinformatics*, 27 (2011) 3276-3285.

Publication 6

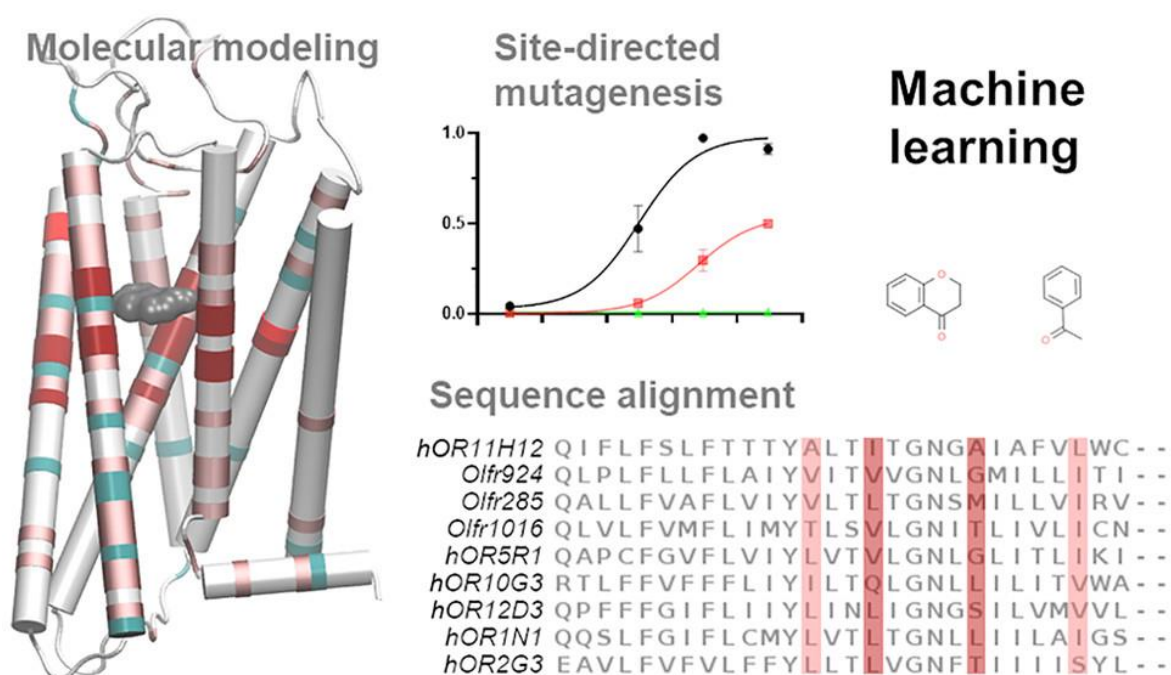
Large-Scale G Protein-Coupled Olfactory Receptor-Ligand Pairing

*Xiaojing Cong, Wenwen Ren, **Jody Pacalon**, Rui Xu, Lun Xu, Xuewen Li, Claire A de March, Hiroaki Matsunami, Hongmeng Yu, Yiqun Yu, Jérôme Golebiowski.*

ACS Cent Sci . 2022 Mar 23;8(3):379-387. doi: 10.1021/acscentsci.1c01495. Epub 2022 Feb 18.

Abstract

G protein-coupled receptors (GPCRs) conserve common structural folds and activation mechanisms, yet their ligand spectra and functions are highly diverse. This work investigated how the amino-acid sequences of olfactory receptors (ORs)—the largest GPCR family—encode diversified responses to various ligands. We established a proteochemometric (PCM) model based on OR sequence similarities and ligand physicochemical features to predict OR responses to odorants using supervised machine learning. The PCM model was constructed with the aid of site-directed mutagenesis, *in vitro* functional assays, and molecular simulations. We found that the ligand selectivity of the ORs is mostly encoded in the residues up to 8 Å around the orthosteric pocket. Subsequent predictions using Random Forest (RF) showed a hit rate of up to 58%, as assessed by *in vitro* functional assays of 111 ORs and 7 odorants of distinct scaffolds. Sixty-four new OR–odorant pairs were discovered, and 25 ORs were deorphanized here. The best model demonstrated a 56% deorphanization rate. The PCM-RF approach will accelerate OR–odorant mapping and OR deorphanization.



Synopsis

Machine learning prediction of protein–ligand pairs using sequence and chemical features: Selecting key residues is an intuitive knowledge-driven method to reduce dimensionality and boost performance.

Introduction

Decoding the sequence–function relationship of proteins is extremely challenging. Slight changes in the sequence may significantly affect the function, whereas proteins with low sequence identity may exhibit similar functions. G protein-coupled receptors (GPCRs) are the most remarkable examples of this phenomenon. They are the largest membrane protein family and the targets for about 40% of marketed drugs [1]. The human genome contains over 800 genes coding for GPCRs [2], which exert differentiated and specific functions in the complex cellular signaling network. Half of these genes are olfactory receptors (ORs) that endow us with fascinating capacities of odor discrimination [3]. Mammalian GPCRs conserve a typical structure of seven transmembrane helices (7TM) that house an orthosteric ligand-binding pocket [4]. They show a conserved signaling mechanism that involves large-scale conformational changes to accommodate their cognate G proteins. The mechanism is encoded in conserved motifs throughout the 7TM, which form a network of inter-TM contacts converging at the cytoplasmic side [5]. Specifically, the “D(E)RY”, “CWLP”, and “NPxxY” motifs in TM3, TM6, and TM7, respectively, are the most conserved hubs of the allosteric communication between the orthosteric pocket and the cytoplasmic side of class A GPCRs [4]. The orthosteric pocket, by contrast, has diversified extensively and resulted in huge variations in the receptors’ function.

This study focuses on the functional heterogeneity of ORs and how this is encoded in the OR sequences. ORs discriminate a vast spectrum of volatile molecules (odorants) and code for an innumerable number of odors perceived in the brain. The many-to-many relationships between ORs and odorants are key to understanding odor perception [6]. Although odorant-binding proteins (OBPs) also contribute to odor detection, they are abundant extracellular proteins that participate in perireceptor events by selecting/carrying odorants [7, 8]. Currently, OR–odorant interactions are mostly measured in heterologous cells, especially for human ORs, which

neglects the effect of OBPs. ORs are also expressed ectopically, and some have emerged as appealing drug targets [9–12]. We sought to predict OR responses to various odorants using OR sequence alignment, proteochemometrics (PCM) [13], and machine learning. The PCM model was based on the OR sequence similarities and the chemical features of the odorants. Sequence-based approaches can handle large protein families and circumvent the difficulties in obtaining high-resolution structures, as is the case for ORs. Machine learning models using protein sequences and ligand chemical similarities have shown great success in predicting drug–target interactions, such as reviewed in refs [14–16]. Attempts to predict OR responses to odorants have also achieved encouraging results [17–20]. However, data scarcity in the immense odor space is a major bottleneck for good predictivities. To date, less than 50% of human ORs (hORs) and 20% of mouse ORs (mORs) have been deorphanized with less than 250 odorants ([Table S1](#)). One effective way to handle data scarcity is dimension reduction, such as by selecting relevant residues in the OR sequences (the so-called feature selection). A recent study on insect and mammalian ORs demonstrated that selecting subsets of 20 residues could indeed increase the model predictivity [20]. However, if one assumes that a given function is mostly encoded by 20 residues out of a GPCR sequence of ~ 300 residues, the binomial coefficient $[300!/20!(300 - 20)!]$ gives more than 10^{30} possible combinations. Therefore, selecting relevant residues is key to constructing an effectual model.

Like other GPCRs, ORs respond to their ligands via allosteric mechanisms, which involve distinct interwound factors: ligand affinity, intrinsic stability of different receptor states, as well as long-range allosteric coupling between the ligand-binding pocket and the cytoplasmic side [21]. Ligand affinity is thought to be dictated by the residues outlining the binding pocket [22, 23]. ORs that respond to the same odorants share higher sequence homology around the pocket than in the rest of the receptor sequence [18].

The OR response to odorants can be drastically altered by mutations that are distant from the pocket [24]. It is nontrivial to select the relevant residues. Here, we combined molecular modeling, site-directed mutagenesis with *in vitro* functional assays, and machine learning to identify the most relevant residues. PCM modeling and random forest (RF) were employed to predict OR responses to prototypical odorants using the relevant residues. Finally, *in vitro* functional assays were performed to assess the selection of relevant residues as well as the predictivity of the PCM-RF model. This approach (outlined in [Figure 1A](#)) largely outperformed existing models by enabling knowledge-based residue selection. It illustrated how the functional heterogeneity of G protein-coupled ORs is encoded in the sequence.

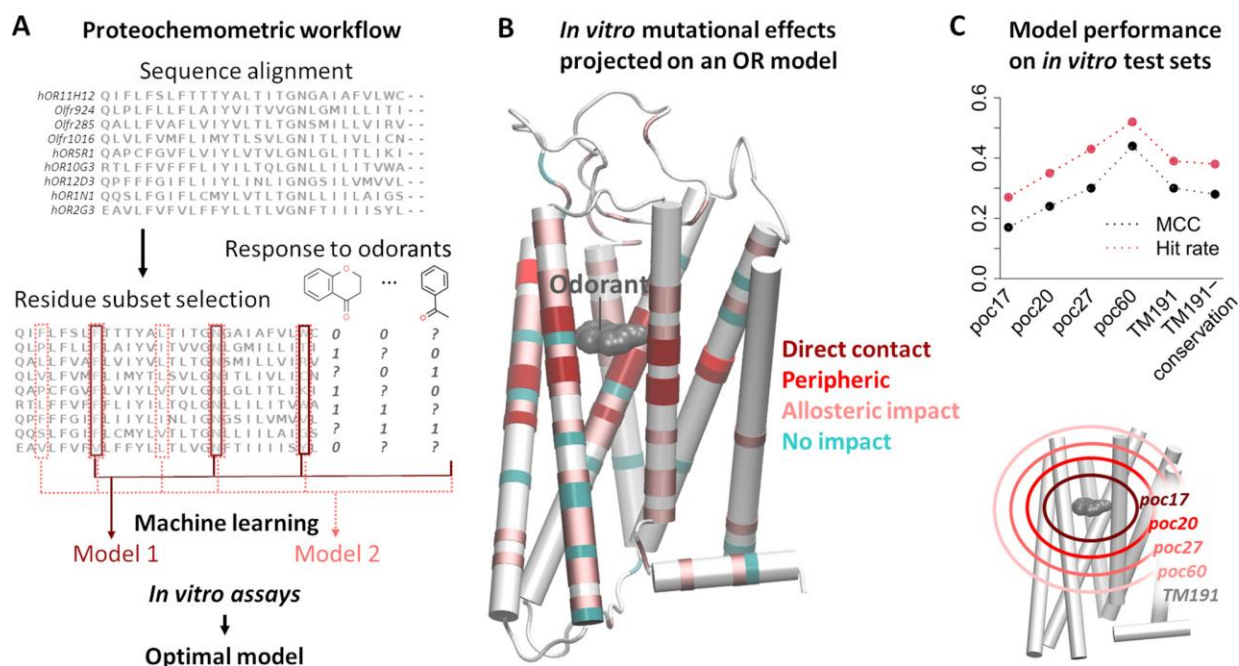


Figure 1: Machine learning protocol and residue selection. (A) Machine learning workflow, in which different residue subsets were extracted from the sequence alignment for the training of different models. The PCM approach combined the OR sequence features, the ligand physicochemical features, and the response data (if available) of each OR–ligand pair. (B) Available site-directed mutagenesis data (including literature data, summarized in ref [24]) projected on the 3D model of mOR256-31. Residues in dark red and red belong to *poc17* and *poc20*, respectively. (C) Matthew's correlation coefficient (MCC) (28) and hit rate of the RF classifiers on the *in vitro* test set.






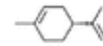


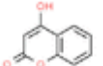

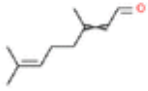



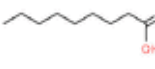
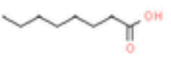
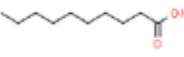
Results

Database of OR–Odorant Pairs for Model Training

We examined all of the literature data of *in vitro* dose-dependent responses of hORs and mORs to diverse odorants. These include 1293 OR–odorant pairs consisting of 390 ORs and 244 odorants. In addition, we included more than 14 400 OR–odorant pairs which have been reported to be nonresponsive *in vitro*. The database ([Data File S1](#)) contains 720 distinct ORs (including 318 orphan ORs) and 244 odorants. Four odorants were considered here as test cases: acetophenone, coumarin, R-carvone, and 4-chromanone. They have been associated with many ORs (dozens to hundreds) in previous studies ([Table 1](#)). To enlarge the training set, we also

included the data of 6 additional odorants that have similar chemical structures to the 4 target odorants.

Table 1: Chemical Structure, PubChem CID, and Training Data^a of the Query Odorants (in Bold) and Their Analogues.

<p>Acetophenone, CID: 7410 <i>P</i>: 11 hORs, 78 mORs <i>N</i>: 183 hORs, 91 mORs</p> 	<p>R-carvone, CID: 439570 <i>P</i>: 7 hORs, 11 mORs <i>N</i>: 7 hORs, 46 mORs</p> 	<p>S-limonene, CID: 439250 (R-carvone analog) <i>P</i>: 4 hORs</p> 
<p>Methyl benzoate, CID: 7150 (acetophenone analog) <i>P</i>: 2 hORs, 22 mORs <i>N</i>: 255 hORs</p> 	<p>S-carvone, CID: 16724 (R-carvone analog) <i>P</i>: 6 hORs, 10 mORs <i>N</i>: 7 hORs, 42mORs</p> 	<p>R-limonene, CID: 440917 (R-carvone analog) <i>P</i>: 12 hORs, 1 mOR <i>N</i>: 186 hORs, 3 mORs</p> 
<p>Isolimonene, CID: 22831540 (R-carvone analog) <i>P</i>: 2 hORs</p> 	<p>Coumarin, CID: 323 <i>P</i>: 7 hORs, 11 mORs <i>N</i>: 8 hORs, 51 mORs</p> 	<p>4-Hydroxycoumarin, CID: 54682930 (coumarin analog) <i>P</i>: 1 mOR <i>N</i>: 10 hORs, 51 mORs</p> 
<p>4-chromanone, CID: 68110 <i>P</i>: 3 hORs, 13 mORs <i>N</i>: 7 hORs, 39 mORs</p> 	<p>Citral, CID: 638011 <i>P</i>: 17 hORs, 1 mOR <i>N</i>: 176 hORs, 5 mORs</p> 	
<p>Nonanal, CID: 31289 <i>P</i>: 9 hORs, 7 mORs <i>N</i>: 9 hORs, 47 mORs</p> 	<p>Octanal, CID: 454 (nonanal analog) <i>P</i>: 4 hORs, 4 mORs <i>N</i>: 231 hORs, 52 mORs</p> 	<p>Decanal, CID: 8175 (nonanal analog) <i>P</i>: 7 hORs, 3 mORs <i>N</i>: 85 hORs, 49 mORs</p> 
<p>Nonanoic acid, CID: 8158 <i>P</i>: 5 hORs, 12 mORs <i>N</i>: 9 hORs, 42 mORs</p> 	<p>Octanoic acid, CID: 379 (nonanoic acid analog) <i>P</i>: 1 hOR, 9 mORs <i>N</i>: 9 hORs, 47 mORs</p> 	<p>Decanoic acid, CID: 2969 (nonanoic acid analog) <i>P</i>: 2 hORs, 9 mORs <i>N</i>: 9 hORs, 43 mORs</p> 

^a *P*: number of responsive (positive) ORs. *N*: number of nonresponsive (negative) ORs. See [Data File S1](#) for the lists of ORs.

Selection of Relevant Residues

Molecular Modeling

Given the existing knowledge of GPCR structures, we first sought for odorant-binding residues within the orthosteric ligand-binding pocket. The mouse OR mOR256-31 (gene name *Olf263*) was chosen as a prototype, since it is a broadly tuned receptor which responds to three of the

four odorants (coumarin, R-carvone, and acetophenone) [25, 26]. We built a 3D homology model of mOR256-31 bound with the odorants using our previously established approaches and molecular dynamics simulations [24, 25, 27]. The 3D model was built under the constraints of conserved amino-acid motifs and site-directed mutagenesis data covering nearly 50% (95 residues) of the TM domain [24]. Seventeen residues were identified within a 5 Å distance of the bound odorants ([Table S2](#)). Fourteen of these residues had been shown to be important for OR responses to odorants by site-directed mutagenesis ([Table S2](#)). These 17 residues were assumed to be in direct contact with the odorants (named *poc17* hereafter, [Figure 1B](#)). However, the relevant residues should include many more than the sole binding pocket.

Site-Directed Mutagenesis

Twenty-four point-mutations were generated within and around *poc17* of mOR256-31. Their impact on the receptor's response to five ligands was measured by *in vitro* dose-dependent responses ([Figure S1](#)). We projected the mutational effect onto the 3D model of mOR256-31, together with all of the OR mutations reported in the literature ([Figure 1B](#)). Twenty residues including *poc17* and 3 peripheral residues ([Figure 1B](#)) delineated a larger orthosteric pocket (*poc20*). Mutations within *poc20* consistently affected the response to most of the odorants. Beyond the region of *poc20*, the mutational effect was less systematic ([Figure 1B](#)).

To determine the best subset of residues for predicting OR responses to odorants, we proceeded in an empirical approach. Namely, we selected 5 small-to-large residue subsets as heuristics, based on the above results: *poc17*, *poc20*, *poc27*, *poc60*, and *TM191*. *poc27* and *poc60* are extensions of the pocket until 6 and 8 Å from the bound odorant, containing 27 and 60 residues, respectively ([Figure 1C](#) and [Table S3](#)). *TM191* contains the whole 7TM region made up of 191 residues. Machine learning models were then built with these residue subsets to compare their predictive power.

PCM and Machine Learning

From the sequence alignment of hORs and mORs, each of the 5 heuristic residue subsets were extracted. PCM models were constructed using the data in [Table 1](#) and physiochemical features of the odorants (see the [Material and Methods](#) section). Each OR–odorant pair was labeled with

the *in vitro* response (responsive or nonresponsive). We trained and assessed supervised support vector machine (SVM) and RF classifiers using 5-fold cross validation. The response probability of each OR–odorant pair was predicted, and a probability >0.5 was classified as responsive. The predictivity was measured by Matthew’s correlation coefficient (MCC) [28]. RF performed better than SVM. The predictivities of the five RF classifiers were not significantly different from one another.

However, they were clearly superior to a naive statistical inference ([Figure S2A](#); see the Supplementary Methods section for the calculation of the statistical inference). The *poc60* classifier performed the best on average ([Figure S2A](#), [Data File S2A,B](#)). Control models built with 60 randomized residues, as expected, showed no predictivity ([Figure S2A](#)). To determine the best residue subset, we constructed five final RF classifiers (*poc17*, *poc20*, *poc27*, *poc60*, and *TM191*) using 100% of the data in [Table 1](#). Each classifier was then used to screen for new ORs for acetophenone, R-carvone, coumarin, and 4-chromanone. The *in silico* screening was performed on 360 ORs (223 hORs and 138 mORs), including 346 orphan ORs. Each classifier predicted and ranked the probabilities of the ORs to respond to each of the 4 odorants ([Data File S2C](#)).

In Vitro Assessment of Relevant Residues

We tested the predictions of all five classifiers in cell functional assays. For each model, we tested all ORs in the responsive class (predicted response probability >0.5 for any odorant) as well as 60 negative control ORs (response probability <0.5 for all odorants). These ORs were tested against all 4 odorants. For instance, in the case of *poc60*, we tested all 20 ORs in the responsive class and 60 randomly picked negative controls from the nonresponsive class ([Figure 2](#)). Similar tests were performed on the other four models ([Figure S3 and Table S4](#), [Data File S2C,D](#)). When significant responses were observed at 300 μ M, dose-dependent responses were measured. Otherwise, the OR–odorant pair was considered nonresponsive. The *poc60* classifier performed the best on the *in vitro* test set ([Figure 1C](#)). It showed 0.39–0.60 hit rates and 0.43–0.48 predictivity (MCC) for the 4 odorants ([Table 2](#)). Therefore, *in vitro* data confirmed that *poc60* is the most relevant residue subset to decode the receptor’s response to odorants. These residues show very low conservation in hORs and mORs ([Figure S2B](#)), suggesting that they have diversified to adapt to various ligands [22, 23]. This implies that amino acid conservations in the OR sequences contain essential information for their

functionality. Thus, we tested an additional model using the amino acid conservations in the TM region. This model turned out to be nearly as predictive as using the amino acid physicochemical features (Figure 1C). This indicates that the type of features used to describe the amino acids is not critical, as long as the features sufficiently convey the sequence differences to the machine learning algorithm.

