

1 – CURRICULUM VITAE

Dr. Jérémie Topin

Address 9 montée Pierre Barelli, 06000 Nice
E-mail jeremie.topin@univ-cotedazur.fr
Personal Details *Nationality:* French
Date of Birth: 10/12/1982
Marital status: married, two children



SUMMARY

Academic experiences

Present Associate Professor University Côte d'Azur
Director MSc Flavor & Fragrance
2016-2019 Research Associate UniCA
2012-2014 Postdoctoral fellow CERMAV
2011-2012 ATER University Nice Sophia-Antipolis
2008-2011 PhD UNS

Industrial experiences

2016-Present Consulting for Expressions Parfumées
2023-Present Consulting for Takasago
2018-2023 Consulting for Robertet
2019 Consulting for Virbac
2018 Consulting for IFF
2017 Consulting for INELDEA
2015-2016 Postdoctoral fellow SANOFI

Fields of expertise

Molecular Modeling of biological systems

- Affinity evaluation
- Kinetic modeling
- Allostery
- Virtual screening

Spectroscopic properties

- CD measurements of molecules and protein in liquid, solid and gas phases
- CD prediction by QM calculations

Sensory Analysis

- Recording/analyzing physiological response
- Constructing and analyzing surveys

Scientific society

2022-Present Scientific council Roudnitska Foundation
2021-Present Scientific council SFB (French Society of Biophysics)
2021-Present Scientific council Neuromod Institute

Mentoring

20 Master students
4 PhD finished, and 3 ongoing
2 post-doctorants

Publications

34 articles, 2 book chapters
4 articles in French scientific journal
40 oral communications (11 invited)
1 patent, 1 database

Grant (PI)

2025 Excellence Chair, IDEX
2024 Bioscale (post-doctorant)
2022 CroPiCro, IDEX
2021-2013 ParfEvo, Region SUD
2020 CST UCA and Region SUD
2020-2022-2024 1-week beamtime (Aarhus)
2019 GEN foundation
2019 CSI UCA
2019 Attractivité du territoire IDEX
2016 City of Nice

Grant (co-I)

2026-2029 ANR PHERECIA
2025-2027 project Explorae invORIA
2024-2029 project PARSAD ARDECO
2020-2024 ANR ChEvoSim
2018-2023 ANR Optinose
2019-2020 CNRS MITI
2018 Research grant from the MSI

RESEARCH INTERESTS

My professional experiences allowed me to apply various methods of molecular modeling and chemoinformatic to investigate at the atomic scale the interactions between ligands and receptors. I apply my expertise to several biological systems: peptides, soluble proteins or trans-membrane proteins, and a large family of ligands: gas (O₂/H₂), odorants (small organic molecules) and sugars. My work is based on the perspective of linking *in silico* observations to *in vitro* or *in vivo* results. I then develop a strong ability to exchange and plan experiments with experimental partners from different fields of expertise.

Summary of Career and Practice Context

Name: Topin **Surnames:** Jérémie, Jacques, Daniel, Grigoire, Evariste
Function : Maître de Conférences / Associate Professor **Grade :** 7
Section : 31 / Chemistry

EDUCATION

- 2006–2008** **Master in Chemical Physics, Joseph Fourier University, Grenoble (France)**
Structural study of protein “ALIX” by mass spectrometry coupled with hydrogen deuterium exchange. (3 and 6 months of internship at the Institut de Biologie Structurale under the supervision of Dr L. Signore).
- Identifying interaction region in protein dimer, LC-ESI-TOF, MALDI-TOF.
- 2008–2011** **Ph.D. in Chemistry funded by the ANR project HYLIOX in chemistry, University of Nice Sophia Antipolis , Nice (France)**
Molecular simulations of gas diffusion through hydrogenases. Funded by the ANR HYLIOX under the supervision of Prof S. Antonczak and Prof J. Golebiowski (Institut de Chimie de Nice). **(2 publications and 240 h of teaching)**
- Metallo-enzyme, statistical analysis of diffusion, free energy calculation, kinetic model