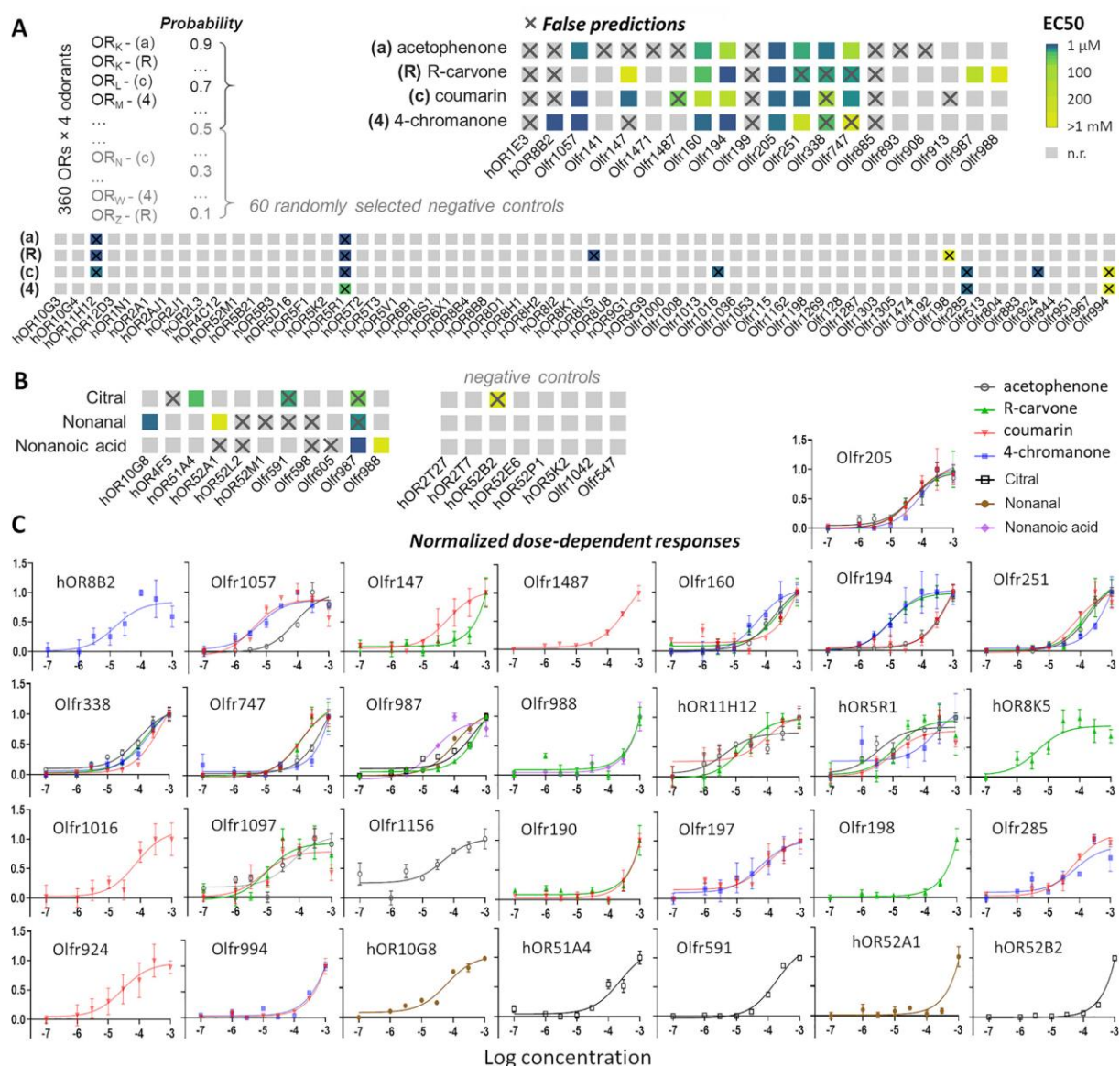


Figure 2: *In vitro* evaluation of machine learning predictions of OR responses to odorants. (A) All of the OR–odorant pairs were ranked by the predicted probability to be responsive. The initial model assessments focused on four odorants. 20 responsive and 60 nonresponsive ORs (negative controls) predicted by the *poc60* model were selected for functional assays. Heatmaps show the *in vitro* EC50 values, in which the false predictions are labeled with ×. Assessments of the other models are provided in Figure S3. (B) *In vitro* assessment of the *poc60* model predictivity for acyclic odorants. (C) Dose-

dependent response curves of all of the responsive OR–odorant pairs identified in this study. Error bars indicate SEM ($n = 3–6$).

Table 2: Performance of the *poc60* Model in Predicting New OR–Odorant Pairs^a

metrics ^b	initial test odorants				additional test odorants		
	acetophenone	R-carvone	coumarin	4-chromanone	citral	nonanal	nonanoic acid
MCC	0.47	0.45	0.43	0.48	0.24	0.48	0.40
hit rate (precision)	0.39	0.6	0.58	0.6	0.50	0.50	0.25
recall (sensitivity)	0.78	0.46	0.47	0.5	0.25	0.67	1.00
F1 score	0.52	0.52	0.52	0.55	0.33	0.57	0.40
specificity	0.85	0.94	0.92	0.94	0.93	0.88	0.65
AUC	0.84	0.72	0.72	0.74	0.58	0.66	0.74
true positives	7	6	7	6	1	2	2
true negatives	60	63	60	64	14	14	11
false positives	11	4	5	4	1	2	6
false negatives	2	7	8	6	3	1	0

^a See [Data File S2C](#) for the raw data.

^b See the [Methods](#) section in the SI for the definitions.

Assessment of Model Utility

Applicability to Other Odorants

While 50% of hORs and 20% of mORs have been deorphanized at the time of this study, only a tiny fraction of the odorant chemical space (<250 odorants) has been tested. The lack of data on odorants is a major restraint on the model utility. To explore this limitation, we generated a learning curve of the *poc60* model predictivity on the external test set versus the amount of training data used ([Figure S4A](#)). The learning curve suggested that a meaningful prediction could be obtained for an odorant with ~15 known ORs. In the current database containing 244 odorants, only 17 (7%) met this criterion, 11 of which contained aromatic or cyclic structures. We attempted three more odorants that contain alkyl chains, citral, nonanal, and nonanoic acid. Following the same procedure, we tested *in vitro* all 11 ORs that were predicted to respond to

any of the three odorants as well as 8 negative control ORs ([Figure 2B](#)). Because the training data lacked responsive ORs for these odorants, the model predicted less responsive pairs than for the 4 cyclic odorants. *In vitro* assays showed that the model performed well on nonanal and nonanoic acid but not on citral ([Table 2](#)). The poor predictivity on citral was likely due to the lack of analogues (thus the lack of data) in the training set ([Table 1](#)) and the fact that citral is a mixture of two isomers, which add ambiguity to the available data. The results demonstrate that the model is generalizable to odorants of different chemical groups, provided enough training data for the odorants in question or their close analogues.

General Model Performance

We evaluated the general performance of the *poc60* model on all of the external test set data, including those tested for the other models and for citral. The test set data were shuffled and split into 5 folds, like in a cross validation. The model predictivity was coherent on the 5 folds of the data set, which gave 0.39–0.46 hit rates and 0.32–0.34 MCC ([Table S6](#)). Blind OR–odorant screening hit rates in Hana3A cells are expected to be lower than 0.1, such as in a pioneer study on 245 hORs and 219 mORs against 93 odorants [19]. Note that the odorants tested here might be more promiscuous than average, since the model requires training data for the query odorants or their analogues. Our test set also enriched more responsive ORs (26%) than in the natural pool of ORs (e.g., 13% in ref [19]), despite the large number of negative-control ORs included. Since many ORs fail to express on the membrane of heterologous cells, it is difficult to estimate the general response rate of ORs to various odorants.

The total external test sets in this work contained 111 ORs and 438 OR–odorant pairs. We identified 63 new OR–odorant pairs with EC50 values in the micromolar to millimolar range, corresponding to 29 ORs ([Figure 2C](#), [Figure S3](#) and [Table S5](#)). Twenty-five ORs were deorphanized in this study, including 9 from the negative control groups. Nevertheless, the deorphanization rate is significantly higher in the predicted positive groups than in the negative control groups ([Figure S4B](#)), which are 56% and 15%, respectively, for the *poc60* model.

Utility for New ORs and Odorants

One important aspect of the model utility is its predictivity on new ORs and odorants that are not part of the training set. While 56 out of the 95 ORs in the external test set are “new”, we recalculated the model performance metrics for this part of the test set. The model still showed good predictivity compared to the full test set ([Table S7](#)). The model predictivity on new odorants was evaluated by the following test: we excluded the 7 odorants one by one from the training set, retrained the model, and calculated the performance metrics on the test set containing only the excluded odorant. In this case, the model only showed predictivity for cyclic odorants, acetophenone, R-carvone, and 4-chromanone ([Table S8](#)). Therefore, the application to new odorants is currently limited by the lack of training data, as already discussed above. New data will gradually enable the application to more odorants. Currently, the model is readily applicable to new ORs for which there are no training data.

Discussion

This work illustrates how the G protein-coupled ORs’ response to ligands can be decoded from their sequence. Sixty residues around the odorant-binding pocket contain the highest signal-to-noise ratio and dictate the variation in the ORs’ response to the odorants ([Figure 3](#)). The ligand-binding pocket of GPCRs has highly diversified during evolution to discriminate various stimuli. It is not surprising that the ORs’ response to the odorants could be predicted by using less than 20% of the sequence, made up with highly variable residues. The results validate previous predictions of pocket residues based on OR sequence analysis [22, 23] and numerous site-directed mutagenesis data [23, 24], which are located in the upper portion of TM3 and TM5–TM7. Here, we highlight 4 residues in TM2 near a conserved allosteric site (centered at D^{2.50}). The allosteric site in nonolfactory class A GPCRs (typically composed of D^{2.50}, N^{3.35}, and S^{3.39}) is known to bind the Na⁺ ion, which modulates the receptors’ activation and affinity/response to ligands (reviewed in ref [29]). Most ORs contain a second acidic residue (E^{3.39}) at this site, which might also accommodate divalent cations [29]. While copper ions play important roles in the recognition of sulfur odorants [30, 31], it remains unclear whether this conserved site in the ORs is involved. The machine learning model established here outperformed existing models using full sequences [17, 19]. The pocket residues are essential

for understanding how chemically similar odorants are differentiated by the OR family with such high specificity/selectivity.

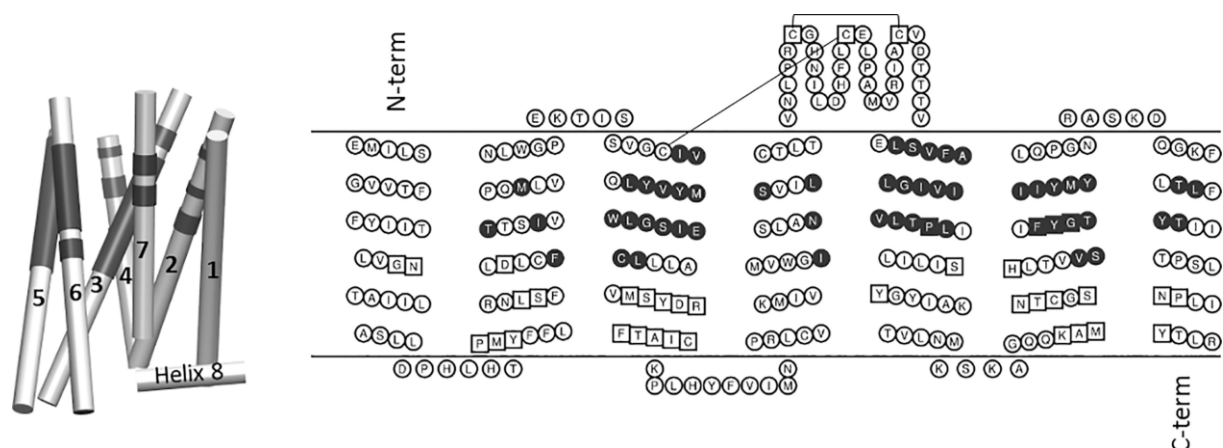


Figure 3: Location of the residues that best encode OR responses to ligands, illustrated with mOR256-31. Conserved motifs in ORs are squared. The N- and C-termini are truncated for clarity.

So far, research focusing on specific OR–ligand recognition has mostly employed molecular modeling (e.g., homology modeling, docking, and molecular simulations) verified by site-directed mutagenesis and functional assays of individual ORs, such as the studies reviewed in ref [32], as well as the more recent work on hOR1A1 for R-/S-carvone enantiomers [33], hOR5AN1 and mOR215-1 for musk odorants [34], zebrafish ORs for bile acids/salts [35], and a virtual screening for new mOR-EG ligands [36]. This approach provides valuable insights into OR–ligand recognition and will continue to generate data for new ORs and ligands. Since it relies on experimental data to generate predictive molecular models, this approach is not suitable for large-scale OR–ligand pairing. The molecular modeling process can be automated to enable large-scale studies [37]; however, the performance has yet to be tested. Ligand QSAR/SAR models using machine learning have also been adopted to predict new OR ligands [38, 39]. This approach allows a rapid virtual screening of large compound databases and is widely used in drug design and drug toxicity prediction [40]. It is limited to the target receptor and the chemical scaffolds of the known ligands. However, the application on ORs will gradually enrich ligand data and reduce the bottleneck of our PCM model.

The machine learning PCM approach established here is readily applicable to the entire mammalian OR family. It will significantly accelerate OR–ligand mapping and OR deorphanization. It is an open loop process where newly identified OR–odorant pairs can be added to continuously improve the model. Because we optimized the model to maximize the

hit rate (to reduce the cost of *in vitro* assays), this consequently gave way to false negatives (Figure S4C). Therefore, repeating the prediction–test loop is necessary to rescue the false negatives by injecting new training data. Note that the lack of response of many orphan ORs might be due to impaired functions in heterologous cells, e.g., lack of cell surface expression [41]. For instance, ~30% of the mORs responding to acetophenone *in vivo* did not show significant responses in heterologous cells [18]. Such cases may be present in the nonresponsive ORs in the *in vitro* test set, the proportion of which is difficult to estimate.

This approach is mostly applicable to large protein families like GPCRs or promiscuous proteins, such as functionally related enzymes [34], odorant/pheromone-binding proteins in insects [35], intrinsically disordered protein regions [36], as well as GPCR-G protein binding partners [37]. The approach focuses on the sequence of the binding region, which overcomes the difficulties in obtaining high-resolution structures or full sequence alignments. It may find applications in, for example, predicting off-target activities in drug design, targeting insect pheromone receptors for pest control, or studies of protein–protein interactions and protein evolution. It requires sequence alignment and a number of known ligands as input data. The selection of relevant residues is important, which enables knowledge-based human intervention to reduce the dimensionality and enhance machine learning on scarce data. Combining *in vitro* functional assays, site-directed mutagenesis, knowledge of GPCR structures and sequences, as well as molecular modeling, we could generate heuristics to decipher how nature has encoded the specific functions of ORs into their varied sequences.

The model is currently limited to the transmembrane domain where the sequence alignment has been established. The loop regions may be addressed for OR subfamilies for which good sequence alignments can be obtained. The discovery of residue subsets associated with given functions could indicate evolutionary hotspots and compensate for existing tools such as phylogenetic analysis based on full sequences.

Materials and Methods

Chemicals and OR Constructs

Odorants were purchased from Sigma-Aldrich. They were dissolved in DMSO to make stock solutions at 1 mM and then freshly diluted in optimal MEM (ThermoFisher) to prepare the

odorant stimuli. The OR constructs were kindly provided by Dr. Hanyi Zhuang (Shanghai Jiaotong University, China). Site-directed mutants were constructed using the Quikchange site-directed mutagenesis kit (Agilent Technologies). The sequences of all plasmid constructs were verified by both forward and reverse sequencing (Sangon Biotech, Shanghai, China). The list of primers used in this study are listed in [Table S9](#).

Cell Culture and Transfection

We used Hana3A cells, a HEK293T-derived cell line that stably expresses receptor-transporting proteins (RTP1L and RTP2), receptor expression-enhancing protein 1 (REEP1), and olfactory G protein ($G\alpha_{olf}$) [42]. The cells were grown in MEM (Corning) supplemented with 10% (v/v) fetal bovine serum (FBS; ThermoFisher) and 100 $\mu\text{g}/\text{mL}$ penicillin–streptomycin (ThermoFisher), 1.25 $\mu\text{g}/\text{mL}$ amphotericin (Sigma-Aldrich), and 1 $\mu\text{g}/\text{mL}$ puromycin (Sigma-Aldrich).

All constructs were transfected into the cells using Lipofectamine 2000 (ThermoFisher). Before the transfection, the cells were plated on 96-well plates (NEST) and incubated overnight in MEM with 10% FBS at 37 °C and 5% CO_2 . For each 96-well plate, 2.4 μg of pRL-SV40, 2.4 μg of CRE-Luc, 2.4 μg of mouse RTP1S, and 12 μg of receptor plasmid DNA were transfected. The cells were subjected to a luciferase assay 24 h after transfection.

Luciferase Assay

The luciferase assay was performed with the Dual-Glo luciferase assay kit (Promega) following the protocol in ref [42]. OR activation triggers the $G\alpha_{olf}$ -driven AC-cAMP-PKA signaling cascade and phosphorylates CREB. Activated CREB induces luciferase gene expression, which can be quantified luminometrically [measured here with a bioluminescence plate reader (MD SPECTRAMAX L)]. Cells were cotransfected with firefly and *Renilla* luciferases where firefly luciferase served as the cAMP reporter. *Renilla* luciferase is driven by a constitutively active simian virus 40 (SV40) promoter (pRL-SV40; Promega), which served as a control for cell viability and transfection efficiency. The ratio between firefly luciferase versus *Renilla* luciferase was measured. Normalized OR activity was calculated as $(L_N - L_{\min}) / (L_{\max} - L_{\min})$, where L_N is the luminescence in response to the odorant, and L_{\min} and L_{\max} are the minimum

and maximum luminescence values on a plate, respectively. The assay was carried out as follows: 24 h after transfection, the medium was replaced with 100 μL of odorant solution (at different doses) diluted in optimal MEM (ThermoFisher), and cells were further incubated for 4 h at 37 $^{\circ}\text{C}$ and 5% CO_2 . After incubation in lysis buffer for 15 min, 20 μL of Dual-Glo luciferase reagent was added to each well of a 96-well plate, and firefly luciferase luminescence was measured. Next, 20 μL of Stop-Glo luciferase reagent was added to each well, and *Renilla* luciferase luminescence was measured. The data analysis followed the published procedure in ref [42]. Three-parameter dose–response curves were fitted with GraphPad Prism 8.

Molecular Modeling

Homology models of mOR256-3, mOR256-8, and mOR256-31 were built using the approach in our previous work [24, 27]. Four X-ray crystal structures of class A GPCRs were used as templates, rhodopsin (1U19), CXCR4 (3ODU), A2aR (2YDV), and CXCR1 (2LNL), to build 100 models with Modeller v9.15 [43]. For docking, we chose the model with the lowest DOPE score. Autodock Vina [44] and the Haddock 2.2 Web server [45] were used to identify a common top-ranked binding pose for each odorant. Residues in the putative ligand-binding pocket were set flexible during docking. Enhanced-sampling all-atom molecular dynamics simulations were performed in a bilayer of an explicit POPC membrane (see the [Methods](#) section in the SI for details). A cluster analysis of the ligand-binding pose was carried out on the simulation trajectories using the Gromacs Cluster tool. The middle structure of the most populated cluster was selected as the final binding pose.

Proteochemometric Machine Learning Model

We assembled the response data of 720 ORs and 244 odorants from the literature to construct the training set ([Data File S1](#)). Ambiguous data records (i.e., OR responses without clear dose-dependent data) were discarded. The full training set contained 1293 responsive OR–odorant pairs (composed of 392 ORs and 244 odorants) and 14 459 OR–odorant pairs that have been reported to be nonresponsive *in vitro* (composed of 550 ORs and 127 odorants, including 318 orphan ORs). Each OR–odorant pair was represented by a vector composed of physicochemical descriptors (features) of the OR sequence and the odorant (see the [Methods](#) section in the SI for

details). The OR–odorant pairs in the training set were labeled “positive” or “negative” according to the response data for supervised machine learning. The test set was constructed in the same manner without labels. The test set contained 360 ORs (including 346 orphan ORs) available in our laboratory, paired with the 7 odorants tested in this study. RF and SVM classification models were built with the Caret package in R [46]. RF performed better than SVM and was chosen for the final model. The R code generated during this study is available as a Jupyter notebook, along with the input and output data, at <https://github.com/chemosim-lab/OlfactoryReceptors> under the GNU General Public License v3.0. The Jupyter notebook illustrates step-by-step the model building, training, and the *in vitro* assessment. The process is illustrated in [Figure S2A](#). More details can be found in the [Methods](#) section in the SI.

Safety Statement

No unexpected or unusually high safety hazards were encountered.

Supporting Information

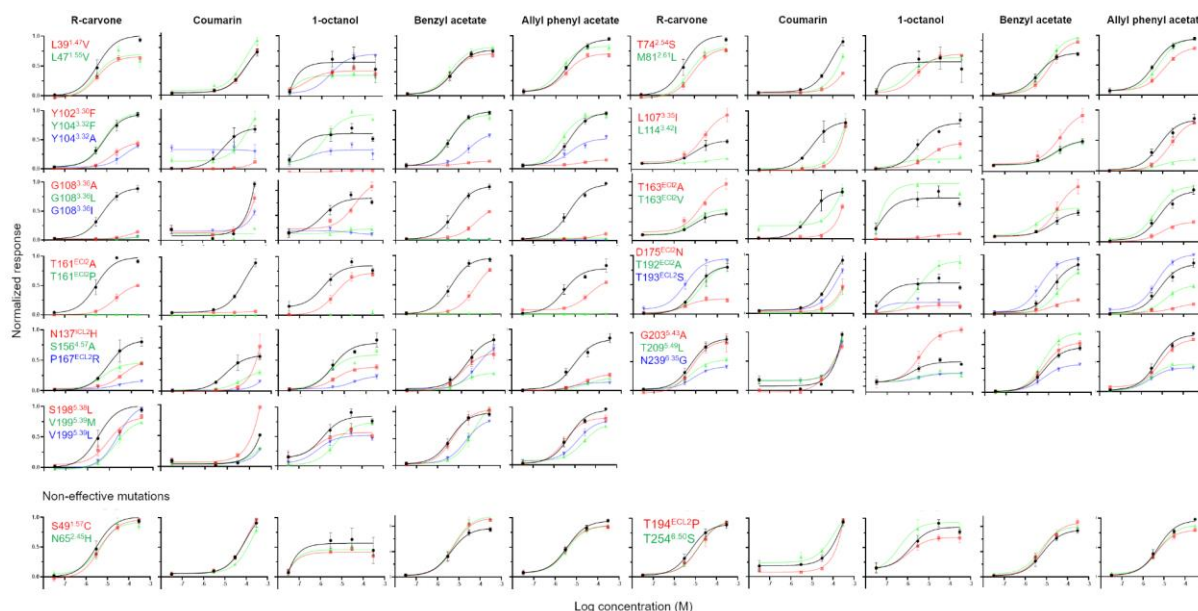


Fig. S1: Dose-dependent response of mOR256-31 mutants to five odorants. The residues were mutated to the corresponding ones in mOR256-8 (red), a narrowly tuned OR that does not

response to these odorants 1. Four other mutations (blue and green) were made to the pocket residues Y104^{3,32}, G108^{3,36} and V199^{5,39}.