RESEARCH EXPERIENCE

- 2011–2012** **Temporary Teaching and Research Associate (ATER), University Nice Sophia Antipolis, Nice (France)**
Molecular simulations of olfactory receptors. In the group of Prof. J. Golebiowski (Institut de Chimie de Nice) in the frame of the Ofactome project certified by 2 competitiveness clusters: PASS and VITAGORA. **(2 publications and 2 book chapters, and 96 hours of teaching)**
- Structure prediction of a class A GPCR, docking and free energy calculation.
- 2012-2015** **Postdoctoral Fellow, Josef Fourier University, Grenoble, (France)**
Understanding recognition and tissue adhesion mechanisms involved in bacterial infections. In the group of Dr. A. Imberty and funded by the Labex ARCANE. **(4 publications)**
- Virtual screening, crystallography, glycan flexibility
- 2015-2016** **Associate researcher, Sanofi-Aventis, Vitry-sur-Seine (France)**
Discover design principles that help in fine-tuning the kinetic parameters of a lead compound. In the group of Dr. H. Minoux, as part of the European IMI consortium, K4DD (Kinetics for Drug Discovery).
- kinetic modelling, metadynamic simulations
- 2016-2019** **Postdoctoral Fellow, Université Côte d’Azur, Nice (France)**
Deciphering the odours/emotions relationships. Funding obtained from the City of Nice to work in Prof. J. Golebiowski’s group (Institut de Chimie de Nice). **(3 publications and 2 publication in a French journal, 60 hours of teaching and 5 M1 internships supervision)**
- Sensory analysis, physiological measurement, protein dynamics
- 2019-present** **Associate Professor, Department of Chemistry, Université Côte d’Azur, Nice (France)**
Deciphering the chemical senses at the molecular level. Co-Head of the research group ChemSenSim at the Institut de Chimie de Nice. **(20 publications, 1 patent, 1 database and 1 publication in a French journal, 192 to 310 hours teaching / year, 6 PhD, 4 M2, 3 M1 and 2 L3 internships supervision)**
- Deciphering the odours/emotions relationships.
- Revealing mechanism of taste and smell receptors activation/inhibition.
- Deciphering the chiroptical properties of life’s building blocks.

Summary of Career and Practice Context

RESEARCH VISITS

Centre for Storage Ring Facilities Aarhus, (Denmark) Beamline CD1: CD & anisotropy spectroscopy in the UV spectral range

- Participation to the project *Amino acid in the gas phase* 7 days of beamtime in 2018 and 2019.
- PI of the project *Following CD spectra of soluble protein during desiccation* 7 days of beamtime in 2020, 2022, 2023, 2024, 2026.

SUPERVISION OF GRADUATE AND POSTDOCTORAL STUDENTS

2016	2 months M1 internship: Julie Breteau (Sorbone Université, France)
2017	2 months M1 internship: Elodie Chiramello (Université Côte d'Azur, France)
2018	2 months M1 internship: Bertille Fourel , Armelle Cugnot , Laura Darner (Université Côte d'Azur, France)
2018	6 months M2 internship: Ourvashi Kissoondoyal (Université Paris Cité, France)
2019	5 months M1 internship: Arianne Salamito (Chimie Paris, France)
2019	2 months M1 internship: Chloé Aubert (Université Côte d'Azur, France)
2019	6 months M2 internship: Eliane Birba (Université Côte d'Azur, France)
2019	6 months M2 internship: Lingling Xiong (Université Côte d'Azur, France)
2020	6 months M2 internship: Clara Devouassoux (Chimie Paris, France)
2020	5 months M1 internship: Mona Behillil (Chimie Paris, France)
2021	6 months M2 internship: Jean-Baptiste Coffin (Université Côte d'Azur, France)
2021	5 months M1 internship: Yuliya Kolomeets (Université Côte d'Azur, France)
2022	2 months M1 internship: Samar Abi Khalil (Université Côte d'Azur, France)
2023	6 months M2 internship: Julie Te (Université Côte d'Azur, France)
2023	5 months M2 internship: Samar Abi Khalil (Université Côte d'Azur, France)
2025	2 months M2 internship: Sebastian Finke (University of Osnabrück, Germany)
2026	6 months M2 internship: Fernando Martinez-Urratia (Université Strasbourg, France)
2018-2022	PhD student (informal co-supervision): Dr. Jody Pacalon , funded by the Roudnitska Foundation, now post-doctorant in the group of Prof. D. Gloriam, University of Copenhagen (Denmark).
2020-2023	PhD student (informal co-supervision) : Dr. Xhino Meleqi , Cifre in collaboration with Robertet, now neuroscientist at Robertet.
2020-2024	PhD student (co-supervision): Dr. Matej Hladiš , funded by the ANR CheMoSim, now post-doctorant in the group of Prof. M. Bronstein, University of Oxford (United Kingdom).
2021-2024	PhD student (informal co-supervision): Dr. Maxence Lalis , funded by the Roudnitska Foundation, now post-doctorant in the group of Ass. Prof. J. Topin, Université Côte d'Azur (France).
2024-present	PhD student (co-supervision): Ulrich Ambassa-Menyié , Cifre in collaboration with Expressions Parfumées.
2024-present	PhD student (co-supervision): Julie Te , funded by the Graduate School of Life & Health Sciences in collaboration with Dr. J. Gilleron from the C3M, Université Côte d'Azur (France)
2025-present	PhD student (informal co-supervision): Marc El-Hasbany , funded by the Roudnitska Foundation.
2025-present	Post-doctorant: Dr. Maxence Lalis (funded by SNA@Bioscale): From molecule to smell.
2026-present	Post-doctorant: Dr. Niccolo Banccone (Funded by the Idex Chair of Excellence): spectroscopic properties of small molecules