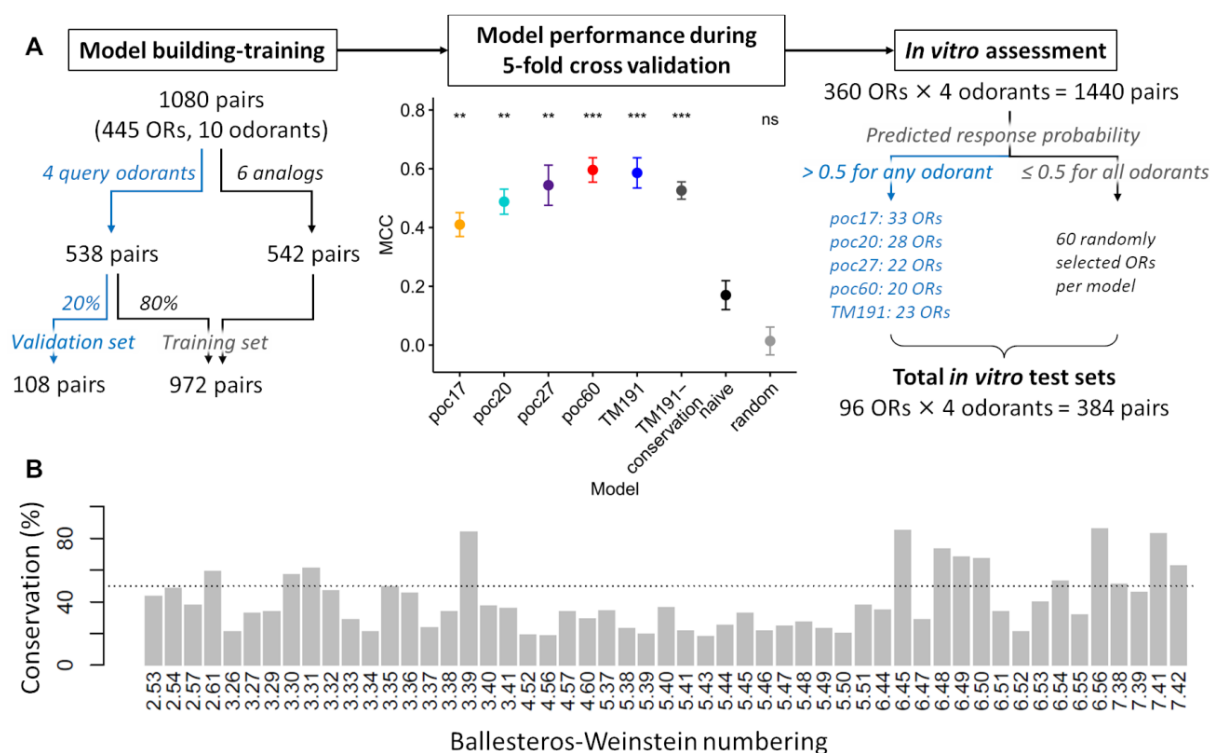


Fig. S2: Workflow and conservation of *poc60* residues. **(A)** Workflow of machine learning model building and training on 4 odorants (through a 5-fold cross validation). Left: 1080 OR-odorant pairs constituted the initial data set for model training, among which 538 pairs were associated with the 4 odorants of interest. The rest of the data referred to 6 analogs of the 4 query odorants, which were included to increase the size of the training set. The validation set contained 20% of the data (108 OR-odorant pairs) corresponding to the 4 query odorants. Middle: The dataset was stratified into training/validation sets 5 times to cross-validate each of the 8 models. Five of the models were built with selected residue subsets from the OR sequence alignment: *poc17*, *poc20*, *poc27*, *poc60*, and *TM191*. The *TM191-conservation* model used the amino acid conservation score in the TM region of the OR sequence alignment, instead of the physicochemical features of the amino acids. The *random* model was built with pseudo OR sequences containing 60 randomized amino acids. The *naive* model was a statistical inference of the response probability of each OR-odorant pair, by calculating their average responsiveness in the training set population (see Materials and Methods for details). Error bars indicate SEM of the 5-fold cross validation. Each model was compared with the *naive* statistical control ($*p < 0.05$, $**p < 0.01$ and $***p < 0.001$). Matthew's correlation coefficient score (MCC) showed that *poc60* led to the best model, although the performance was not statistically different from the other 4 residue subsets. The *naive* model was slightly better than *random*, both of which performed poorly. Right: *in vitro* assessments were performed on the 5 models built with the selected residue subsets. The

5 models predicted varied numbers of responsive ORs (response probability > 0.5 for any of the 4 odorants), e.g., 20 ORs by *poc60* and 33 by *poc17*. These ORs all underwent *in vitro* functional assays on the 4 odorants. As negative controls for each model, 60 ORs were randomly picked from the model predicted non-responsive ORs (response probability ≤ 0.5 for all the 4 odorants). These added up to a total of 384 *in vitro* functional tests. **(B)** Conservation of the most frequent amino acid at *poc60*, which govern OR response to odorants in human and mouse ORs. Seven residue positions that are absent in some of the ORs are not shown.

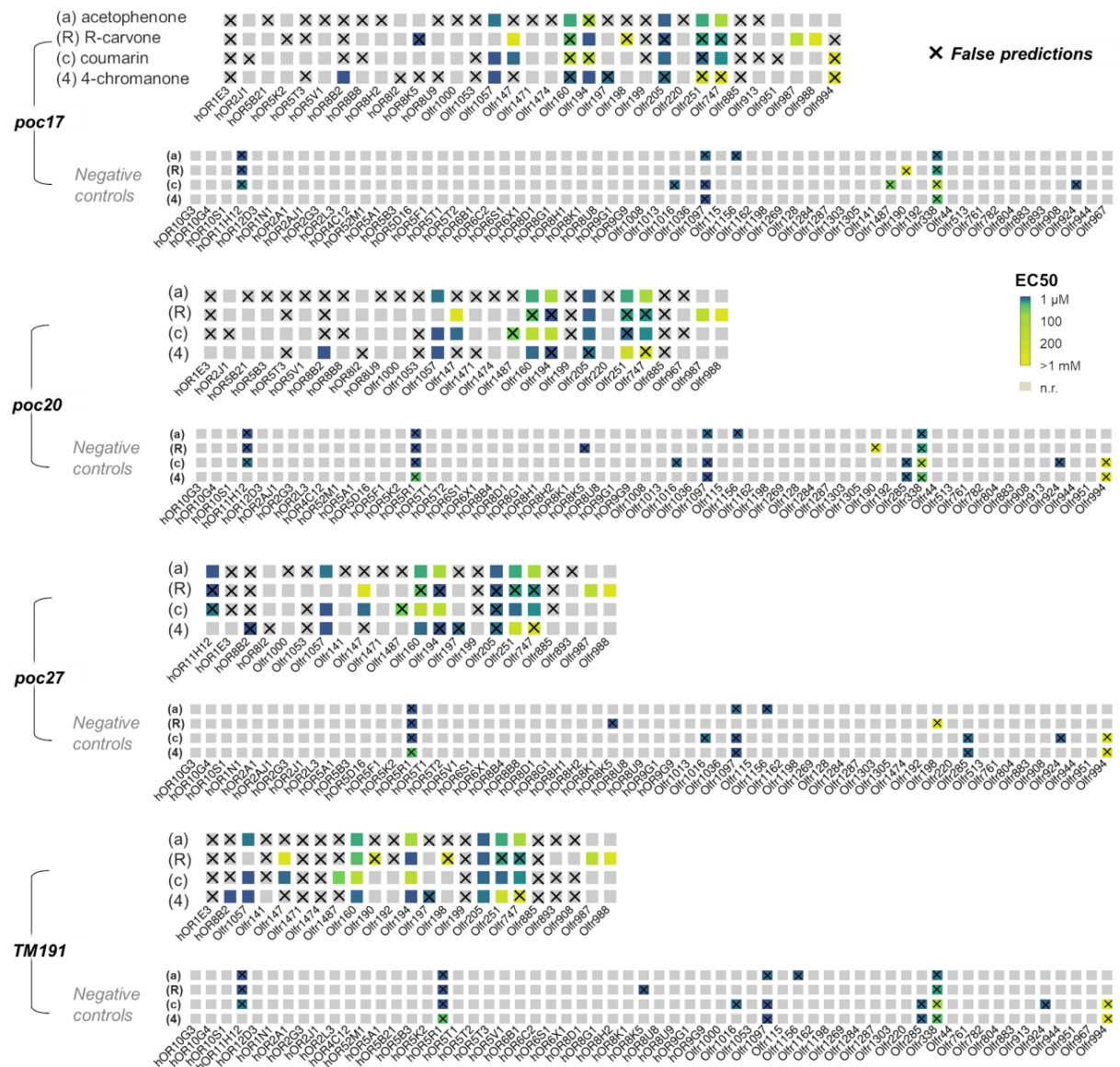


Fig. S3: *In vitro* test sets of different models on 4 odorants. Heatmaps show the EC₅₀ values, in which the false predictions are labeled with a “x” sign.

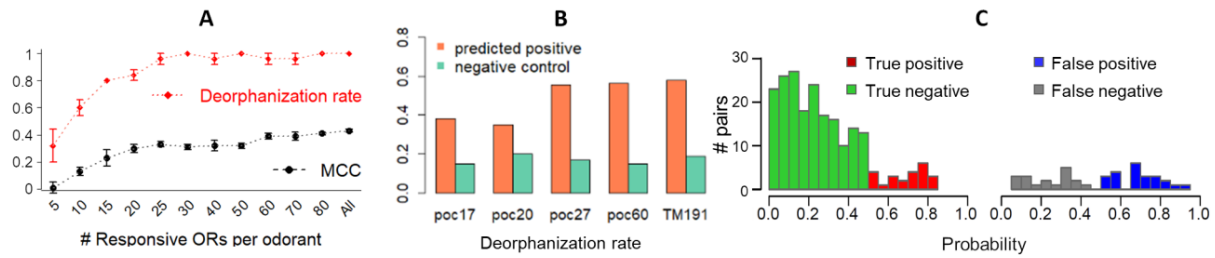


Fig. S4: Analysis of the RF classifiers' predictivity on OR response to odorants. **(A)** Predictive power of the *poc60* model as a function of the number of responsive ORs included in the training data set. The responsive ORs were randomly sampled from the full training set in Table 1. Note that the right side of the plot is dictated by acetophenone which is the only odorant with > 30 responsive ORs. The deorphanization rate (red) refers to the ratio of deorphanized ORs between a model using part of the training data and the one using the full data. For example, the model trained with 20 responsive ORs per odorant could deorphanize 80% of all the deorphanized ORs in this study. Error bars indicate SEM ($n = 5$). **(B)** Deorphanization rate in the predicted positive group and the negative control group of the *in vitro* test set for each model. **(C)** Distribution of predicted probabilities to be responsive OR-odorant pairs in the *in vitro* test set. Probability > 0.5 was considered positive (responsive).

Table S1: Deorphanized human and mouse ORs to date.

hOR51E2, hOR2AT4, hOR11H4, hOR11H6, hOR11H7, hOR12D2, hOR51E1, hOR5A1, hOR8A1, hOR9A2, hOR1G1, hOR1D2, hOR2T34, hOR5B17, hOR2T11, hOR56A5, hOR2AG1, hOR5AN1, hOR1A1_R128H, hOR11A1, hOR52D1, hOR2W1, hOR10G3, hOR10G3_S73G, hOR10G4, hOR10G4_A9V_M134V_V195E_R235G_K295Q, hOR2G2, hOR2J2, hOR2J2_T111A, hOR2T10, hOR51S1, hOR5AC2, hOR1A2, hOR2B3, hOR4D6, hOR4D9, hOR5A2, hOR1A1_V233M, hOR2W1_D296N, hOR4E2_V118M_Q234R, hOR51L1, hOR51L1_I281M, hOR51L1_T196I_A207V, hOR10J3, hOR1S1, hOR3A4_I42V_S86R_E103Q_I120V_P168T_N175S, hOR51B2, hOR51B4_V36I_M147T, hOR52N5, hOR52W1, hOR52H1, hOR52K1, hOR52L1, hOR52R1, hOR56B4, hOR1A1, hOR2J3, hOR2C1, hOR11H1, hOR14A16_Indel, hOR2A25, hOR2A25_S75N, hOR2A25_S75N_A209P, hOR2T35_V31I, hOR4M1, hOR5J2, hOR8B2, hOR4A47_I104L_V145M, hOR4K15_N65S_E88V_S93A_A231E_L280P_I286M, hOR52R1_I129T_N201Y_S245A, hOR5M11, hOR8K3_L122R, hOR1A1_P285S, hOR5P3, hOR8B3_H20R_Q24R_V34I_M114I, hOR51B5, hOR8D1, hOR1N2_W37R_V244G_T301M, hOR2C1_C149W, hOR5K1, hOR10J5, hOR1J4, hOR51T1, hOR52A5, hOR52K2, hOR7A5, hOR3A1, hOR4Q3, hOR10G7, hOR5D18, hOR10G6, hOR2J3_I228V_M261I, hOR10A6, hOR11H13, hOR11L1_G108S_F117L_A142T_R171P, hOR2L2, hOR2L8, hOR2T1_P132L, hOR2Y1_V200L, hOR4D1_R54Q, hOR51D1, hOR51E1_S11N, hOR5A11, hOR5A11, hOR5D14, hOR5P2, hOR6C65, hOR7G2, hOR8J1_Y120C, hOR1E1, hOR2A4, hOR2A7, hOR6P1, hOR4P4, hOR56A4, hOR1C1, hOR2B11, hOR10G7_T13M, hOR10G7_T90A, hOR11G2, hOR1D5, hOR2D2, hOR4D10, hOR4F17, hOR4K5, hOR4N5, hOR4Q3_F238L, hOR56A3_M51T, hOR5L1, hOR11H2, hOR13C9, hOR2AK2, hOR2B2, hOR4A16, hOR10G4_L24P_V195E, hOR2M7, hOR2T11_C119R_Q309R, hOR2M4, hOR2W1_M81V, hOR9G1, hOR13C5, hOR14J1, hOR8K3, hOR52I1, hOR10A6_L287P, hOR10A6_V140P_L287P, hOR2J3_I228V, hOR2J3_R226Q_I228V_M261I, hOR2J3_T113A_R226Q_I228V_M261I, hOR4A5, hOR4F4, hOR51A2, hOR52E5_P234L, hOR6C68_A45T, hOR52J3, hOR2C3_T20A_P68S_L289V, hOR5H2_E11Q_I225V_I281V, hOR7D4, hOR7D4_L17F, hOR7D4_R88W_T133M, hOR1E3, hOR52E8, hOR4E2, hOR2M3, hOR8G5, hOR2C1_G16S_C149W_R229H, hOR2C1_P58S_C149W, hOR6K2, hOR2F1, hOR10G9_A90T_M134V, hOR4D2_C97S_L187F, hOR6S1_T42I_V156I_R237H, hOR1E2_C27R, hOR6C1_V74I_C130Y_L161F_H165D_V246I_R291K, hOR13C8, hOR52N1, hOR10C1_F60L_P174Q_M246V, hOR10K1, hOR10V1_Q101L_V117A_Q123W, hOR14L1, hOR1J1, hOR2T29, hOR2T8, hOR2W3_Indel, hOR51F1, hOR51J1_R75H, hOR52N2_M228T, hOR5B12, hOR6J1_R121H_T124A_R226H_P250S_L274S_V279L, hOR9Q1_A2V, hOR10J5_R233W, hOR5B3, hOR7C1, hOR7C1_S99G_V126I_E171K_S210P, hOR7C1_V126I_E171K_S210P, hOR4N2	Ref.
Olfr362, Olfr15, Olfr167, Olfr168, Olfr895, Olfr599, Olfr638, Olfr1079, Olfr1352, Olfr171, Olfr340, Olfr124, Olfr1362, Olfr1377, Olfr1411, Olfr476, Olfr734, Olfr90, Olfr992, Olfr1062, Olfr1341, Olfr1325, Olfr1356, Olfr1413, Olfr24, Olfr459, Olfr569, Olfr1019, Olfr447, Olfr514, Olfr715, Olfr1264, Olfr311, Olfr632, Olfr64, Olfr979, Olfr1324, Olfr1395, Olfr323, Olfr532, Olfr109, Olfr174, Olfr508, Olfr556, Olfr644, Olfr1032, Olfr62, Olfr202, Olfr790, Olfr406, Olfr1093, Olfr110, Olfr131, Olfr1427, Olfr180, Olfr181, Olfr195, Olfr20, Olfr30, Olfr31, Olfr376, Olfr449, Olfr463, Olfr464, Olfr491, Olfr527, Olfr57, Olfr96, Olfr161, Olfr554, Olfr653, Olfr683, Olfr685, Olfr1328, Olfr796, Olfr983, Olfr749, Olfr1104, Olfr876, Olfr1080, Olfr1095, Olfr1126, Olfr1130, Olfr13, Olfr1353, Olfr1443, Olfr1502, Olfr1504, Olfr1511, Olfr1512, Olfr156, Olfr206, Olfr209, Olfr231, Olfr25, Olfr27, Olfr339, Olfr414, Olfr424, Olfr434, Olfr47, Olfr518, Olfr739, Olfr898, Olfr920, Olfr958, Olfr960, Olfr961, Olfr978, Olfr982, Olfr985, Olfr616, Olfr911-ps1, Olfr620, Olfr214, Olfr145, Olfr1477, Olfr378, Olfr478, Olfr494, Olfr502, Olfr874, Olfr875, Olfr888, Olfr889, Olfr1009, Olfr1044, Olfr1047, Olfr1054, Olfr1094, Olfr1170, Olfr1238, Olfr133, Olfr1333, Olfr136, Olfr1370, Olfr143, Olfr1444, Olfr1448, Olfr1463, Olfr1469, Olfr1484, Olfr151, Olfr160, Olfr166, Olfr19, Olfr191, Olfr196, Olfr201, Olfr203, Olfr205, Olfr346, Olfr347, Olfr429, Olfr430, Olfr44, Olfr466, Olfr490, Olfr51, Olfr60, Olfr736, Olfr744, Olfr746, Olfr887, Olfr890, Olfr904, Olfr923, Olfr935, Olfr1442, Olfr1461, Olfr150, Olfr1501, Olfr211, Olfr357, Olfr539, Olfr61, Olfr919, Olfr123, Olfr1364, Olfr561, Olfr67, Olfr1446, Olfr2, Olfr33, Olfr5, Olfr535, Olfr549, Olfr617, Olfr628, Olfr677, Olfr702, Olfr558, Olfr609, Olfr611	2-4, 5, 16, 18-43

Table S2: mOR256-31 residues in direct contact with the odorants.

Ballesteros-Weinstein nomenclature	Occupation (%) during MD ^a			Literature data (mutagenesis)	
	Coumarin	R-carvone	Acetophenone	Receptor ^b	Ref.
2.53	100	38	44		
3.29	0	72	49	hOR1A1	
3.30	0	17	0	mOR256-3, mOR256-8	
3.32	100	100	96	hOR1A1, hOR2AG1, mOR-EG, mOR256-3	
3.33	100	100	100	hOR1A1, mOR-EG, mOR42-3	
3.36	100	99	99	hOR1A1, hOR1A2, mOR-EG, mOR42-3, mOR256-3	
3.37	9	81	97	hOR1A1, hOR1A2, mOR42-3, mOR244-3	
3.38	0	11	1		
5.42	26	0	0	hOR1A1, mOR-EG, mOR42-3	1, 18, 22, 44-50
5.43	0	74	99	mOR42-3, mOR256-3	
5.46	94	71	59	hOR1A1, hOR1A2, hOR2AG1, mOR-EG, mOR42-3	
5.47	0	89	97	hOR1A1, mOR-EG	
6.48	100	85	97	hOR1A1, hOR7D4, mOR256-3	
6.51	100	100	73	hOR1A1, hOR1A2, mOR-EG, mOR42-3	
6.52	0	96	100		
6.55	0	97	97	mOR-EG	
7.42	100	30	1	hOR1A1, hOR1A2, hOR2AG1, mOR-EG	

^a Percentage of the simulation trajectory where the residue was within 5 Å distance of the odorants.

^b Receptor response to odorants was affected upon mutation.[24, 27, 47, 47–50]

Table S3: Six subsets of residues tested in machine learning of OR response to odorants.

Residue subset	Number of residues	Ballesteros-Weinstein nomenclature
<i>poc17</i>	17	2.53 3.29 3.30 3.32 3.33 3.36 3.37 3.38 5.42 5.43 5.46 5.47 6.48 6.51 6.52 6.55 7.42
<i>poc20</i>	20	2.53 2.54 3.29 3.30 3.32 3.33 3.36 3.37 3.38 5.38 5.39 5.42 5.43 5.46 5.47 6.48 6.51 6.52 6.55 7.42
<i>poc27</i>	27	2.53 2.54 3.29 3.30 3.31 3.32 3.33 3.36 3.37 3.38 3.40 3.41 4.56 5.37 5.38 5.39 5.42 5.43 5.46 5.47 6.44 6.48 6.51 6.52 6.55 7.38 7.42
<i>poc60</i>	60	2.53 2.54 2.57 2.61 3.25 3.26 3.27 3.28 3.29 3.30 3.31 3.32 3.33 3.34 3.35 3.36 3.37 3.38 3.39 3.40 3.41 4.52 4.53 4.56 4.57 4.60 5.35 5.36 5.37 5.38 5.39 5.40 5.41 5.42 5.43 5.44 5.45 5.46 5.47 5.48 5.49 5.50 5.51 6.44 6.45 6.47 6.48 6.49 6.50 6.51 6.52 6.53 6.54 6.55 6.56 7.35 7.38 7.39 7.41 7.42
<i>TM191</i>	191	1.32–1.65, 2.39–2.67, 3.25–3.55, 4.40–4.61, 5.37–5.65, 6.32–6.56, 7.35–7.55

Table S4: Performance of different RF classifiers in predicting new OR-odorant pairs (see Fig. S3 and Data file S2C for the raw data).

Model	Metrics	Acetophenone	R-carvone	Coumarin	4-chromanone
<i>poc17</i>	MCC	0.26	0.16	0.21	0.07
	Precision	0.24	0.3	0.36	0.2
	Recall	0.6	0.23	0.31	0.17
	F1	0.34	0.26	0.33	0.18
	Specificity	0.77	0.91	0.88	0.9
	AUC	0.78	0.68	0.66	0.7
	True positives	6	3	5	2
	True negatives	64	73	68	73
	False positives	19	7	9	8
	False negatives	4	10	11	10
<i>poc20</i>	MCC	0.23	0.28	0.26	0.25
	Precision	0.25	0.44	0.43	0.4
	Recall	0.55	0.31	0.35	0.31
	F1	0.34	0.36	0.39	0.35
	Specificity	0.77	0.93	0.89	0.92
	AUC	0.8	0.65	0.63	0.7
	True positives	6	4	6	4
	True negatives	59	70	63	69
	False positives	18	5	8	6
	False negatives	5	9	11	9
<i>poc27</i>	MCC	0.41	0.28	0.26	0.31
	Precision	0.37	0.5	0.45	0.5
	Recall	0.7	0.25	0.31	0.31
	F1	0.48	0.33	0.37	0.38
	Specificity	0.83	0.96	0.91	0.94
	AUC	0.85	0.67	0.72	0.75
	True positives	7	3	5	4
	True negatives	60	67	60	65
	False positives	12	3	6	4
	False negatives	3	9	11	9
<i>TM191</i>	MCC	0.33	0.29	0.31	0.29
	Precision	0.32	0.4	0.47	0.4
	Recall	0.64	0.43	0.44	0.43
	F1	0.42	0.41	0.46	0.41
	Specificity	0.79	0.87	0.86	0.87
	AUC	0.75	0.7	0.76	0.74
	True positives	7	6	8	6
	True negatives	57	60	56	60
	False positives	15	9	9	9
	False negatives	4	8	10	8

Table S5: Newly identified OR-odorant pairs and EC50 (with 95% confidence interval) ^a in Hana3A cells.