INNOVATION AND TECHNOLOGY TRANSFER

2017-present Consulting for [Expressions Parfumées](#) (Odor & Emotions) (30 k€ + 150 k€ for a Cifre)

Summary of Career and Practice Context

2018	Consulting for IFF (Bibliographic research) (5 k€)
2018-2023	Consulting for Robertet (Odor & Emotions) (10 k€ + 115 k€ for a Cifre)
2019	Consulting for Virbac (QM calculation) (6 k€)
2019-present	Consulting for Edeniste (Analysis of perfume-evoked emotions) (10 k€)
2024-2025	Consulting pour Takasago (Application development and numerical modelling) (18 k€)
2024-present	Creation of a start-up Nylix (member of the scientific board) (180 k€)

FELLOWSHIPS & AWARDS

2017	Young Scientist Fellowship Award , ISEO: International Symposium of Essential Oils, Pécs (Hungary)
2019	The Gen Foundation Grants
2024-present	Performance-Based Bonus RIPEC C3
2024	Winner of the DREAM Challenge: Olfactory Mixtures Prediction Challenge
2025	Chair of the Advanced Research Program, IDEX. InibOR : Deciphering the Molecular Basis of Activation and Inhibition in Olfactory Receptors, Academy 4, Université Côte d'Azur (75 k€ + 96h of teaching exemption)
2025	Winner task2 of the DREAM Challenge: Olfactory Mixtures Prediction Challenge

ARTISTIC PRODUCTIONS

[Musée International de la Parfumerie, Grasse \(France\)](#)

2022	Draw Me an Odour research project co-produced between Emma Febvre-Richards from Massey University (New Zealand) and Jérémie Topin, exhibition <i>Respirez l'Art</i> . A multisensory, multimedia, interactive installation around four olfactory compositions.
2024	STEEL LIFE by Carla Barkatz in the frame of the Programme Pilote, an innovative initiative that brings together contemporary artistic creation and scientific research, in partnership with the Villa Arson and the Institut de Chimie de Nice at Université Côte d'Azur.
2025	Draw Me an Odour v2 . In this collaborative artistic project, Emma Febvre Richards and Jérémie Topin invite us to reflect on the invisible. Through the interplay of colours and shapes collected during a previous exhibition, their immersive installation projects olfactory molecules into a visual universe, revealing the fragments of the infinitely small that weave our sensory relationship with the world.

ORGANIZATION OF SCIENTIFIC MEETINGS AND CONFERENCES

2019	Organizing Committee GGMM : Groupe de Graphisme et Modélisation Moléculaire, Nice (France)
2021	Session Chair, GGMM : Groupe de Graphisme et Modélisation Moléculaire, Lille (France)
2024	Symposium Organizer <i>Odors, Sensors, Perception: Convergence with AI</i> , ISOT : International Symposium on Olfaction and Taste, Reykjavik (Iceland)
2024	Organizing Committee Réunion Annuelle du GDR O3 2024 : Groupement de Recherche Odorant, Odeur, Olfaction, Nice (France)
2025	Symposium Organizer Biologically Inspired and bio-based sensors in Olfaction and Taste, ECRO : European Chemoreception Research Organisation, Bilbao (Spain)

INSTITUTIONAL RESPONSIBILITIES

2021-Present	Selected for the Scientific Council of the Neuromod Institute, Université Côte d'Azur The NeuroMod Institute for Modeling in Neuroscience and Cognition aims at promoting modeling as an approach for integrating brain mechanisms and cognitive functions. The institute relies on the interdisciplinary resources available within the Université Côte d'Azur: more than 250 researchers and 16 different laboratories and research entities.
2021-present	Director of the MSc Management of the Flavor & Fragrance Industry, Université Côte d'Azur The MSc F&F is a 100% English-taught program structured to meet the needs of the flavor

Summary of Career and Practice Context

and fragrance industry. It offers two distinct tracks: a two-year Science Track providing in-depth technical training in chemistry, and a one-year Marketing Track focused on business, marketing, and strategic management. The program enrolls 45 students annually, two-thirds of whom are international.