EC50 (μM)	acetophenone	R-carvone	coumarin	4-chromanone
hOR8B2	-	-	-	15 (2–101)
Olf1057	83 (39–170)	-	4 (1–125)	6 (3–13)
Olf147	-	~0.4 mM	0.6 (0.1–31)	-
Olf1487	-	407 (231–794)	-	-
Olf160	183 (82–442)	270 (55–2494)	1964 (423–?)	58 (18–194)
Olf194	775 (335–?)	10 (5–20)	1164 (359–2708)	11 (2–57)
Olf205	47 (20–106)	45 (19–115)	51 (22–125)	93 (53–165)
Olf251	193 (130–289)	159 (50–575)	71 (39–128)	1293 (416–?)
Olf338	995 (50–190)	157 (97–257)	509 (280–1085)	217 (95–528)
Olf747	858 (363–4395)	142 (60–347)	119 (70–204)	~0.6 mM
Olf987	-	883 (542–1701)	-	-
Olf988	-	~0.7 mM	-	-
hOR11H12	5 (0.7–30)	16 (6–48)	92 (17–507)	-
hOR5R1	3 (1–137)	12 (3–33)	10 (2–64)	218 (11–4506)
hOR8K5	-	5 (1–21)	-	-
Olf1016	-	-	72 (14–336)	-
Olf1097	57 (15–227)	-	7 (1–54)	10 (2–50)
Olf1156	46 (10–2238)	-	-	-
Olf285	-	-	67 (32–138)	738 (15–367)
Olf924	-	-	28 (5–173)	-
Olf190	-	~2 mM	-	-
Olf197	-	-	-	52 (4–635)
Olf198	-	~1 mM	-	-
Olf994	-	-	4532 (953–?)	2199 (748–?)
EC50 (μM)	Citral	Nonanal	Nonanoic acid	
hOR10G8	-	61 (36–105)	-	
hOR51A4	235 (87–823)	-	-	
hOR52A1	-	~0.5 M	-	
hOR52B2	~18 M	-	-	
Olf591	184 (141–230)	-	-	
Olf987	384 (140–1223)	105 (76–147)	22 (12–39)	
Olf988	-	-	~0.6 mM	

^a in μM unless otherwise indicated. “?” means undetermined. “-” means no significant response up to 1 mM.

Table S6: Performance of the *poc60* model on 5 folds of the *in vitro* test set.

Test set	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
MCC	0.34	0.34	0.32	0.32	0.34
Precision	0.43	0.46	0.40	0.40	0.39
Recall	0.46	0.43	0.46	0.46	0.54
F1	0.44	0.44	0.43	0.43	0.45
Specificity	0.89	0.91	0.88	0.88	0.85
AUC	0.74	0.69	0.64	0.72	0.73
TP	6	6	6	6	7
TN	66	67	66	65	64
FP	8	7	9	9	11
FN	7	8	7	7	6

Table S7: Performance of the *poc60* model on “new” ORs that are not part of the training set.

	Acetophenone	R-carvone	Coumarin	4-chromanone	Citral	Nonanal	Nonanoic acid	All
MCC	0.50	0.50	0.32	0.21	0.52	0.26	0.38	0.34
Precision	0.33	0.67	0.50	0.33	1.00	0.33	0.29	0.41
Recall	1.00	0.50	0.45	0.38	0.33	0.50	1.00	0.54
F1	0.50	0.57	0.48	0.35	0.50	0.40	0.44	0.47
Specificity	0.76	0.95	0.86	0.84	1.00	0.80	0.50	0.83
AUC	0.90	0.78	0.67	0.59	0.69	0.52	0.65	0.70
TP	5	4	5	3	1	1	2	21
TN	31	36	30	32	9	8	5	151
FP	10	2	5	6	0	2	5	30
FN	0	4	6	5	2	1	0	18

Table S8: Performance of the *poc60* model on “new” odorants by excluding the query odorant from the training set.

	Acetophenone	R-carvone	Coumarin	4-chromanone	Citral	Nonanal	Nonanoic acid
MCC	0.31	0.3	NA	0.31	0.04	NA	NA
Precision	0.67	0.5	NA	0.4	0.25	NA	NA
Recall	0.18	0.29	0	0.43	0.25	0	0
F1	0.29	0.36	NA	0.41	0.25	NA	NA
Specificity	0.99	0.95	1	0.89	0.79	1	1
AUC	0.45	0.73	0.43	0.69	0.54	0.19	0.26
TP	2	4	0	6	1	0	0
TN	84	78	78	73	11	16	15
FP	1	4	0	9	3	0	0
FN	9	10	18	8	3	2	3

Table S9: List of primers used for site directed mutagenesis of mOR256-31.

Mutant	Forward primer (5'-3')	Reverse primer (5'-3')
N137H	CTTGCACTACTTTGTAATCATGCACCCACGCTCATGCGTCAAGATGA	TCATCTTGACGCATAGACGTGGGTGCATGATTACAAGTAATGCAAG
L107I	GCTCTATGTGTATATGTGGATTGGCTCCATTGAGTGTCTGC	GCAGACACTCAATGGAGCCAATCCACATATACACATAGAGC
T192A	GTCAGGATAGCTTGTGTAGACGCTACAACAGTTGAATTGTCTGTT	AACAGACAATTCAACTGTTGTAGCGTCTACACAAGCTATCCTGAC
Y102F	GGCTGATTTGTCAGCTCTTCGTGTATATGTGGCTGGGCT	AGCCAGCCACATATACACGAAGAGCTGAACAATACAGCC
S156A	GGGGTATTAGTTTGGCCAAACGCGTAATATTATGCACACTCAC	GTGAGTGTGCATAATATTACGGCGTTGGCCAAACTAATACCCC
D175N	ATGTGGACACAACATCCTGAATCACTTCCTGTGTGAGTTGC	GCAACTCACACAGGAAGTGATTAGGATGTTGTGCCACAT
G108A	GTGTATATGTGGCTGGCCCTCCATTGAGTGTCTG	CAGACACTCAATGGAGGCCACCATATACAC
T161P	CCAACCTCTGTAATATTATGCCCACTCACTGTGAATTTGCCCT	AGGCAAATTCACAGTGTGAGTGGGCATAATATTACAGAGTTGG
P167R	ACACTCACTGTGAATTTGGCAGCATGTGGACACAACATCC	GGATGTTGTGCCACATCGTCGCAAATTCACAGTGTGAGTGT
L114I	GGGCTCCATTGAGTGTCTGATTCTAGCTGTCTATGCTCCTATG	CATAGGACATGACAGCTAGAATCAGACACTCAATGGAGCCC
T163A	CTGTAATATTATGCACACTCGCAGTGAATTTGCCCTGATGTGG	CCACATCGAGGCAAATTCAGTGTGCATAATATTACAG
L39V	GTTGTCACCTCTTTTACATTATTACAGTGGTGAAGCAACT	AGTGTATCCCACTGTAATAATGTAAGAAGAGGTTGACAAC
L47V	GGTGGGTAACACTGCTATAATTGTTGCATCCCTCC	GGAGGGATGCAACAATTATAGCAGTGTATCCACC
Y104F	GTTACGCTCTATGTGTTATGTGGCTGGGCTCC	GGAGCCAGCCACATAAACACATAGAGCTGAAC
Y104A	GTATTGTTACGCTCTATGTGGCTATGTGGCTGGGCTCCATTG	CAATGGAGCCCAGCCACATAGCCACATAGAGCTGAACAATAC
G108L	CTATGTGTATATGTGGCTGCTCCATTGAGTGTCTGCTC	GAGCAGACACTCAATGGAGAGCAGCCACATATACACATAG
G108I	CTATGTGTATATGTGGCTGATCTCCATTGAGTGTCTGCTC	GAGCAGACACTCAATGGAGATCAGCCACATATACACATAG
T161A	CCAACCTCTGTAATATTATGCCCACTCACTGTGAATTTGCCCT	AGGCAAATTCACAGTGTGAGTGGGCATAATATTACAGAGTTGG
S198L	GACAATGCCTAGACAAAAACAAGCAATTCACCTGTTGCTGTCTA	TAGACACGACAACAGTTGAATTTGCTGTTTTTGTCTAGGCATTGTC
V199M	CACGACAACAGTGAATTGCTATGTTTGTCTTAGGCATTGCTATTG	CAATGACAATGCCTAGAGCAAACATAGACAATTCACCTGTTGCTGTG
V199L	CGACAACAGTGAATTGCTCTTTTTGCTCTAGGCATTGTC	GACAATGCCTAGAGCAAAAAGAGACAATTCACCTGTTGCTG
S49C	ACACTGCTATAAATCTTGCATGCCCTCTGGATCC	GGATCCAGGAGGCATGCAAGAATTATAGCAGTGT
N65H	CCAATGTAATTTTCTCAGGCATCTATCTTCTCGGATTTGT	ACAATCCAGGAAAGATAGATGCCTGAGGAAAAAGTACATTGG
T74S	TTCTGGATTGTGTTTCTCAACAAGCAITGTCCTC	GAGGACAATGCTTTGTTGAGAAAACAAAATCCAGGAA
M81L	CAAGCAITGTCCTCAGCTGTGGTTAACTTGTGG	CCACAAGTTAACCAGCAGCTGAGGGACAATGCTTG
T163V	TCCACATCGAGGCAAATTCACAACGAGTGTGCATAATATTACAGAG	CTCTGTAATATTATGCACACTCGTTGTGAATTTGCCCTCGATGTGGA
T193S	GATAGCTTGTGTAGACACGTCACAGTTGAATTTGCTGT	ACAGACAATTCACCTGTTGACGTGTCTACACAAGCTATC
G203A	GTGTAAGGACAATGACAATGGCTAGAGCAAAAACAGACAAT	ATTGCTGTTTTTGTCTAGCCATTGCTATTGCTCTTACAC
T209L	AATAAGGATGAGAGGTAAGGACAATGACAATGCCTAGAGCAAAAAC	GTTTTTGTCTAGGCATTGTCAATTGCTTTTTTACCTCTCATCTTATT
N239G	GAGATGGGATCCACAGGTACCCATTGCTTTTTTGTGCCCT	AGGGCAACAAAAAGCAATGGGTACCTGTGGATCCCATCTC
T194P	ATAGCTTGTGTAGACACGACACAGTTGAATTTGCTGTTTTTG	CAAAAACAGACAATTCACCTGGTGTGCTGTCTACACAAGCTAT
T254S	GTGGTCTCAATTTCTATGATCTATCATCTACATGATCTAC	GTAGATACATGATGATAGATCCATAGAATATGGAGACCAC

Supplementary Methods

Molecular dynamics simulations.

Each OR-odorant complex was embedded in a bilayer of POPC using PACKMOL-Memgen [51]. The system was solvated in a periodic $75 \times 75 \times 105 \text{ \AA}^3$ box of explicit water and neutralized with 0.15 M of Na^+ and Cl^- ions. Effective point charges of the ligands were obtained by RESP fitting [52] of the electrostatic potentials calculated with the HF/6-31G* basis set using Gaussian 09 [53]. The Amber 99SB-ildn [54], lipid 14 [55] and GAFF [56] force fields were used for the proteins, the lipids and the ligands, respectively. The TIP3P [57] and the Joung-Cheatham [58] models were used for the water and the ions, respectively. After energy minimization, all-atom MD simulations were carried out using Gromacs 5.1 [59] patched with the PLUMED 2.3 plugin [60]. Each system was gradually heated to 310 K and pre-equilibrated during 10 ns of brute-force MD in the *NPT*-ensemble. The replica exchange with solute scaling (REST2) [61] technique was employed to enhance the sampling with 48 replicas in the *NVT* ensemble. The protein and the ligands were considered as “solute” in the REST2 scheme—force constants of their van der Waals, electrostatic and dihedral terms were subject to scaling. The effective temperatures used for generating the REST2 scaling factors ranged from 310 K to 700 K, following a distribution calculated with the Patriksson-van der Spoel approach [62]. Exchange between replicas was attempted every 1000 simulation steps. This setup resulted in an average exchange probability of ~40%. The original unscaled replica (at 310 K effective temperature) was collected and analyzed. The first 10 ns were discarded for equilibration.

Machine learning.

Human and mouse ORs were presented by their aligned amino acid sequences from our previous work [23]. The sequence alignment of 1733 ORs including 282 polymorphisms and mutants were hand curated. Amino acid positions that contain gaps in the sequence alignment were removed. The remaining sequence alignment contained 191 residues in the TM domain and 23 residues in the ECL2. Each amino acid was converted to 3 physicochemical features [63] amino acid composition, polarity and volume. Features showing low variance (< 5%) or high correlation (Pearson’s correlation coefficient > 0.8) across the entire OR set were removed.

Finally, 398 features remained, and the full OR data set was a 1733×398 matrix. The odorant data set contained 244 odorants from OR-odorant functional assays in the literature (Data file S1). Chemical similarities among the odorants were calculated with the Tanimoto similarity index (ranging from 0 to 1) to identify analogs. The Dragon software [64] ⁶⁵ was used to calculate 3850 physicochemical features for each odorant from the SMILES string. Thus, the full odorant data set was a 244×3850 matrix. To construct the training set for each machine learning model, the query odorants and their analogs are extracted from the odorant data set matrix. Note that the Tanimoto cutoff for analogs is a model parameter that can be adjusted according to available data: higher cutoff leads to more relevant analogs but less training data. For the odorants studied here, we used the cutoffs ranging from 0.6 to 0.7 as good tradeoffs. After removing low-variance and highly correlated features from the odorant data set, it was merged with the OR set to generate a PCM matrix of OR-odorant pairs. Each row of the matrix corresponded to a unique OR-odorant pair. The last column of the PCM matrix contained the labels of the response data, “positive” for responsive pairs and “negative” for non-responsive pairs. The test set of OR-odorant pairs was presented in the same manner in a PCM matrix, without the column that contained the response labels.

Model training and parameter tuning were carried out by nested 5-fold cross validation, in which the validation sets contained only the query odorants (not their analogs) and their paired ORs (Supplementary Fig. 2A). This was because the models were to be assessed with the query odorants and not their analogs. The analogs were used only for model training, to augment the amount of training data. We first trained 5 independent models using different residue subsets, *poc17*, *poc20*, *poc27*, *poc60*, and *TM191*. For each model, only the corresponding residue columns in the PCM matrix were used, together with the columns containing the odorant features and the response label (Fig. 1A). An alternative model using amino acid conservations in the TM region was built by replacing the amino acid features in the *TM191* model by the conservation scores at each amino acid column in the OR sequence alignment. These were compared with two control models: a “random” model and a “naive” statistical inference (Supplementary Fig. 2A). The “random” model was generated with randomized pseudo sequences of 60 amino acids, which should contain no signal or predictive power. The naive baseline was calculated by assuming that the responsiveness of an OR-odorant pair is the average of the OR's responsiveness to all the odorants and the odorant's activity on all the ORs. In other words, this assumed that promiscuous ORs and odorants had high probabilities to form responsive pairs. For example, among all the responsive OR-odorant pairs in the dataset, hOR1A1 responds to 25% of the odorants, while acetophenone activates 14% of the ORs, thus

the hOR1A1-acetophenone pair has assumedly $(25\% + 14\%) / 2 = 19.5\%$ probability to be responsive. Normalizing this probability (by setting the most promiscuous OR-odorant pair to 1 and the least promiscuous to 0) gave the naive model in Supplementary Fig. 2A.

Each of the 5 models was used to predict and rank the ORs by their probability to respond to each odorant. A probability > 0.5 was classified as responsive. The prediction of each model was then assessed independently by *invitro* functional assays. For each model, the predicted responsive ORs were tested on all the query odorants. Negative control ORs were randomly picked from those with response probabilities ≤ 0.5 for all the odorants. They were also tested on all the query odorants. The procedure was repeated on all the models built with different residue subsets (Supplementary Fig. 2A). This led to *in vitro* functional assays on a total of 384 OR-odorant pairs.

To estimate how the model predictivity varies with the amount of available training data, we generated a learning curve (Supplementary Fig. 4) of the *poc60* model (the best model). We focused on acetophenone, R-carvone, coumarin and 4-chromanone for which there were enough learning data. Each of the test models was trained with randomly sampled subsets of the training data. Each data subset contained responsive and non-responsive OR-odorant pairs in a 1:3 ratio, in order to mimic the ratio in the full data base. The performance of these models was evaluated with the *in vitro* test data, to estimate the minimum amount of training data needed to obtain a predictive model.

The following performance metrics were used to evaluate the models. Matthew's correlation coefficient (MCC) [28] was used as the main metrics of predictivity for the classification models. It is generally considered more informative than other confusion matrix measures (e.g. accuracy, precision, recall and F1 score), especially for highly imbalanced data as is the case here. MCC returns a value between -1 and $+1$, where $+1$ indicates a perfect prediction, 0 a random prediction, and -1 an inverse prediction. It is calculated as:

$$(TP \times TN - FP \times FN) / [(TP + FP)(TP + FN)(TN + FP)(TN + FN)]^{1/2},$$

where *TP*, *TN*, *FP* and *FN* are the numbers of true positives, true negatives, false positives and false negatives, respectively. Other metrics used here were hit rate (or precision) = $TP / (TP + FP)$, recall = $TP / (TP + FN)$, specificity = $TN / (TN + FP)$, and AUC (area under the receiver operating characteristics curve).

Data file S1. Database of *in vitro* responses of 708 ORs to 244 odorants.

Data file S2. Prediction and *in vitro* assessments of OR responses to the 4 odorants.

Author Information

Corresponding Authors

Xiaojing Cong - Université Côte d'Azur, CNRS, Institut de Chimie de Nice UMR7272, Nice 06108, France; Present Address: X.C.: Institut de Génomique Fonctionnelle, University of Montpellier, CNRS, INSERM, 34094 Montpellier, France; Orcid<https://orcid.org/0000-0002-5051-2392>; Email: xiaojing.cong@igf.cnrs.fr

Yiqun Yu - Ear, Nose & Throat Institute, Department of Otolaryngology, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai 200031, People's Republic of China; Clinical and Research Center for Olfactory Disorders, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai 200031, People's Republic of China; Email: yu_yiqun@fudan.edu.cn

Jérôme Golebiowski - Université Côte d'Azur, CNRS, Institut de Chimie de Nice UMR7272, Nice 06108, France; Department of Brain and Cognitive Sciences, Daegu Gyeongbuk Institute of Science and Technology, Daegu 711-873, South Korea; Orcid<https://orcid.org/0000-0002-3675-1952>; Email: jerome.golebiowski@gmail.com

Authors

Wenwen Ren - Institutes of Biomedical Sciences, Fudan University, Shanghai 200031, People's Republic of China

Jody Pacalon - Université Côte d'Azur, CNRS, Institut de Chimie de Nice UMR7272, Nice 06108, France

Rui Xu - School of Life Sciences, Shanghai University, Shanghai 200444, People's Republic of China

Lun Xu - Ear, Nose & Throat Institute, Department of Otolaryngology, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai 200031, People's Republic of China

Xuewen Li - School of Life Sciences, Shanghai University, Shanghai 200444, People's Republic of China

Claire A. de March - Department of Molecular Genetics and Microbiology, and Department of Neurobiology, and Duke Institute for Brain Sciences, Duke University Medical Center, Research Drive, Durham, North Carolina 27710, United States

Hiroaki Matsunami - Department of Molecular Genetics and Microbiology, and Department of Neurobiology, and Duke Institute for Brain Sciences, Duke University Medical Center, Research Drive, Durham, North Carolina 27710, United States

Hongmeng Yu - Ear, Nose & Throat Institute, Department of Otolaryngology, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai 200031, People's Republic of China; Clinical and Research Center for Olfactory Disorders, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai 200031, People's Republic of China; Research Units of New Technologies of Endoscopic Surgery in Skull Base Tumor, Chinese Academy of Medical Sciences, Beijing 100730, People's Republic of China

Author Contributions

X.C. and W.R. contributed equally. X.C., H.M., Y.Y., and J.G. designed the research. X.C. and J.P. collected and analyzed literature data. X.C., J.P., and J.G. performed and analyzed *in silico* experiments. W.R., R.X., L.X., X.L., C.A.d.M., and Y.Y. performed *in vitro* experiments. W.R., R.X., L.X., H.Y., C.A.d.M., H.M., and Y.Y. analyzed *in vitro* data. X.C., Y.Y., and J.G. wrote the paper. All authors have given approval to the final version of the manuscript.

Notes

The authors declare the following competing financial interest(s): H.M. has received royalties from ChemCom, research grants from Givaudan, and consultant fees from Kao.

Acknowledgments

This work received funding from the French National Research Agency (ANR, grant NEUROLF to J.G.) as part of the ANR-NSF-NIH Collaborative Research in Computational Neuroscience; the French government, through the UCA-Jedi project managed by the ANR (ANR-15-IDEX-01 to X.C. and J.G.) and, in particular, by the interdisciplinary Institute for Modeling in Neuroscience and Cognition (NeuroMod) of the Université Côte d'Azur; the German Research Foundation (DFG, grant CO 1715/1-1 to X.C.); the Roudnitska Foundation (France, a PhD fellowship to J.P.); GIRACT (Switzerland, a research fellowship to J.P.); the National Institutes of Health (NIH, grants DC016224 to H.M. and K99/R00DC018333 to C.D.M.); the National Science Foundation (NSF, grant 1515801 to H.M.); the National Natural Science Foundation of China (Grants 32070996 and 31771155 to Y.Y., 31900714 to W.R.); Science and Technology Commission of Shanghai Municipality (grant 21140900600 to Y.Y.); Shanghai Municipal Human Resources and Social Security Bureau (Shanghai Talent Development Fund 2018112 to Y.Y.); Eye, Ear, Nose and Throat Hospital, Fudan University (Excellent Doctors-Excellent Clinical Researchers Program SYB202002 to Y.Y.); and the New Technologies of Endoscopic Surgery in Skull Base Tumor: CAMS Innovation Fund for Medical Sciences (CIFMS; 2019-I2M-5-003 to H.Y.). This work was granted access to the HPC resources of CINES under the allocation 2018–2019 A0040710477 made by GENCI. We also thank Dr. Sébastien Fiorucci, Dr. Jérémie Topin, Dr. Cédric Bouysset, Priyanka Meesa, and Shiyi Jiang for critically reading and editing the manuscript.

References

1. Santos R, Ursu O, Gaulton A, et al (2017) A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* 16:19–34. <https://doi.org/10.1038/nrd.2016.230>
2. Fredriksson R, Lagerström MC, Lundin L-G, Schiöth HB (2003) The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol Pharmacol* 63:1256–1272. <https://doi.org/10.1124/mol.63.6.1256>
3. Bushdid C, Magnusco MO, Vosshall LB, Keller A (2014) Humans Can Discriminate More than 1 Trillion Olfactory Stimuli. *Science* 343:1370–1372. <https://doi.org/10.1126/science.1249168>

4. Cong X, Topin J, Golebiowski J Class A GPCRs: Structure, Function, Modeling and Structure-based Ligand Design. *Current Pharmaceutical Design* 23:4390–4409
5. Venkatakrishnan AJ, Deupi X, Lebon G, et al (2013) Molecular signatures of G-protein-coupled receptors. *Nature* 494:185–194. <https://doi.org/10.1038/nature11896>
6. Malnic B, Hirono J, Sato T, Buck LB (1999) Combinatorial Receptor Codes for Odors. *Cell* 96:713–723. [https://doi.org/10.1016/S0092-8674\(00\)80581-4](https://doi.org/10.1016/S0092-8674(00)80581-4)
7. Tcatchoff L, Nespoulous C, Pernollet J-C, Briand L (2006) A single lysyl residue defines the binding specificity of a human odorant-binding protein for aldehydes. *FEBS Letters* 580:2102–2108. <https://doi.org/10.1016/j.febslet.2006.03.017>
8. Briand L, Eloit C, Nespoulous C, et al (2002) Evidence of an Odorant-Binding Protein in the Human Olfactory Mucus: Location, Structural Characterization, and Odorant-Binding Properties. *Biochemistry* 41:7241–7252. <https://doi.org/10.1021/bi015916c>
9. Lee S-J, Depoortere I, Hatt H (2019) Therapeutic potential of ectopic olfactory and taste receptors. *Nat Rev Drug Discov* 18:116–138. <https://doi.org/10.1038/s41573-018-0002-3>
10. Kang N, Kim H, Jae Y, et al (2015) Olfactory Marker Protein Expression Is an Indicator of Olfactory Receptor-Associated Events in Non-Olfactory Tissues. *PLOS ONE* 10:e0116097. <https://doi.org/10.1371/journal.pone.0116097>
11. Kang N, Koo J (2012) Olfactory receptors in non-chemosensory tissues. *BMB Reports* 45:612–622. <https://doi.org/10.5483/BMBRep.2012.45.11.232>
12. Ferrer I, Garcia-Esparcia P, Carmona M, et al (2016) Olfactory Receptors in Non-Chemosensory Organs: The Nervous System in Health and Disease. *Frontiers in Aging Neuroscience* 8:
13. Lapinsh M, Prusis P, Gutcaits A, et al (2001) Development of proteo-chemometrics: a novel technology for the analysis of drug-receptor interactions. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1525:180–190. [https://doi.org/10.1016/S0304-4165\(00\)00187-2](https://doi.org/10.1016/S0304-4165(00)00187-2)
14. Rifaioglu AS, Atas H, Martin MJ, et al (2019) Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. *Briefings in Bioinformatics* 20:1878–1912. <https://doi.org/10.1093/bib/bby061>
15. Mousavian Z, Masoudi-Nejad A (2014) Drug-target interaction prediction via chemogenomic space: learning-based methods. *Expert Opin Drug Metab Toxicol* 10:1273–1287. <https://doi.org/10.1517/17425255.2014.950222>
16. Ezzat A, Wu M, Li X-L, Kwoh C-K (2019) Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey. *Brief Bioinform* 20:1337–1357. <https://doi.org/10.1093/bib/bby002>
17. Liu X, Su X, Wang F, et al (2011) ODORactor: a web server for deciphering olfactory coding. *Bioinformatics* 27:2302–2303. <https://doi.org/10.1093/bioinformatics/btr385>

18. Jiang Y, Gong NN, Hu XS, et al (2015) Molecular profiling of activated olfactory neurons identifies odorant receptors for odors in vivo. *Nat Neurosci* 18:1446–1454. <https://doi.org/10.1038/nn.4104>
19. Saito H, Chi Q, Zhuang H, et al (2009) Odor Coding by a Mammalian Receptor Repertoire. *Science Signaling* 2:ra9–ra9. <https://doi.org/10.1126/scisignal.2000016>
20. Chepurwar S, Gupta A, Haddad R, Gupta N (2019) Sequence-Based Prediction of Olfactory Receptor Responses. *Chemical Senses* 44:693–703. <https://doi.org/10.1093/chemse/bjz059>
21. Chen K-YM, Keri D, Barth P (2020) Computational design of G Protein-Coupled Receptor allosteric signal transductions. *Nat Chem Biol* 16:77–86. <https://doi.org/10.1038/s41589-019-0407-2>
22. Man O, Gilad Y, Lancet D (2004) Prediction of the odorant binding site of olfactory receptor proteins by human–mouse comparisons. *Protein Science* 13:240–254. <https://doi.org/10.1110/ps.03296404>
23. de March CA, Kim S-K, Antonczak S, et al (2015) G protein-coupled odorant receptors: From sequence to structure. *Protein Science : A Publication of the Protein Society* 24:1543–1548. <https://doi.org/10.1002/pro.2717>
24. de March CA, Topin J, Bruguera E, et al (2018) Odorant Receptor 7D4 Activation Dynamics. *Angewandte Chemie International Edition* 57:4554–4558. <https://doi.org/10.1002/anie.201713065>
25. Yu Y, de March CA, Ni MJ, et al (2015) Responsiveness of G protein-coupled odorant receptors is partially attributed to the activation mechanism. *Proceedings of the National Academy of Sciences* 112:14966–14971. <https://doi.org/10.1073/pnas.1517510112>
26. Nara K, Saraiva LR, Ye X, Buck LB (2011) A Large-Scale Analysis of Odor Coding in the Olfactory Epithelium. *J Neurosci* 31:9179–9191. <https://doi.org/10.1523/JNEUROSCI.1282-11.2011>
27. de March CA, Yu Y, Ni MJ, et al (2015) Conserved Residues Control Activation of Mammalian G Protein-Coupled Odorant Receptors. *J Am Chem Soc* 137:8611–8616. <https://doi.org/10.1021/jacs.5b04659>
28. Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure* 405:442–451. [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9)
29. Katritch V, Fenalti G, Abola EE, et al (2014) Allosteric sodium in class A GPCR signaling. *Trends in Biochemical Sciences* 39:233–244. <https://doi.org/10.1016/j.tibs.2014.03.002>
30. Duan X, Block E, Li Z, et al (2012) Crucial role of copper in detection of metal-coordinating odorants. *Proceedings of the National Academy of Sciences* 109:3492–3497. <https://doi.org/10.1073/pnas.1111297109>
31. Wang J, Luthey-Schulten ZA, Suslick KS (2003) Is the olfactory receptor a metalloprotein? *Proceedings of the National Academy of Sciences* 100:3035–3039. <https://doi.org/10.1073/pnas.262792899>

32. Launay G, Sanz G, Pajot-Augy E, Gibrat J-F (2012) Modeling of mammalian olfactory receptors and docking of odorants. *Biophys Rev* 4:255–269. <https://doi.org/10.1007/s12551-012-0080-0>
33. Geithe C, Protze J, Kreuchwig F, et al (2017) Structural determinants of a conserved enantiomer-selective carvone binding pocket in the human odorant receptor OR1A1. *Cell Mol Life Sci* 74:4209–4229. <https://doi.org/10.1007/s00018-017-2576-z>
34. Sato-Akuhara N, Horio N, Kato-Namba A, et al (2016) Ligand Specificity and Evolution of Mammalian Musk Odor Receptors: Effect of Single Receptor Deletion on Odor Detection. *J Neurosci* 36:4482–4491. <https://doi.org/10.1523/JNEUROSCI.3259-15.2016>
35. Cong X, Zheng Q, Ren W, et al (2019) Zebrafish olfactory receptors ORAs differentially detect bile acids and bile salts. *Journal of Biological Chemistry* 294:6762–6771. <https://doi.org/10.1074/jbc.RA118.006483>
36. Yuan S, Dahoun T, Brugarolas M, et al (2019) Computational modeling of the olfactory receptor Olfr73 suggests a molecular basis for low potency of olfactory receptor-activating compounds. *Commun Biol* 2:1–10. <https://doi.org/10.1038/s42003-019-0384-8>
37. Launay G, Téletchéa S, Wade F, et al (2012) Automatic modeling of mammalian olfactory receptors and docking of odorants. *Protein Engineering, Design and Selection* 25:377–386. <https://doi.org/10.1093/protein/gzs037>
38. Caballero-Vidal G, Bouysset C, Grunig H, et al (2020) Machine learning decodes chemical features to identify novel agonists of a moth odorant receptor. *Sci Rep* 10:1655. <https://doi.org/10.1038/s41598-020-58564-9>
39. Bushdid C, de March CA, Fiorucci S, et al (2018) Agonists of G-Protein-Coupled Odorant Receptors Are Predicted from Chemical Features. *J Phys Chem Lett* 9:2235–2240. <https://doi.org/10.1021/acs.jpcclett.8b00633>
40. Muratov EN, Bajorath J, Sheridan RP, et al (2020) QSAR without borders. *Chem Soc Rev* 49:3525–3564. <https://doi.org/10.1039/D0CS00098A>
41. Ikegami K, de March CA, Nagai MH, et al (2020) Structural instability and divergence from conserved residues underlie intracellular retention of mammalian odorant receptors. *Proceedings of the National Academy of Sciences* 117:2957–2967. <https://doi.org/10.1073/pnas.1915520117>
42. Zhuang H, Matsunami H (2008) Evaluating cell-surface expression and measuring activation of mammalian odorant receptors in heterologous cells. *Nat Protoc* 3:1402–1413. <https://doi.org/10.1038/nprot.2008.120>
43. Eswar N, Webb B, Marti-Renom MA, et al (2006) Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics* 15:5.6.1–5.6.30. <https://doi.org/10.1002/0471250953.bi0506s15>
44. Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31:455–461. <https://doi.org/10.1002/jcc.21334>

45. van Zundert GCP, Rodrigues JPGLM, Trellet M, et al (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* 428:720–725. <https://doi.org/10.1016/j.jmb.2015.09.014>
46. Kuhn M (2008) Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28:1–26. <https://doi.org/10.18637/jss.v028.i05>
47. Sekharan S, Ertem MZ, Zhuang H, et al (2014) QM/MM Model of the Mouse Olfactory Receptor MOR244-3 Validated by Site-Directed Mutagenesis Experiments. *Biophys J* 107:L5–L8. <https://doi.org/10.1016/j.bpj.2014.07.031>
48. Wolf S, Jovancevic N, Gelis L, et al (2017) Dynamical Binding Modes Determine Agonistic and Antagonistic Ligand Effects in the Prostate-Specific G-Protein Coupled Receptor (PSGR). *Sci Rep* 7:16007. <https://doi.org/10.1038/s41598-017-16001-4>
49. Katada S, Hirokawa T, Oka Y, et al (2005) Structural Basis for a Broad But Selective Ligand Spectrum of a Mouse Olfactory Receptor: Mapping the Odorant-Binding Site. *J Neurosci* 25:1806–1815. <https://doi.org/10.1523/JNEUROSCI.4723-04.2005>
50. Abaffy T, Malhotra A, Luetje CW (2007) The molecular basis for ligand specificity in a mouse olfactory receptor: a network of functionally important residues. *J Biol Chem* 282:1216–1224. <https://doi.org/10.1074/jbc.M609355200>
51. Schott-Verdugo S, Gohlke H (2019) PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. *J Chem Inf Model* 59:2522–2528. <https://doi.org/10.1021/acs.jcim.9b00269>
52. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* 21:1049–1074. [https://doi.org/10.1002/1096-987X\(200009\)21:12<1049::AID-JCC3>3.0.CO;2-F](https://doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F)
53. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A., et al., (2009) Gaussian 09, Revision A.02. Wallingford CT,
54. Lindorff-Larsen K, Piana S, Palmo K, et al (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78:1950–1958. <https://doi.org/10.1002/prot.22711>
55. Dickson CJ, Madej BD, Skjevik ÅA, et al (2014) Lipid14: The Amber Lipid Force Field. *J Chem Theory Comput* 10:865–879. <https://doi.org/10.1021/ct4010307>
56. Wang J, Wolf RM, Caldwell JW, et al (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174. <https://doi.org/10.1002/jcc.20035>
57. Jorgensen WL, Chandrasekhar J, Madura JD, et al (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935. <https://doi.org/10.1063/1.445869>

58. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations | The Journal of Physical Chemistry B. <https://pubs.acs.org/doi/10.1021/jp8001614>. Accessed 16 Sep 2022
59. Van Der Spoel D, Lindahl E, Hess B, et al (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701–1718. <https://doi.org/10.1002/jcc.20291>
60. Tribello GA, Bonomi M, Branduardi D, et al (2014) PLUMED 2: New feathers for an old bird. *Computer Physics Communications* 185:604–613. <https://doi.org/10.1016/j.cpc.2013.09.018>
61. Wang L, Friesner RA, Berne BJ (2011) Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J Phys Chem B* 115:9431–9438. <https://doi.org/10.1021/jp204407d>
62. Patriksson A, Spoel D van der (2008) A temperature predictor for parallel tempering simulations. *Phys Chem Chem Phys* 10:2073–2077. <https://doi.org/10.1039/B716554D>
63. Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* 185:862–864. <https://doi.org/10.1126/science.185.4154.862>
64. Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R., DRAGON software: An easy approach to molecular descriptor calculations. *MATCH Communications in Mathematical and in Computer Chemistry* 2006, 56, 237-248

Publication 7

Smell and taste changes are early indicators of the COVID-19 pandemic and political decision effectiveness

*Denis Pierron, Veronica Pereda-Loth, Marylou Mantel, Maëlle Moranges, Emmanuelle Bignon, Omar Alva, Julie Kabous, Margit Heiske, **Jody Pacalon**, Renaud David, Caterina Dinnella, Sara Spinelli, Erminio Monteleone, Michael C Farruggia, Keiland W Cooper, Elizabeth A Sell, Thierry Thomas-Danguin, Alyssa J Bakke, Valentina Parma, John E Hayes, Thierry Letellier, Camille Ferdenzi, Jérôme Golebiowski, Moustafa Bensafi*

Nat Commun. 2020 Oct 14;11(1):5152. doi: 10.1038/s41467-020-18963-y.

Abstract

In response to the COVID-19 pandemic, many governments have taken drastic measures to avoid an overflow of intensive care units. Accurate metrics of disease spread are critical for the reopening strategies. Here, we show that self-reports of smell/taste changes are more closely associated with hospital overload and are earlier markers of the spread of infection of SARS-CoV-2 than current governmental indicators. We also report a decrease in self-reports of new onset smell/taste changes as early as 5 days after lockdown enforcement. Cross-country comparisons demonstrate that countries that adopted the most stringent lockdown measures had faster declines in new reports of smell/taste changes following lockdown than a country that adopted less stringent lockdown measures. We propose that an increase in the incidence of sudden smell and taste change in the general population may be used as an indicator of COVID-19 spread in the population.

Introduction

Following similar decisions in China and Italy, a strict lockdown was enforced in France beginning on March 17, 2020 to block the progression of COVID-19 and alleviate pressure on hospitals. One issue currently faced by governments is how to conduct the progressive relaxation of the lockdown [1], which needs to be conducted systematically and carefully to prevent subsequent outbreaks while facilitating economic activity and recovery. On May 7, 2020, the French government categorized each geographical area as being red or green, depending on their COVID-19 prevalence. Compared to green areas, red areas were characterized by: (i) higher active circulation of the virus, (ii) higher level of pressure on hospitals (i.e., CCRU occupancy), and (iii) reduced capacity to test new cases (Fig. 1a). In each area, red/green labels were used to define steps associated with the local relaxation of lockdown. The French Ministry of Health used the ratio of consultations for suspected cases of COVID-19 to general consultations at the emergency room (ER) in hospitals as an indicator to assess the active circulation of the virus (detailed in “Methods” section). Concurrently, changes in smell and taste are prominent symptoms of COVID-19 [2–5], as has consistently been demonstrated in many countries (e.g., Iran [6], Spain [7], France [8], Italy [9], Germany [10], and the UK [2], among others). More critically, these chemosensory changes generally occur

earlier than other symptoms [9] and may constitute more specific symptoms than fever or dry cough [2, 11]. Accordingly, monitoring self-reported changes in smell and taste could thus provide early and specific information on the spread of COVID-19 in the general population and support health system monitoring to avoid daily CCRU admission overflows. Using data from a global, crowd-sourced study deployed in 30+ languages (Global Consortium for Chemosensory Research survey, GCCR, see “Methods” section), we tested whether changes in smell/taste at the population level could be used as an early indicator for local COVID-19 outbreaks. As pre-registered (see “Methods” section), our primary aim was to test the association between self-reported smell and taste changes and indicators of pressure in hospitals (COVID-related hospitalizations, CCRU admissions, and mortality rates) for each French administrative region over the last 3 months. Our secondary aim was to examine temporal relationships between the peak of smell and taste changes in the population and the peak of COVID-19 cases and the application of lockdown measures. The potential for self-reported smell and taste loss to serve as an early indicator of the number of COVID-19 cases—and hence hospital stress—was tested in a natural experiment by comparing France with Italy and the UK, which implemented lockdown with different timing and levels of stringency. Here, we show that self-reports of smell/taste changes are closely associated with hospital overload and are early markers of the spread of infection of SARS-CoV-2.

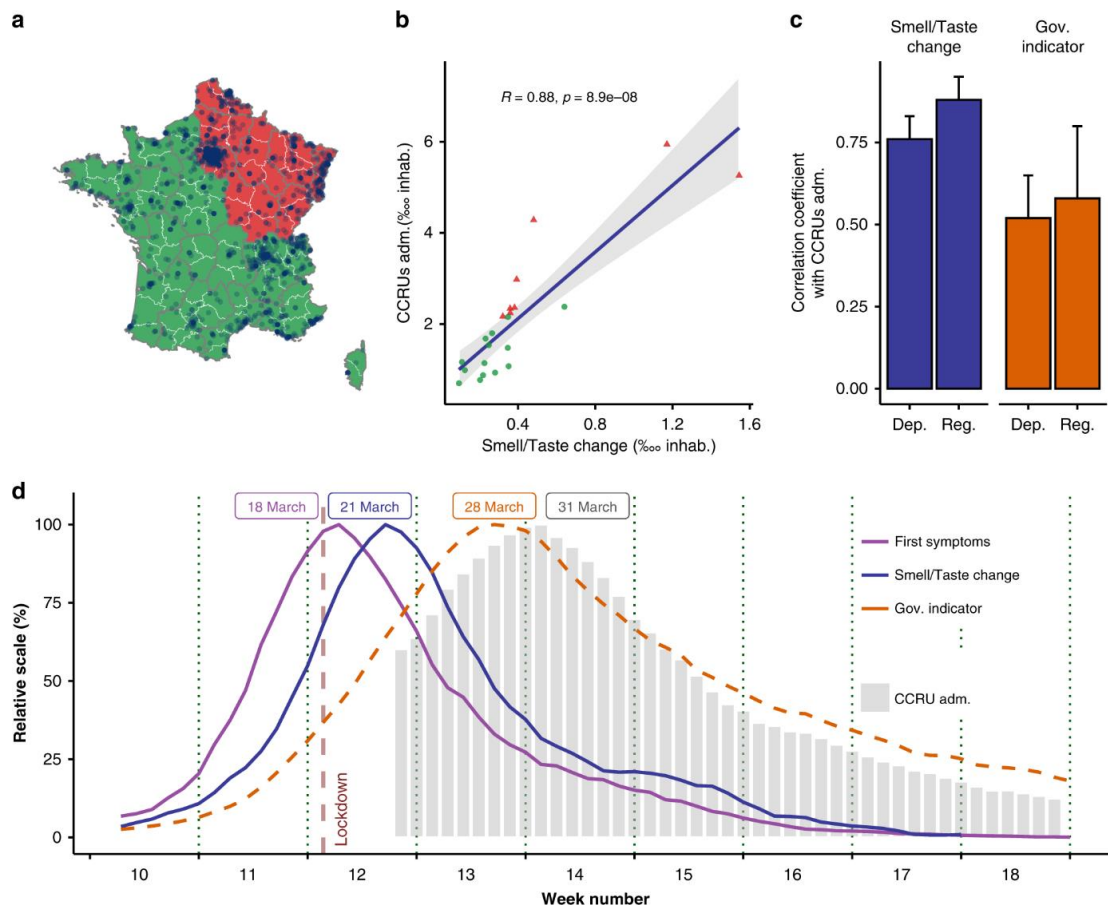


Figure 1: **a** French regions were assigned a green or red status by the French government to guide local relaxation of lockdown protocols. Dots represent people self-reporting smell and taste changes in a web-based survey. Base map is from OpenStreetMap and OpenStreetMap Foundation. **b** The number of COVID-19-related CCRU admissions (as of May 11, 2020) correlated with the number of self-reported chemosensory changes (between March 1 and May 11, 2020, total $n = 3832$). Green dots correspond to regions with a post-lockdown level labeled green, and red triangles indicate regions considered red. Values are standardized based on the number of inhabitants (inhab.) for each regions. The two red triangles with CCRU admissions >5 are Alsace and Ile de France. The gray band represent the confidence interval of the linear smooth (formula ‘ $y \sim x$ ’) R and p represent value of the test for association between paired samples, using one of Pearson’s product moment correlation coefficient, without correction for multiple comparisons. **c** Colored bar represent the value of computed correlation coefficients (confidence intervals are depicted as thin black bars) between the number of CCRU admissions per area and i) the number of people reporting smell and taste changes ($n = 3832$, blue), and ii) the governmental indicator (Gov. indicator), ratio of ER consults for COVID-19 (orange). Analyses were done both at the level of metropolitan regions (Reg) and departments (Dep). **d** Temporal relationships in France between smell/taste change symptom onset (blue solid line, $n = 1476$), the governmental indicator (orange dashed line), and COVID-19 cases in CCRUs (gray bars) around the lockdown period (vertical dashed line). Data are 7-day running averages, normalized to the day with the highest value.

Results

Changes in smell and taste are associated with overwhelmed healthcare systems

The relationship between self-reported changes in smell and taste by French residents (diagnosed as COVID-19+ or not, see “Methods” section and Supplementary Table 1) and estimators of local healthcare system stress was evaluated geographically. Figure 1a depicts the geographical distribution in red and green regions (as defined by the French government) and participants who self-reported changes in their smell and taste. Red areas of France account for 40.8% of the population. Green areas are clustered into a group with both a low number of self-reported chemosensory changes and a low number of admissions to CCRUs (Fig. 1b). Red areas show an opposite trend (Chi-square $< 1 \times 10^{-200}$ and Biserial correlations $p < 1.3 \times 10^{-2}$). A strong relationship exists between self-reported changes in smell and taste and the number of admissions to CCRUs ($R_{\text{smell}} = 0.88$, $p = 8.9 \times 10^{-08}$). This correlation remained significant even after removing the two most impacted areas (Alsace and Ile de France, $R_{\text{smell}} = 0.72$; $p < 3 \times 10^{-04}$), indicating that the significant relationship is not driven solely by these two regions.

Strikingly, use of self-reported chemosensory changes produced a stronger correlation than the current governmental indicator of virus circulation (Fig. 1c). Overall, smell/taste changes are better correlated with the number of COVID-19 admissions to hospitals than the current governmental indicator i.e., the ratio of ER consultations for suspicion of COVID-19 to general ER consultations ($R_{\text{smell}} = 0.81$, $p = 6.71 \times 10^{-06}$ vs. $R_{\text{gov}} = 0.60$, $p = 3.35 \times 10^{-03}$); the same pattern was found for the number of COVID-19 related deaths ($R_{\text{smell}} = 0.75$, $p = 5.62 \times 10^{-05}$ vs. $R_{\text{gov}} = 0.58$, $p = 4.97 \times 10^{-03}$ see Supplementary Table 2). Further, when smaller geographical areas were considered (France is divided into 96 administrative units, called departments), these correlations remained highly significant (e.g., admissions to CCRUs: $R_{\text{smell}} = 0.76$, $p < 5 \times 10^{-19}$) (Fig. 1c). Moreover, the three relationships (change in smell/taste versus COVID-19-related hospitalization, resuscitations, and death) also remained highly significant when considering only individuals who were not clinically diagnosed by a medical professional but considering themselves showing some symptoms of COVID-19 (e.g., admissions to CCRUs: $R_{\text{smell}} = 0.83$, $p = 1.65 \times 10^{-06}$). Potential sampling bias due to regional media

coverage of our survey (Supplementary Table 3) and self-reported chemosensory changes by region was ruled out by confirming these variables were not correlated ($R < 0.01$, $p > 0.9$).

Notably, relationships between pandemic markers and online searches related to chemosensation were also significant in France. Google queries related to smell or taste loss (“perte odorat,” “perte goût” in French) were correlated with the three measures of an overwhelmed healthcare system described above (e.g., CCRU admissions: $R_{\text{smell}} = 0.8$, $p < 4 \times 10^{-03}$, see Supplementary Table 2).

Changes in smell and taste are early markers of the effectiveness of political decisions

Next, we examined the temporal dynamics in France of self-reported changes in smell/taste, the current governmental indicator (ratio of ER consults), and the number of CCRU admissions due to COVID-19 before and after the lockdown period. As shown in Fig. 1d, the peak of the onset of changes in smell/taste appeared 4 days after the lockdown and for these individuals, the first reported COVID-19 symptoms occur even earlier. Conversely, the governmental indicator of ER consults only peaked 11 days after the lockdown, while the peak of CCRU admissions was shifted later by 14 days. This is consistent with emerging data showing that COVID-19-related changes in smell and taste occur in the first few days after infection [6, 12–14]. The robustness of smell and taste changes over time was assessed in two ways. First, we showed the peak of smell/taste changes remained the same regardless of our survey’s completion date (Supplementary Fig. 1a). Second, we observed the exact same peak when analyzing a separate French survey performed on 950 individuals and focusing on smell alterations in the French population independently of COVID-19 (see “Methods” section): the peak of olfactory changes again occurred 4 days after the lockdown decision, and this was independent of survey completion dates (Supplementary Fig. 1B). The robustness of smell and taste changes was also observed over age (Supplementary Fig. 2A) and gender (Supplementary Fig. 2B). Finally, we also show that the observed peak does not correspond to seasonal occurrence of allergies in France based on the ratio of consultations for Allergy to general consultations at the emergency room (Supplementary Fig. 3).

Further, analyses of Google searches confirm this temporal relationship: on the same days where survey participants report experiencing their first symptoms (around March 18, 2020), there was a peak of Google queries for terms associated with early COVID-19 symptoms (fever, cough, aches, Supplementary Fig. 4A). A few days later, the peak of online queries for “taste loss” and “smell loss” is synchronized with the report of smell and taste changes (Supplementary Fig. 4B). One week later, queries for shortness of breath preceded the peak of CCRU admissions (Supplementary Fig. 4C). Collectively, these results indicate a significant fraction of French COVID-19 patients followed the same symptom time course, experiencing initial symptoms at the very start of the lockdown, which might be representative of a peak of infection a few days before the lockdown. This is consistent with the ultimate goal of the lockdown, which was to decrease the number of new infections following implementation. Thus, the period immediately prior to lockdown represents the expected peak of new infections. In France, a large population may have been infected two days before lockdown because that weekend was crowded and sunny and occurred over the course of election day. Further, there were busier train stations and supermarkets in anticipation of a shortage of supplies during lockdown [12].

These data suggest that the short-term efficacy of a lockdown could be monitored by tracking changes in smell and taste in the population. To assess whether such a prediction might generalize to other countries, we performed parallel analyses with data from Italy and the UK, where the lockdown measures were established with different levels of severity (see Fig. 2). We monitored the dynamics of confirmed COVID-19 cases, self-reported first symptoms, and self-reported taste and smell changes, and compared them as a function of the governmental stringency index. Immediately after lockdown, we found that the two countries with the higher stringency index experienced a more rapid decrease in both self-reported smell and taste changes and COVID-19 symptoms. Further, as expected, the evolution of confirmed COVID-19 cases differs according to the stringency index. The governments of Italy and France rapidly increased their stringency index, which led to a sharp decrease in COVID-19 symptoms and cases. In contrast, in the UK, the number of people in the UK reporting symptoms showed a slower decrease, presumably due to a less severe lockdown policy, and the number of confirmed cases remained high during the observation window. In each country, self-reported smell and taste changes can be regarded as a useful metric to predict the dynamics of confirmed COVID-19 cases. That is, when the number of new onsets of chemosensory changes decreases sharply

(France and Italy), the number of confirmed COVID-19 cases also decreases, albeit with a lag of two weeks. On the contrary, a slow decrease in the number of new onset chemosensory changes is associated with a plateau of confirmed cases (UK).

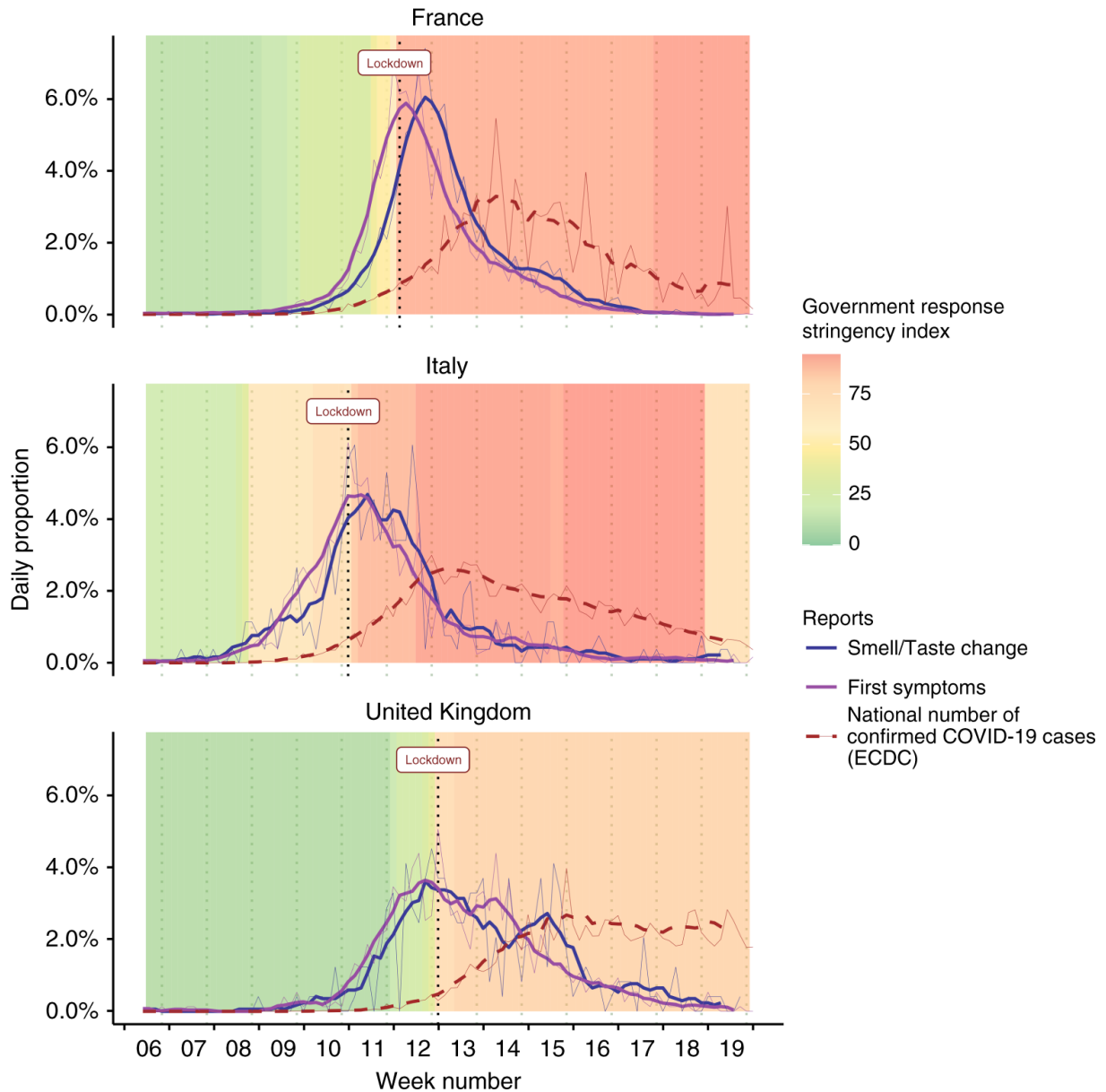


Figure 2: The daily proportion of first symptoms is shown as a violet line (France, $n = 4720$, Italy, $n = 1241$, UK, $n = 750$). The daily proportion of smell/taste changes is shown as a blue line (France, $n = 1487$, Italy, $n = 264$, UK, $n = 263$). The daily proportion of COVID-19 confirmed cases from the European Centre for Disease Prevention and Control (ECDC) is shown as a red dashed line. Each panel shows both raw data (thin line) and the corresponding 7-day running average (thick line). The government response stringency index is shown as the background color.

Discussion

The present analyses reveal a strong spatial and temporal relationships between self-reported smell and taste changes and multiple indices of health care system stress, such as admissions to CCRUs. This is consistent with cumulative evidence showing a high prevalence of chemosensory alterations in patients affected by COVID-19 in Europe (France [8, 14], Italy [9], UK[2, 15, 16]). Participants endorsed smell and taste changes only 3-4 days after their first symptoms. Such early chemosensory estimators may represent a cost-effective and easy way to implement alternative surveillance methods to large-scale virology tests, which are difficult to perform, costly, and time-consuming, especially during a pandemic.

A prominent question raised by these findings is whether the smell and taste changes observed in our study are solely related to COVID-19 or whether they can be explained by other temporal patterns, like seasonal illnesses or allergies. To the best of our knowledge, there are no existing studies that have explored the dynamics of sudden anosmia (as in COVID-19) throughout the year in France. Relationship between olfactory disturbances and seasons have been reported in Korea, Germany or US with a moderate increase of anosmia prevalence in spring [17–19]. Although the cyclical pattern of smell/taste changes might overlap, the amplitude of reported changes (either due to allergy or viral affection) were very limited compared to the present report. To further rule out the possibility, we examined whether the annual peak of allergies in France could explain the peak of smell and taste changes observed here. In analyzing existing French governmental data, we found that the annual peak of allergies in France occurred around week 30 (beginning of summer), multiple weeks after the observation window of the present study (from week 5 to week 20, Supplementary Fig. 3). Further, the French national aerobiological surveillance network (RNSA, <https://pollens.fr>), which follows pollen concentration in the atmosphere, has also indicated the first week of lockdown was very low risk for seasonal allergies. In addition, when considering Google Trends data, we did not observe any similar peaks in queries for smell/taste loss in the corresponding time period in previous years. Finally, a comparative study in Israel [20] showed that in COVID-19 suspected patient the frequency of smell change is almost ten time higher in a COVID-19 positive patients (68%) than in COVID-19 negative (8%). Considering that most of the participants of the present study are diagnosed with COVID-19 and that their description of a sudden loss of smell/taste is consistent with the now typical presentation of COVID-19 symptoms, it is highly probable that COVID-19 infection is the main reason of their smell and taste change. Collectively, these

data suggest the peak of smell and taste changes studied here are more consistent with sudden COVID-19 viral infections rather than an artifact due to seasonal illnesses.

The time lag between the onset of COVID-19-related symptoms and their declaration by the respondents of our study also deserves comment. Although immediate reporting of symptoms would have been ideal, such reporting is not possible within the context of the sudden first wave of a new viral pandemic. A similar time lag has been observed in other large-scale studies focusing on olfaction and COVID-19 [21]. Indeed, this time lag is inevitable given the preparation time required for scientists and clinicians design and launch such a survey, with appropriate ethics approval, once anosmia and ageusia began to emerge as cardinal symptoms of COVID-19. The vast majority of participants completed the survey between April 10th and April 19th, 2020, and most of them declared a date of onset of their symptoms roughly a month earlier (although a small fraction of participants did indicate onset prior to 2020). A possible consequence of a time lag between survey completion and the effective date of symptom onset is that subjects' statements may have been influenced by major societal events such as the lockdown decision, potentially creating some recall bias. To examine whether the date of a major event like the lockdown might bias dates of reported smell and taste loss, we explored narrative descriptions provided by our participants. By analyzing responses to the optional open-ended question "Please describe the progression or order you noticed your symptoms", we observed that, for France, a mere 11 of 3705 people (who have filled the optional question) used the term "confinement" ("lockdown") in their description of the onset date. Separately, another factor that mitigates concerns about a potential recall bias is the stable nature of participant's statements, regardless of their date of completion. That is, logic suggests, the longer the time between the onset date of smell and taste loss and the reporting date, the greater the recall bias should be. However, our data clearly show that regardless of the date of completion, the onset date falls within the same period (Supplementary Fig. 1). Finally, other evidence against a potential recall bias comes from Google Trends data. Analyzing real-time Google queries in March, we observed a very particular trend in France (Supplementary Fig. 4). We first observed a peak of queries for terms associated with early COVID-19 symptoms (fever, cough, aches) synchronized with the declared onset of the first symptoms in the survey (around March 18th). A few days later, a peak of online queries for "taste loss" and "smell loss" was seen, and this was synchronized with the date reported of smell and taste changes in our survey. The striking concurrence between Google queries and reports in our survey argues against the idea that a recall bias could be driving the effects described here.

Another important factor to consider in our survey is the way the press and media might have influenced our findings. Indeed, when the survey was launched, smell and taste changes were reported as symptoms of COVID-19 in the national and local media, which might have influenced respondents to remind themselves of such symptoms and to then report these changes on the survey. Such an emphasis on smell and taste loss would have biased attempts to explore the prevalence of chemosensory deficits in COVID-19. However, the primary aim of the present investigation was not to focus on the prevalence of anosmia and ageusia with COVID-19, but rather to explore use of reported smell and taste loss as indicators of COVID-19 pandemic. Still, the media coverage of our survey could also have biased the selection of participants geographically, as some French regions received more media coverage than others. However, as reported above, there was no correlation between the number of participants in a given region and the intensity of media and press coverage for the survey in that same region. Finally, when participants were asked to describe the chronology of their symptoms, they did not refer to the media coverage as a prominent element influencing their awareness of their smell/taste changes. While this does not exclude an implicit and non-verbalized bias due to media coverage, this pattern suggests a genuine report of symptoms with a high occurrence of COVID symptoms just after the lockdown.

An interesting question raised by our findings is what impact they might have on government strategies in a pandemic. Following lockdown, the rapid decrease of self-reported changes in smell and taste in France may be representative of the effectiveness of this decision in reducing infection rates. Similarly, data from Italian participants show highly similar patterns, but with a one-week difference compared to the French data. This might reflect highly similar responses by the Italian and French governments. Conversely, the prevalence of chemosensory changes in the UK shows a more gradual decrease. The UK government began with advice to avoid pubs, clubs and theaters, and to work from home from March 16, with restrictions around March 18. However, a lockdown was not declared until March 23, and this was less stringent than those in France or Italy. Notably, new COVID-19 cases in the UK showed a plateau phase which is not observed in either France or Italy. Accordingly, we conclude that collecting online information about changes in smell and taste from residents (even retrospectively) may be a valuable metric of the effectiveness of reopening strategies related to the COVID-19 pandemic. Practically, in areas where smell and taste changes are notable COVID-19 symptoms, the proportion of individuals who self-report changes in their ability to smell or taste might be an early indicator of subsequent demand for healthcare. If confirmed, continuous monitoring of

changes in smell and taste perception would then be a highly cost-effective, minimally invasive, and reliable way to track future COVID-19 outbreaks. When used this way, we caution that particular attention must be paid to potential selection bias. That is, self-report studies online can be impacted by multiple selection biases, including socioeconomic status, fluency with technology and willingness and interest in participating in scientific research. When considering the present data, at least 3 parameters may contribute to a selection bias in our sample: (1) the age, (2) the gender of the participants, and (3) the format and the advertising of the survey. Regarding participant' age, our study cohort (mean 40.7 years, $sd = 12.4$) was quite similar to the French population mean (41.1 years, according to INSEE, <https://www.insee.fr/fr/statistiques/1893198>); however, we did only include individuals over 18 due to issues of consent, and administrative reasons, and seniors were also less represented. For gender, our sample contained a greater proportion of women (67%) compared to men, which might influence the results. However, additional analysis showed no differences in peaks of smell/taste changes across age or gender, minimizing concerns that such selection biases may have influenced present results (See Supplementary Fig. 2). We also tested the potential selection bias due to format and the advertising of the survey, by comparing the GCCR dataset with an independent second study performed on French residents (see “Methods” section). Remarkably we observed highly similar results across studies where advertising, inclusion criteria, and survey format were different.

Based on the present findings, we highlight the paramount importance and robustness of associations between smell/taste changes and COVID-19 and we strongly endorse the need for additional large-scale validation studies to assess the causality between the observed association between smell/taste changes and indicators of the COVID-19 pandemic. This could be achieved by setting up a simplified interface where selection biases are controlled for (age, gender, motivation, media coverage, socioeconomic level, etc.) through both traditional and online media—and whereby real time information about changes in smell and taste in the general population may be available to decision-makers. Subjects' participation in the questionnaire and the reliability of the answers should also be considered. In particular, if a participant knows how their answers may influence enforcement of lockdown, their answers might become less truthful. This motivation can be expressed through different forms of behavior. Whereas some individuals may tend to provide statements that minimize their symptoms in order to avoid strict containment measures, others will maximize their declaration to maintain the lockdown, or will provide honest answers in order to participate in the collective effort to better understand the COVID-19 pandemic. These motivational factors are a recurrent risk in online studies and

different strategies should be held to control for them in future predictive studies. Based on the above, a large implementation of the study of smell and taste changes in institutional models should allow for monitoring of COVID-19 spread. This might be especially relevant in areas in which testing proves difficult or delayed and for future outbreaks that may overlap with other seasonal viral diseases which share many of the symptoms (fever, cough etc.) but whose treatment or prevention (vaccination) are less demanding in terms of critical care than COVID-19. We advocate that self-report surveys should be used to enhance other strategies such as large-scale PCR tests and COVID-19 symptom assessments (including anosmia and ageusia) in primary/secondary care.

In summary, we propose that an increase in the incidence of sudden smell and taste change in the general population may be used as a valuable minimally invasive indicator of coronavirus spread in the population. To formally test the temporal relationship between chemosensory changes and spread of the disease, we recommend that a large-scale causal study in different countries be conducted on real-time monitoring of self-reported changes in the ability to smell or taste. Such a prospective study will allow for the creation of statistical models that can assist in prediction of future hospital admissions for COVID-19. Further, it could also help decision-makers take important measures at the local level, either in catching new outbreaks sooner, or in guiding the relaxation of local lockdowns, given the strong impact of lockdown on economic and social activities.

Methods

Online survey

This study is mainly based on data from the Global Consortium for Chemosensory Research survey (GCCR, <https://gcchemosensr.org/>) – a global, crowd-sourced online study deployed in 30+ languages [22]. The data analyzed here were collected from April 7 to May 14, 2020. The protocol complies with the revised Declaration of Helsinki and was approved as an exempt study by the Office for Research Protections at The Pennsylvania State University (Penn State) in the U.S.A. (STUDY00014904; PI Hayes).

Participants in the GCCR questionnaire were recruited by word of mouth, as well as through social and traditional media (flyers, social media, television, radio) during the COVID-19

pandemic. It was well covered by the French press, as over 70 articles mentioned the project, at both the regional and national level (see Supplementary Table 3). Respondents received no monetary incentive for their participation. Inclusion criteria were as follows. (i) Questionnaire completion was allowed only to participants who indicated they had suffered from a respiratory disease in the past two weeks, whether they noticed a change in their taste/smell or not. (ii) Participants aged 18 years old or younger were excluded.

For the analyses conducted in this article, only individuals reporting a change in smell and/or taste perception were included, based on the question “Have you had any of the following symptoms with your recent respiratory illness or diagnosis?”. Moreover, to exclude unreliable entries, participants must have reported a quantitative decrease of at least 5 on a 0-to-100 rating scale between their ability to smell and/or taste before and during their recent respiratory illness or diagnosis. Therefore, Due to this inclusion criteria, “smell/taste change” is equivalent to a quantitative decrease of participant ability to smell and/or taste. We then extracted individuals from the full dataset who reported living in France, Italy or the UK. As the country of residence was completed as a text entry, we allowed for typical variations (e.g., “United Kingdom” or “UK”), spelling mistakes, use of different languages (e.g., “Italie” or “Italia”), as well as subdivisions (e.g., “Scotland”) and major cities (“Paris”). Metropolitan France was split into 13 so-called “regions” in 2016. However, we considered the former system where France was split into 22 regions here, since the organization of the health system mostly remains based on the structure built before 2016. An alternative, finer granularity, splits metropolitan France into 96 so-called “departments.” To retrieve the French department and region of the participants, we used the city of residence they reported in the questionnaire and combined them with the French public website (data.gouv.fr, after a semi-manual correction of spelling). Participants came from all metropolitan departments but three (Mayenne, Creuse, Cantal). Consequently, the number of responses analyzed in France was between $n = 1476$ and 4720 depending on the analysis conducted (i.e., on whether the information of interest was present or missing and the date range of analysis, see Supplementary Table 1 for details). For comparison, between 264 to 1241 participants from Italy and between 243 to 750 participants from the UK were included. Most participants were women (FR:66.38%, IT:69.3%, UK:76.0%), and the mean age was around 40 [FR = 40.7 (sd = 12.4), IT = 41.1 (sd = 11.4), UK = 41.09 (sd = 12.1)]. In the French data, a total of 15% of individuals tested positive for COVID-19 (lab result) and 44% were diagnosed clinically by a medical professional from their symptoms. The remaining 41% were not diagnosed for COVID-19 but declared a change in perception of either smell or taste. The number of participants was normalized by region, by using the number of inhabitants in each

region as estimated by the French public statistics office, INSEE. Finally, the time of onset of smell and taste change was assessed via responses to several optional open-ended questions. These included: “Please describe the progression or order you noticed your symptoms” and the time of onset of recent disease by the question: “At what date did you first notice symptoms of your recent respiratory illness? Provide your best guess or leave blank if you do not remember.”

Complementary and independent French Survey

The data of another online survey were used to evaluate the robustness of the temporal evolution of smell and taste changes. This survey was conducted in the French population between April 8 and May 8, 2020 and aimed at characterizing chemosensory disorders in people with and without COVID-19, as well as their consequences on quality of life. The data of 950 respondents were eligible for comparison with data from the GCCR survey, i.e., responses where both the date of completion and the date of smell loss onset were provided. Only responses that were complete and from people who were responding to the questionnaire for the first time and were over age 18 were included. This survey was approved by the CNRS ethics committee. Data collection was strictly anonymous. The protocol complies with the revised Declaration of Helsinki and the study was approved by the ethics committee of the Institute of Biological Sciences of the CNRS on the 3rd of April 2020 (DPO #TRRECH-467). All individuals provided informed consent when participating in the survey.

Online trends

Trends of online queries by French region were performed using Google Trends, a tool returning the popularity of a search term in a specific state or region. Google is by far the most used search engine in France (>90% of internet searches, according to StatCounter Global Stats). We looked for the popularity of terms (listed in Supplementary Fig. 3, using default selection of “All categories” and “Web search”), within the timeframe of February 1, 2020 to May 10, 2020 (from the month of the first official COVID-related death in Europe to the end of lockdown in France). It should be noted that Google Trends does not provide the actual numbers of searches but rather a relative score from 0 to 100 (100 corresponding to the day with the greatest number of searches during the specified time period). To compare Google

Trends scores between French regions, we transformed them by computing the relative number of queries per day in the region of interest. For example, despite a value of 100, the peak day might represent only 5% of the total number of queries related to the topic across the timeframe of interest (see above).

Healthcare system data

The French governmental indicator to estimate the circulation of the virus was calculated from the ratio of consultations for suspected COVID-19 to general consultations at the emergency room (ER) in hospitals. This ratio corresponds to the medical diagnostic for COVID-19 suspicion (codes CIM10: U07.1, U07.10, U07.11, U07.12, U07.14, U07.15, U04.9, B34.2, B97.2). The definition of COVID-19 has evolved rapidly during the lockdown period but the diagnosis is principally based on symptoms of COVID-19 considered as common such as fever, cough, and dyspnea (difficulty breathing). To the best of our knowledge, anosmia and ageusia were officially considered in France as putative symptoms of COVID-19 from a letter of the Direction Générale de la Santé (April 1st) and communication of the Haut Conseil de la Santé Publique (a letter dated April 8, published online April 15, following a letter from the CNP-ORL dated March 20). Areas with values of the French governmental indicator higher than 10% are considered having a high virus circulation. This indicator contributes to the assignment of a red/green label. Allergies incidence in previous years were calculated from the ratio of consultations for Allergy to general consultations at the emergency room (ER) in hospitals.

Data dealing with the health status across countries (number of COVID-19 cases and deaths for each day) were downloaded on May 22, 2020 from the European Centre for Disease Prevention and Control databank (ECDC, <https://www.ecdc.europa.eu/en>). Data regarding healthcare system stress in France (hospitalizations, CCRU entries and deaths) were also downloaded on May 22 from the French Public Health website (Géodes, Santé Publique France, <https://geodes.santepubliquefrance.fr/#c=home>). Here, we use the term CCRU (Critical Care Resuscitation Unit) to translate the French hospital service of “Réanimation.” Raw data were normalized across regions with regard to their number of inhabitants as estimated by INSEE. The temporal evolution of the stringency of government response was retrieved from the Oxford COVID-19 (<https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>). Here, the stringency level of a country is computed according to which measures of a list of items (e.g., school closures, cancellation of public events,

international travel controls, etc.) are undertaken. For the post-lockdown situation, the color assigned by the French government to each department was downloaded on May 12 from the government website. Only data before May 11 (the initial lift of the lockdown) were included in the analyses.

Statistical analyses

Statistical analyses were pre-registered at the Open Science Framework (OSF). Data were analyzed using R software (4.0) and its standard packages (maps, ggplot, etc.). Data were grouped at the national level (France, Italy, UK). In France they were also grouped at the regional level (according to the division into 22 regions in place prior to the 2016 reform). The rationale behind this is that the healthcare system is still structured following this organization, with University Hospitals in regional main cities serving patients of the surrounding departments. Participants from overseas French territories were not included in the geographical analysis because of too few data ($n < 10$). The relationship between (1) GCCR responses (or online queries), and (2) public health data was determined using parametric (e.g., Pearson correlations) statistics as allowed by the normal distribution of the variable of interest. The association between GCCR participant and *red/green* post-lockdown status was tested using Chi-square tests and Biserial correlations. Complementary analyses not planned in the pre-registration included: (i) the analysis using the independent French online survey (see section “Complementary and independent French Survey” of the methods), (ii) the correlation between regional media coverage and the number of responses to the online survey per region, (iii) the correlation at the level of department, (iv) the correlation excluding extreme points, and (v) the correlation with the government indicator. Pre-registered statistical analyses not presented here include: (i) Mann-Kendall trend test and Change-point Detection test to detect time series changes, and (ii) part of the Google Trends analysis.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this article.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its [supplementary information](#) files. (Source Data file). [Source data](#) are provided with this paper.

Code availability

R scripts are available on the osf server (<https://osf.io/gew7p/>).

References

1. Gilbert M, Dewatripont M, Muraille E, et al (2020) Preparing for a responsible lockdown exit strategy. *Nat Med* 26:. <https://doi.org/10.1038/s41591-020-0871-y>
2. Menni C (2020) Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 26:. <https://doi.org/10.1038/s41591-020-0916-2>
3. Eliezer M (2020) Sudden and complete olfactory loss function as a possible symptom of COVID-19. *JAMA Otolaryngol Neck Surg* 146:. <https://doi.org/10.1001/jamaoto.2020.0832>
4. Gautier J-. F, Ravussin Y (2020) A new symptom of COVID-19: loss of taste and smell. *Obesity* 28:. <https://doi.org/10.1002/oby.22809>
5. Pellegrino R (2020) Corona viruses and the chemical senses: past, present, and future. *Chem Senses* 45:. <https://doi.org/10.1093/chemse/bjaa031>
6. Bagheri SH (2020) Coincidence of COVID-19 epidemic and olfactory dysfunction outbreak in Iran. *Med J Islam Repub Iran MJIRI* 34:
7. Beltrán-Corbellini Á (2020) Acute-onset smell and taste disorders in the context of Covid-19: a pilot multicenter PCR-based case-control study. *Eur J Neurol* 27:. <https://doi.org/10.1111/ene.14273>
8. Bénézit F (2020) Utility of hyposmia and hypogeusia for the diagnosis of COVID-19. *Lancet Infect Dis* 20:. [https://doi.org/10.1016/S1473-3099\(20\)30297-8](https://doi.org/10.1016/S1473-3099(20)30297-8)
9. Giacomelli A (2020) Self-reported olfactory and taste disorders in SARS-CoV-2 patients: a cross-sectional study. *Clin Infect Dis* 71:. <https://doi.org/10.1093/cid/ciaa330>

10. Haehner A, Drafi J, Dräger S, et al (2020) Predictive value of sudden olfactory loss in the diagnosis of COVID-19. *ORL* 82:.. <https://doi.org/10.1159/000509143>
11. Yan CH, Faraji F, Prajapati DP, et al (2020) Self-reported olfactory loss associates with outpatient clinical course in Covid-19. *Int Forum Allergy Rhinol* 10:.. <https://doi.org/10.1002/alr.22592>
12. Salje H (2020) Estimating the burden of SARS-CoV-2 in France. *Science* 369:.. <https://doi.org/10.1126/science.abc3517>
13. Levinson, R. et al. Anosmia and dysgeusia in patients with mild SARS-CoV-2 infection. Preprint at, <https://doi.org/10.1101/2020.04.11.20055483> (2020).
14. Klopfenstein T (2020) Features of anosmia in COVID-19. *Med Mal Infect* 50:.. <https://doi.org/10.1016/j.medmal.2020.04.006>
15. Lechien JR (2020) Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. *Eur Arch Otorhinolaryngol* 277:.. <https://doi.org/10.1007/s00405-020-05965-1>
16. Moein ST (2020) Smell dysfunction: a biomarker for COVID-19. *Int Forum Allergy Rhinol* 10:.. <https://doi.org/10.1002/alr.22587>
17. Lotsch J, Hummel T (2020) A data science-based analysis of seasonal patterns in outpatient presentations due to olfactory dysfunction. *Rhinology* 58:
18. Potter MR, Chen JH, Lobban N-. S, Doty RL (2020) Olfactory dysfunction from acute upper respiratory infections: relationship to season of onset. *Int Forum Allergy Rhinol* 10:.. <https://doi.org/10.1002/alr.22551>
19. Kang JW, Lee YC, Han K, et al (2020) Epidemiology of anosmia in South Korea: a nationwide population-based study. *Sci Rep* 10:.. <https://doi.org/10.1038/s41598-020-60678-z>
20. Karni, N. et al. Self-rated smell ability enables highly specific predictors of COVID-19 status: a case control study in Israel. Preprint at <https://doi.org/10.1101/2020.07.30.20164327> (2020).
21. Iravani B (2020) Relationship between odor intensity estimates and COVID-19 prevalence prediction in a Swedish Population. *Chem Senses* 45:.. <https://doi.org/10.1093/chemse/bjaa034>
22. Parma, V. et al. More than smell—COVID-19 is associated with severe impairment of smell, taste, and chemesthesis. *Chem. Senses* bjaa041. <https://doi.org/10.1093/chemse/bjaa041> (2020).

Acknowledgements

This work was supported financially by EXTREM-O (CNRS MITI), CONFINEZ2 (CNES), the CORODORAT grant (IDEX-Lyon-Université de Lyon), by the French government, through the UCAJEDI “Investments in the Future” project managed by the ANR grant No. ANR-15-IDEX-01. The authors thank Sébastien Buthion, Conceição Silva, Brigitte Perucca, Clément Blondel, Stéphanie Younès, François Maginot, and their teams from the CNRS communication units, as well as the Global Consortium for Chemosensory Research (GCCR). Deployment of the GCCR survey in multiple languages by the Pennsylvania State University (Penn State) was partially supported by funds from James and Helen Zallie and the Penn State Sensory Evaluation Center.

Author information

These authors contributed equally: Denis Pierron, Veronica Pereda-Loth, Camille Ferdenzi, Jérôme Golebiowski, Moustafa Bensafi.

These authors jointly supervised this work: Denis Pierron, Veronica Pereda-Loth, Camille Ferdenzi, Jérôme Golebiowski, Moustafa Bensafi.

Authors and Affiliations

Équipe de Médecine Evolutive Faculté de chirurgie dentaire; UMR5288; CNRS/Université Paul-Sabiater Toulouse III, Toulouse, 31400, France

Denis Pierron, Veronica Pereda-Loth, Omar Alva, Julie Kabous, Margit Heiske & Thierry Letellier

Lyon Neuroscience Research Center, CNRS UMR5292, INSERM U1028, University Claude Bernard Lyon 1, Bron, France

Marylou Mantel, Maëlle Moranges, Camille Ferdenzi & Moustafa Bensafi

Université Côte d’Azur, CNRS, Institut de Chimie de Nice UMR7272, Nice, France

Emmanuelle Bignon, Jody Pacalon & Jérôme Golebiowski

Université Côte d'Azur, CHU de Nice, Nice Memory Clinic, Nice, France

Renaud David

University of Florence, Florence, Italy

Caterina Dinnella, Sara Spinelli & Erminio Monteleone

Interdepartmental Neuroscience Program, Yale University, 333 Cedar Street, New Haven, CT, 06520, USA

Michael C. Farruggia

Department of Neurobiology and Behavior, University of California, Irvine, CA, 92697, USA

Keiland W. Cooper

Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Blvd, Philadelphia, PA, 19104, USA

Elizabeth A. Sell

Centre des Sciences du Goût et de l'Alimentation, INRAE, CNRS, AgroSup-Dijon, University Bourgogne Franche-Comté, Dijon, France

Thierry Thomas-Danguin

The Pennsylvania State University, Philadelphia, PA, 19104, USA

Alyssa J. Bakke & John E. Hayes

Temple University, Philadelphia, PA, 19122, USA

Valentina Parma

Department of Brain and Cognitive Sciences, Daegu Gyeongbuk Institute of Science and Technology, Daegu, 711-873, South Korea

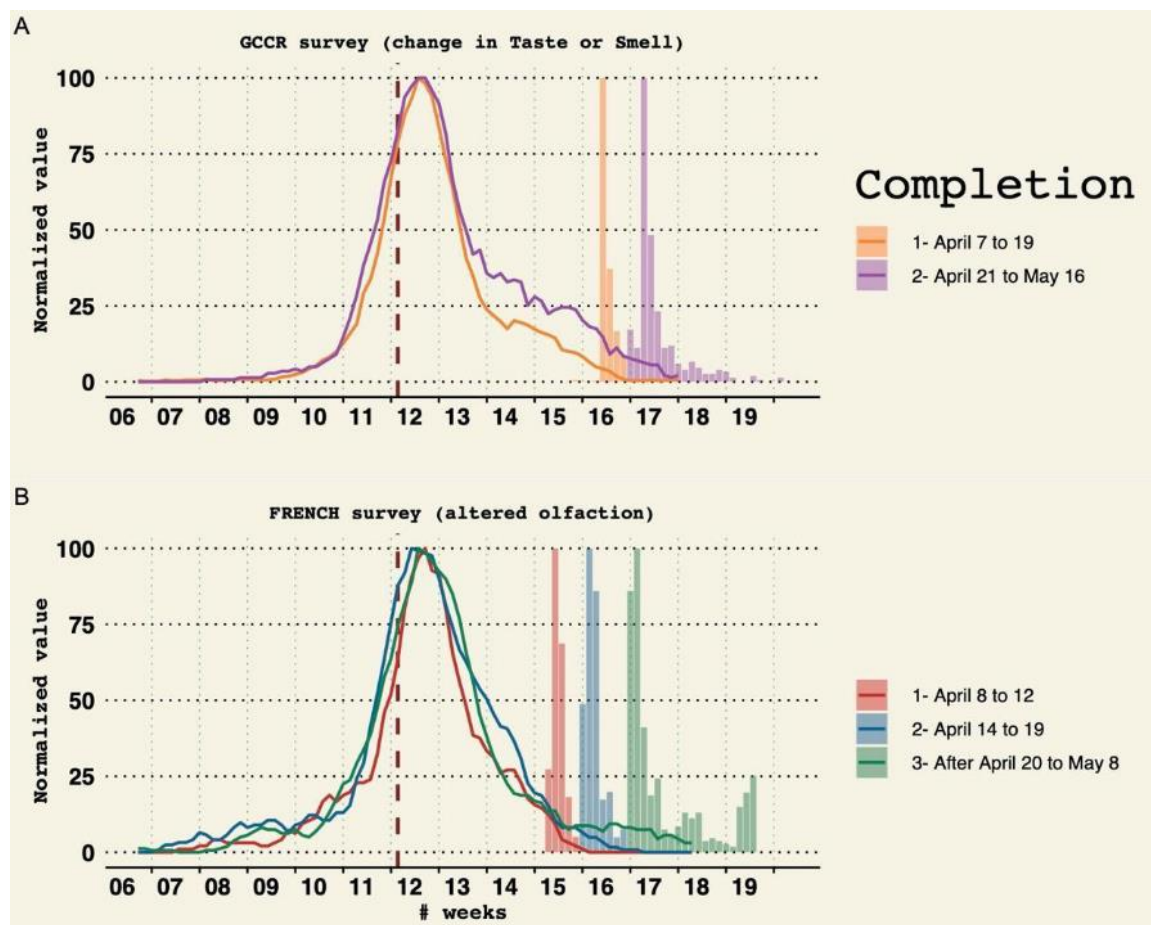
Jérôme Golebiowski

Contributions

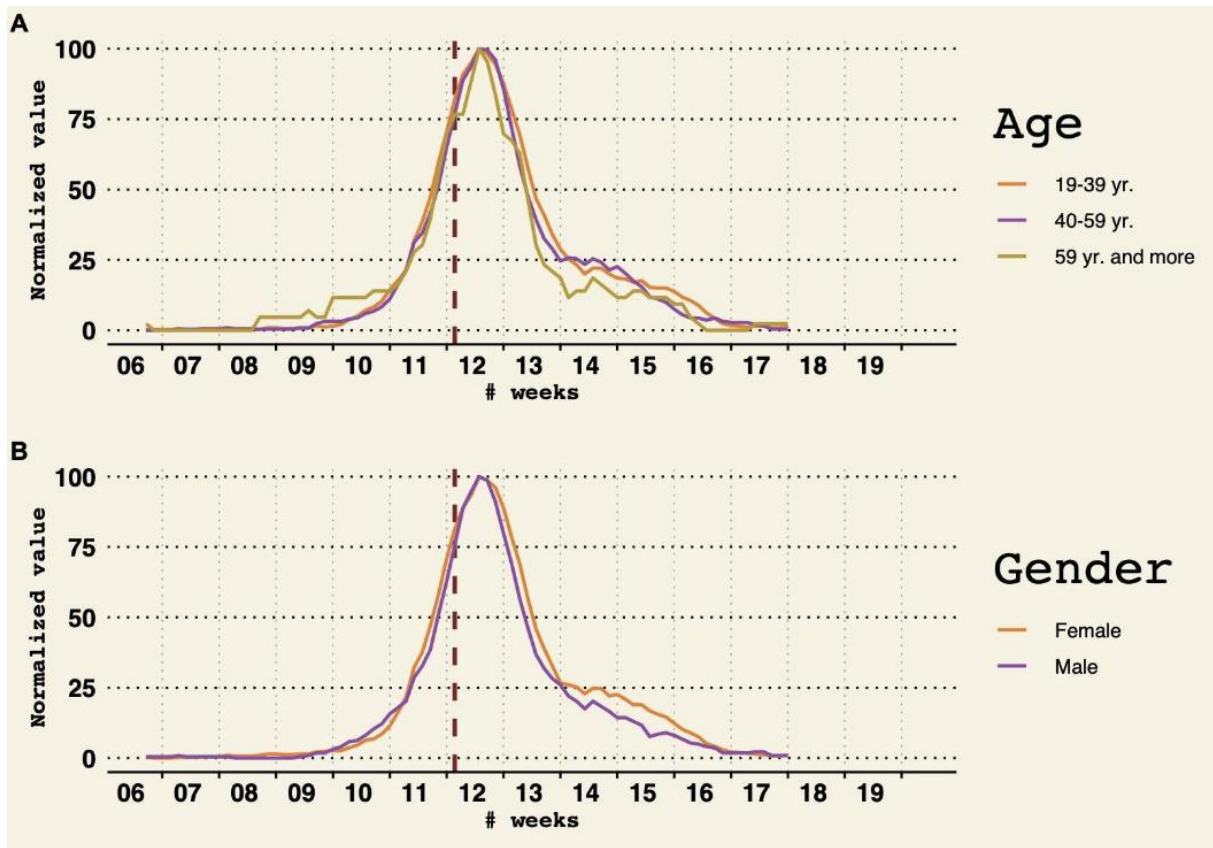
Conceived, designed the study, and wrote the paper: D.P., V.P.-L., C.F., J.G., M.B. Data acquisition and curation: D.P., V.P., M.M., M.M., E.B., O.A., J.K., M.H., J.P., R.D., C.D., S.S., E.M., M.C.F., K.W.C., E.A.S., T.T.D., A.J.B., V.P., J.E.H., T.L., C.F., J.G., M.B. Performed analysis: D.P., V.P.-L., C.F., J.G., M.B., M.M.a., M.M.o., E.B. Edited and approved the final

Supplementary information

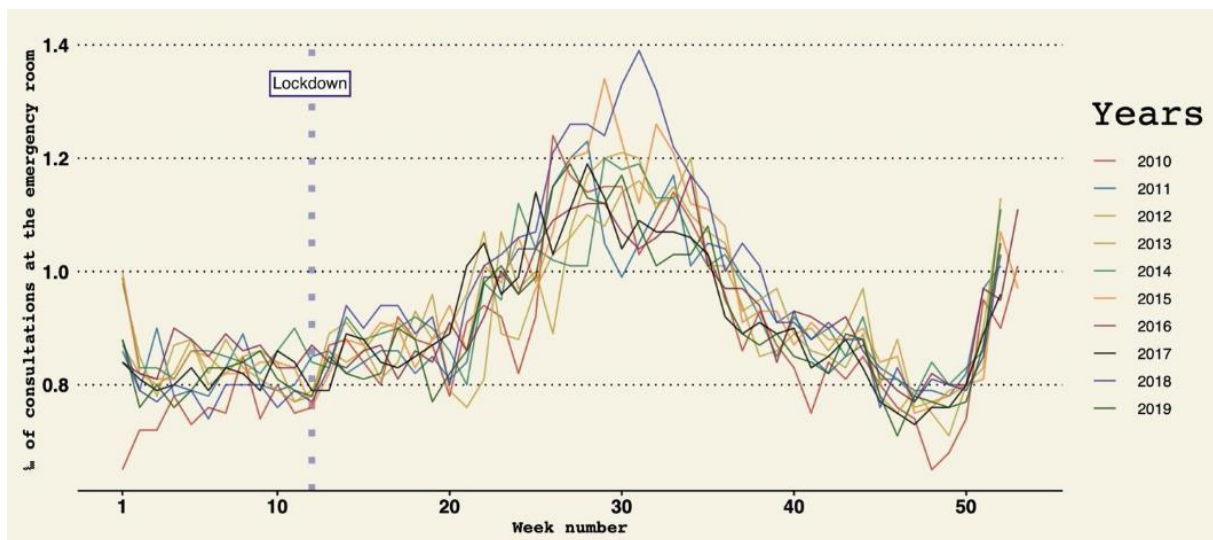
Supplementary Figures



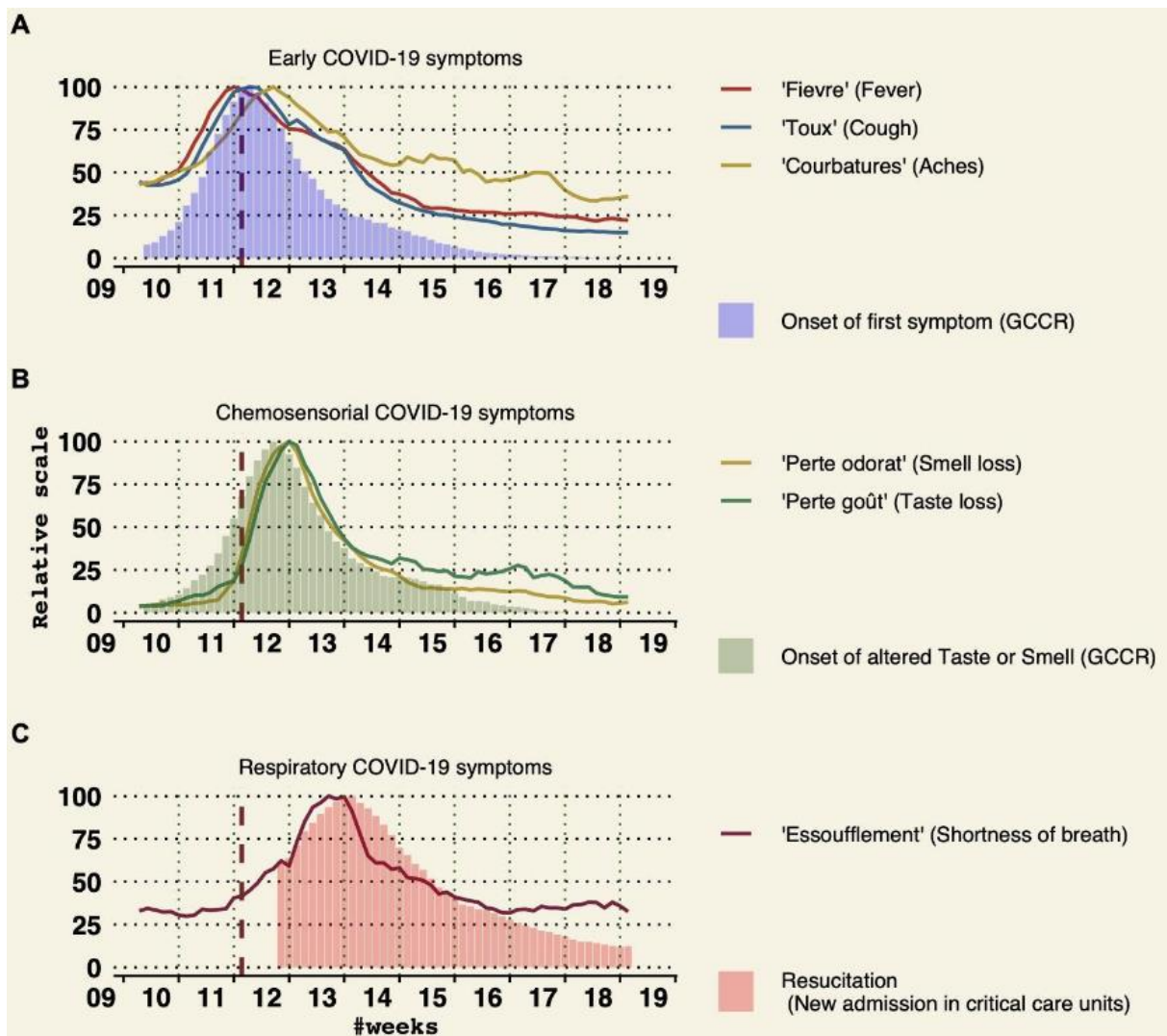
Supplementary Figure 1: Evaluation of the robustness of reported smell and taste changes over time. Colored lines represent the proportion of reported onset of smell and taste change normalized to the highest value of each series. The barplot indicates the proportion of the date of completion of the survey (normalized to the highest value of each series). Colors represent groups of participants according to completion time. French lockdown is represented by the vertical brown dashed line. A) Peaks of smell/taste changes in the French participants who answered the GCCR questionnaire according to different dates of completion (before or after April 20). B) Date distribution of an independent French survey performed on 950 individuals and focusing on smell alterations in the French population independently of COVID-19 (see Methods).



Supplementary Figure 2: Time of peaks of smell/taste changes in French participants who answered the GCCR questionnaire according to their age or gender. Similarly to figure S1, colored lines represent the proportion of reported onset of smell and taste normalized to the highest value of each series. Colors represent groups of participants according to age (S2.A) or gender (S2.B).



Supplementary Figure 3: Evolution of the ratio of consultation for allergies in the emergency room (for 10.000 consultation) over time during the last 10 years. The start of the 2020 lockdown is represented by the vertical blue line.



Supplementary Figure 4: Google trends in France for COVID-19 symptoms. The French lockdown start is shown by a brown dashed line. A) The peak of the onset of first symptoms declared in the GCCR survey (blue bars, 1 day after the lockdown) corresponds to peaks of online queries for fever (red line), cough (blue line) and aches (yellow line). B) The peak of onset of altered smell/taste (green bars, 4 days after the lockdown) corresponds to peaks of online queries for “smell loss” (yellow) and “taste loss” (green). C) Online queries for “shortness of breath” (red curve) preceded the peak of the number of admissions to CCRU by three days.

Supplementary Tables

Supplementary Table 1: Data exclusions and Sampling strategy.

	n
Total GCCR participants	41759
Answering Smell change question	18118
Change in SMELL or TASTE	15872
Residents in France - ITALY - UK	7929
With date onset of first symptom	6784
FIGURE 1 :	
French resident with onset of first symptom between 03-01-2020 and 05-11-2020	3832
France resident with information regarding the onset of Smell/taste change	1476
FIGURE 2 :	
FRANCE With date onset of first symptom between 02-01-2020 and 05-11_2020	4720
ITALY With date onset of first symptom between 02-01-2020 and 05-11_2020	1241
UK With date onset of first symptom between 02-01-2020 and 05-11_2020	750
FRANCE with date onset of SMELL/TASTE change between 02-01-2020 and 05-11_2020	1487
ITALY with date onset of SMELL/TASTE change between 02-01-2020 and 05-11_2020	264
UK with date onset of SMELL/TASTE change between 02-01-2020 and 05-11_2020	243

Supplementary Table 2: Test of associations between putative indicators. Test for association between paired samples, using one of Pearson's product moment correlation coefficient, without correction for multiple comparisons (IDF= Ile de France, capital region of France, "GCCR participants" =participant to the present study).

variableX	variableY	p	cor	type	level
proportion of GCCR participants without medical diagnostic	ratioER	7.32e-01		0,04	without IDF an departement
proportion of GCCR participants without medical diagnostic	ratioER	7.54e-01		0,07	without IDF an region
proportion of GCCR participants	ratioER	1.21e-01		0,17	without IDF an departement
proportion of GCCR participants	ratioER	4.13e-01		0,19	without IDF an region
proportion of GCCR participants without medical diagnostic	hospital death	1.91e-02		0,25	without IDF an departement
proportion of GCCR participants without medical diagnostic	hospitalization	1.3e-02		0,27	without IDF an departement
ratioER	CCRU admision	2.96e-03		0,32	without IDF an departement
ratioER	hospitalization	1.16e-03		0,34	without IDF an departement
ratioER	hospital death	1.05e-03		0,35	without IDF an departement
courbatures_GOOGLE	ratioER	3.2e-01		0,38	without IDF an region
proportion of GCCR participants without medical diagnostic	ratioER	8.01e-05		0,39	all departement
proportion of GCCR participants	hospital death	6.42e-05		0,42	without IDF an departement
proportion of GCCR participants	ratioER	1.6e-05		0,42	all departement
proportion of GCCR participants without medical diagnostic	hospital death	5.58e-02		0,43	without IDF an region
ratioER	CCRU admision	6.06e-02		0,43	without IDF an region
courbatures_GOOGLE	ratioER	2.12e-01		0,43	all region
ratioER	hospital death	5.02e-02		0,44	without IDF an region
proportion of GCCR participants without medical diagnostic	hospitalization	4.73e-02		0,45	without IDF an region
proportion of GCCR participants without medical diagnostic	ratioER	3.43e-02		0,45	all region
proportion of GCCR participants without medical diagnostic	hospital death	3.19e-06		0,46	all departement
ratioER	hospitalization	4.09e-02		0,46	without IDF an region
courbatures_GOOGLE	CCRU admision	2.16e-01		0,46	without IDF an region
proportion of GCCR participants	hospitalization	5.3e-06		0,47	without IDF an departement
proportion of GCCR participants without medical diagnostic	CCRU admision	4.55e-06		0,47	without IDF an departement
courbatures_GOOGLE	hospitalization	1.97e-01		0,47	without IDF an region
proportion of GCCR participants	ratioER	2.23e-02		0,48	all region
courbatures_GOOGLE	hospital death	1.86e-01		0,48	without IDF an region
proportion of GCCR participants	hospital death	1.86e-07		0,5	all departement
proportion of GCCR participants without medical diagnostic	hospitalization	2.49e-07		0,5	all departement
ratioER	hospital death	1.86e-07		0,5	all departement
ratioER	hospitalization	1.14e-07		0,51	all departement
ratioER	CCRU admision	6.9e-08		0,52	all departement
proportion of GCCR participants	hospitalization	6.08e-09		0,55	all departement
proportion of GCCR participants without medical diagnostic	CCRU admision	1.21e-02		0,55	without IDF an region
ratioER	CCRU admision	4.82e-03		0,58	all region
ratioER	hospital death	4.97e-03		0,58	all region
perte gout_GOOGLE	ratioER	8.1e-02		0,58	all region
perte gout_GOOGLE	ratioER	1.01e-01		0,58	without IDF an region
perte odorat_GOOGLE	hospital death	9.79e-02		0,59	without IDF an region
ratioER	hospitalization	3.35e-03		0,6	all region
perte odorat_GOOGLE	CCRU admision	8.91e-02		0,6	without IDF an region
proportion of GCCR participants	hospital death	3.16e-03		0,63	without IDF an region
fièvre_GOOGLE	hospital death	7.15e-02		0,63	without IDF an region
proportion of GCCR participants	CCRU admision	2.56e-11		0,64	without IDF an departement
perte odorat_GOOGLE	hospitalization	5.52e-02		0,66	without IDF an region
courbatures_GOOGLE	hospitalization	3.4e-02		0,67	all region
courbatures_GOOGLE	CCRU admision	3.45e-02		0,67	all region
courbatures_GOOGLE	hospital death	3.51e-02		0,67	all region
fièvre_GOOGLE	CCRU admision	4.96e-02		0,67	without IDF an region
toux_GOOGLE	ratioER	4.88e-02		0,67	without IDF an region
proportion of GCCR participants	hospitalization	1.03e-03		0,68	without IDF an region
proportion of GCCR participants without medical diagnostic	hospital death	3.9e-04		0,69	all region
toux_GOOGLE	ratioER	2.71e-02		0,69	all region
fièvre_GOOGLE	hospitalization	3.58e-02		0,7	without IDF an region
fièvre_GOOGLE	CCRU admision	2.43e-02		0,7	all region
fièvre_GOOGLE	hospital death	2.44e-02		0,7	all region
proportion of GCCR participants without medical diagnostic	CCRU admision	1.69e-16		0,72	all departement
proportion of GCCR participants	CCRU admision	3.02e-04		0,72	without IDF an region
proportion of GCCR participants without medical diagnostic	hospitalization	8.13e-05		0,74	all region
fièvre_GOOGLE	hospitalization	1.43e-02		0,74	all region
proportion of GCCR participants	hospital death	5.62e-05		0,75	all region
proportion of GCCR participants	CCRU admision	4.16e-19		0,76	all departement
toux_GOOGLE	CCRU admision	1.08e-02		0,76	all region
perte gout_GOOGLE	CCRU admision	1.42e-02		0,78	without IDF an region
perte odorat_GOOGLE	hospital death	8.23e-03		0,78	all region
perte gout_GOOGLE	hospital death	1.12e-02		0,79	without IDF an region
perte odorat_GOOGLE	ratioER	5.43e-03		0,8	all region
proportion of GCCR participants	hospitalization	5.71e-06		0,81	all region
perte odorat_GOOGLE	CCRU admision	3.97e-03		0,82	all region
toux_GOOGLE	CCRU admision	7.2e-03		0,82	without IDF an region
proportion of GCCR participants without medical diagnostic	CCRU admision	1.65e-06		0,83	all region
perte odorat_GOOGLE	hospitalization	3.07e-03		0,83	all region
toux_GOOGLE	hospital death	2.7e-03		0,83	all region
toux_GOOGLE	hospitalization	2.62e-03		0,84	all region
perte gout_GOOGLE	hospitalization	3.78e-03		0,85	without IDF an region
toux_GOOGLE	hospital death	4.01e-03		0,85	without IDF an region
proportion of GCCR participants	CCRU admision	8.95e-08		0,88	all region
perte gout_GOOGLE	CCRU admision	7.66e-04		0,88	all region
perte gout_GOOGLE	hospital death	7.31e-04		0,88	all region
toux_GOOGLE	hospitalization	1.85e-03		0,88	without IDF an region
fièvre_GOOGLE	ratioER	4.55e-04		0,9	all region
fièvre_GOOGLE	ratioER	8.67e-04		0,9	without IDF an region
perte gout_GOOGLE	hospitalization	1.74e-04		0,92	all region
perte odorat_GOOGLE	ratioER	3.56e-04		0,92	without IDF an region

Supplementary Table 3: National and regional media coverage of 1 the GCCR study in France

type	medium	distribution	date	link
presse	BFM TV	national	14-avr	https://www.bfmtv.com/sante/coronavirus-une-etude-internationale-lancee-
presse	orange	national	14-avr	https://actu.orange.fr/france/lancement-d-une-vaste-enquete-sur-la-perde-
presse	msn	national	14-avr	https://www.msn.com/fr-fr/actualite/coronavirus/coronavirus-une-
presse	BFM / RMC	national	14-avr	https://rmc.bfmtv.com/emission/un-questionnaire-en-ligne-pour-participer-a-
presse	femme actuel	national	14-avr	https://www.femmeactuelle.fr/sante/news-sante/coronavirus-comment-
presse	huffingtonpos	national	14-avr	https://www.huffingtonpost.fr/entry/vous-avez-eu-le-coronavirus-aidez-les-
presse	france info	national	14-avr	https://www.francetvinfo.fr/sante/maladie/coronavirus/coronavirus-une-
presse	20 minutes	national	14-avr	https://www.20minutes.fr/sante/2760607-20200414-coronavirus-enquete-
presse	nouvel obs	national	14-avr	https://www.nouvelobs.com/sante/20200414_OBS27524/coronavirus-perde-
presse	le Figaro	national	14-avr	https://www.lefigaro.fr/flash-actu/covid-19-enquete-internationale-sur-la-
presse	CNRS	national	14-avr	http://www.cnrs.fr/fr/covid-19-lancement-d-une-enquete-internationale-sur-la-
presse	libération	national	14-avr	https://www.liberation.fr/direct/element/une-enquete-internationale-lancee-
blog	medisite	national	14-avr	https://www.medisite.fr/coronavirus-covid-19-la-perde-du-gout-et-de-lodorat-
blog	1001 infos	national	14-avr	https://1001infos.net/france/coronavirus-une-enquete-internationale-lancee-
blog	la gazette du	national	14-avr	https://www.gazettelabo.fr/breves/9706Covid19-enquete-internationale-
blog	myscience	national	14-avr	https://www.myscience.fr/news/wire/covid_19_lancement_d_une_enquete_i
blog	yahoo sport	national	14-avr	https://fr.sports.yahoo.com/news/coronavirus-enqu%C3%AAte-scientifique-
blog	le blob	national	14-avr	https://leblob.fr/actualites/covid-19-une-enquete-mondiale-pour-decrire-
blog	lactu24	national	14-avr	https://lactu24.com/26376-2/
radio	radio scoop	Auvergne / Rh	14-avr	https://radioscoop.com/infos/participez-a-une-enquete-sur-la-perde-de-l-
presse	lyonmag	Rhône-Alpes	14-avr	https://www.lyonmag.com/article/106956/coronavirus-une-etude-sur-la-perde-
radio	impact fm	Rhône-Alpes	14-avr	https://www.impactfm.fr/infos/participez-a-une-enquete-sur-la-perde-de-l-
presse	lyon capitale	Rhône-Alpes	14-avr	https://www.lyoncapitale.fr/actualite/coronavirus-lyon-participe-a-une-
presse	la depeche du	Languedoc-Ro	14-avr	https://www.ladepêche.fr/2020/04/14/covid-19-enquete-internationale-sur-la-
presse	golfe de St Tr	PACA	14-avr	https://golfedesainttrophez.maville.com/actu/actudet_-coronavirus-une-
presse	ma ville dragu	PACA	14-avr	https://draguignan.maville.com/actu/actudet_-coronavirus-une-enquete-
presse	brignoles - m	PACA	14-avr	https://brignoles.maville.com/actu/actudet_-coronavirus-une-enquete-
presse	la provence	PACA	14-avr	https://www.laprovence.com/article/sante/5961521/covid-19-enquete-
blog	cres PACA	PACA	14-avr	http://www.cres-
presse	ouest france	Bretagne / Ce	14-avr	https://www.ouest-france.fr/sante/virus/coronavirus/coronavirus-une-
presse	sante magazin	national	15-avr	https://www.santemagazine.fr/actualites/actualites-sante/coronavirus-une-
presse	notre temps	national	15-avr	https://www.notretemps.com/sante/covid-19-enquete-internationale-sur-la-
presse	latribune	national	15-avr	https://www.latribune.ca/covid-19/participants-recherches-pour-une-
presse	science et ave	national	15-avr	https://www.sciencesetavenir.fr/sante/coronavirus-le-point-du-jour-sur-le-
blog	technoscience	national	15-avr	https://www.techno-science.net/actualite/covid-19-lancement-enquete-
blog	presse19	national	15-avr	http://presse19.centerblog.net/4098-etude-internationale-sur-la-perde-de-
radio	tonic radio	Rhône-Alpes	15-avr	https://www.tonicradio.fr/covid-19-une-etude-lancee-face-a-la-perde-dodorat-
presse	centre jacques	Rhône-Alpes	15-avr	https://centrefjacquescartier.com/actualites/detail/actu/covid-19-traitements-
presse	france region	Rhône-Alpes	15-avr	https://france3-regions.francetvinfo.fr/auvergne-rhone-alpes/coronavirus-
radio	KISS FM	PACA	15-avr	http://www.kissfm.fr/page.asp?id=90
presse	france 3 PACA	PACA	15-avr	https://france3-regions.francetvinfo.fr/provence-alpes-cote-d-
TV	france 3 PACA	PACA	15-avr	https://france3-regions.francetvinfo.fr/provence-alpes-cote-d-
presse	futurasciences	national	16-avr	https://www.futura-sciences.com/sante/breves/coronavirus-covid-19-
presse	yahoo	national	16-avr	https://fr.news.yahoo.com/covid-19-participez-%C3%A0-enqu%C3%AAta-
blog	anosmie.org	national	16-avr	https://www.anosmie.org/2020/04/16/enquetes-publiques-etudes-cliniques-
presse	Le Monde	national	20-avr	https://www.lemonde.fr/sciences/article/2020/04/20/combien-de-cas-de-
radio	France inter	national	17-avr	https://www.franceinter.fr/environnement/detox-la-perde-du-gout-et-de-l-
radio	RCF	national	17-avr	https://rcf.fr/la-matinale/trois-questions-jerome-golebiowski
blog	pourquoi doct	national	17-avr	https://www.pourquoidocteur.fr/Articles/Question-d-actu/32168-Coronavirus-
radio	RFI	national	20-avr	http://www.rfi.fr/ru/%D1%84%D1%80%D0%B0%D0%BD%D1%86%D0%B8
presse	top santé	national	20-avr	https://www.topsante.com/medecine/maladies-
presse	inrae	national	20-avr	https://www.inrae.fr/covid-19
blog	nez - la revue	national	21-avr	https://www.nez-larevue.fr/magazine/actualites/covid-19-le-jour-ou-le-monde-
presse	UCA	PACA	21-avr	http://univ-cotedazur.fr/contenus-riches/actualites/fr/covid-19-lancement-
blog	breakingnews	national	22-avr	https://www.breakingnews.fr/sante/avez-vous-perdu-lodeur-ou-le-gout-des-
blog	seronet	national	23-avr	https://seronet.info/breve/enquete-sur-la-perde-de-lodorat-et-du-gout-87429
presse	actu toulouse	Midi-Pyrénées	23-avr	https://actu.fr/occitanie/toulouse_31555/toulouse-perde-gout-de-lodorat-
presse	france3 region	Languedoc-Ro	27-avr	https://france3-regions.francetvinfo.fr/occitanie/haute-
blog	CPU	national	02-mai	http://www.cpu.fr/actualite/covid-19-la-piste-du-gout-et-de-lodorat/
presse	tribune de lyo	Rhône-Alpes	03-mai	https://tribunedelyon.fr/salade-lyonnaise/article/sommeil-et-confinement-
presse	nice matin	PACA	03-mai	https://www.pressreader.com/france/nice-matin-nice-littoral-et-
presse	doctissimo	national	?	https://www.doctissimo.fr/sante/news/coronavirus-enquete-internationale-
blog	crumpe	national	?	https://www.crumpe.com/2020/04/une-enquete-scientifique-internationale-
presse	cnrs occitanie	Languedoc-Ro	?	https://www.cnrs.fr/occitanie-ouest/actualites/article/alerte-presse-covid-19-
blog	alternatif bien	national	?	https://alternatif-bien-etre.com/urgence-sante-coronavirus-des-chercheurs-
presse	univ lyon	Rhône-Alpes	?	https://popsociences.universite-lyon.fr/ressources/covid-19-odorat-et-qualite-
presse	cnrs centre es	Alsace / Champagne-Ardenn	?	http://www.dr6.cnrs.fr/centre-est/spip.php?rubrique26
presse	cnrs bretagne	Bretagne / Pa	?	https://www.dr17.cnrs.fr/
presse	cnrs aquitaine	Aquitaine	?	http://www.cnrs.fr/aquitaine/
presse	SCF	national	?	https://www.societechimiquedefrance.fr/Le-milieu-academique-se-mobilise-