2021-present **Member of the Chemistry Department Steering Committee, Université Côte d'Azur**

The role is to ensure that human resources in chemistry are available and aligned with the specific needs of the programs offered at Université Côte d'Azur.

SCIENTIFIC REVIEWING ACTIVITY

2024 **Topic Editor, Frontiers in Computer Science**

Cutting-Edge Technologies for Multi-Sensory Research at the Frontier Between Disciplines

2018-present **Reviewer** for multi-disciplinary journal: *Plos One*, Modeling journal: *Computational Biology and Chemistry*, *Journal of Chemical Information and Modeling*, Physico-chemistry: *Physical Chemistry Chemistry Physics*

SCIENTIFIC SOCIETY

2022-present **Designated as a member of the Scientific Council of the Roudnitska Foundation, under the aegis of the Fondation de France**

2021-2025 **Elected at the Scientific Council of the SFB: Société Française de Biophysique (French Society of Biophysics)**

PHD JURY PARTICIPATION

2022 **Examiner** for Ines Adrar's PhD thesis: *Pourquoi aimons-nous tant le parfum de la rose ?* CNRL : Centre de Recherche en Neurosciences de Lyon (France)

2023 **Examiner** for Valentin Ghibaudo's PhD thesis : *Le système olfactif, une porte d'entrée pour apaiser le corps et l'esprit* CNRL : Centre de Recherche en Neurosciences de Lyon (France)

SEMINARS

2024 **Olfaction under the lens of a computer microscope** Hanyang Bio Mini Symposium, Hanyang University, Seoul, Republic of Korea.

2025 **Deorphanizing the Human Olfactory Receptor Repertoire: Lessons from M2OR**, Institute of Physics University of Freiburg, Germany.

2025 **Deorphanizing the Human Olfactory Receptor Repertoire: Lessons from M2OR**, National University of Seoul Republic of Korea.

FUNDINGS AS PRINCIPAL INVESTIGATOR

2016 **Post-doctoral Fellowship** from the city of Nice (100 k€)

2018 **InsectOR**, *Étude structure-fonction computationnelle et expérimentale des récepteurs olfactifs d'insectes*, MITI (Mission for interdisciplinary initiatives) CNRS (15 k€)

2019 **Territorial attractiveness** from the Academy 2, Université Côte d'Azur (15 k€)

2019 **StimOlf**, *Stimulation Olfactive pour le traitement de l'apathie*, funded by CSI (Incentive-based research), Université Côte d'Azur (15 k€)

2020 **A la découverte des sens chimiques !**, Université Côte d'Azur (3 k€), Region SUD (3 k€)

2021 **ParfEmo**, *Élaboration de Compositions Odorantes Modulatrice de l'Émotion*, Region Sud (27 k€)

2022 **CroPiCro**, *Conception Rationnelle de réPulsifs ciblant les Récepteurs Olfactifs d'insecte*, Booster « Junior » Academy 2, Université Côte d'Azur (18 k€)

2022 **Travel Grant**, Neuromod, Université Côte d'Azur (1 k€)

2022 **PRED-IA-TAAR**, *Déchiffrer le code combinatoire d'activation des récepteurs aux « traces amines »*, Neuromod, Université Côte d'Azur (7.5 k€)

2023 **PhD International Mobility Program « DocWalker »**, Two-month stay at the Monell Chemical

Summary of Career and Practice Context

	Senses Center (M. Hladiš and M. Lalis) (6 k€)
2023	FUNOBP , <i>Functional study of OBP</i> , CSI (Université Côte d'Azur), 5k€
2023	KillRab , <i>Identify compounds able to repress a Rab protein-dependent negative feedback loop on PPARα transcription to improve MAFLD therapies</i> , Academy 4, Université Côte d'Azur (24 k€)
2024	Jeune Docteur Innovant (Promising young scientist), <i>From molecule to smell</i> , Region SUD (37 k€) and SATT (12 k€)
2024-2025	KinetChemSen , <i>Molecular kinetics in chemical senses</i> , PHC PROCOPE in collaboration with Steffen Wolf (Freibourg University) (10 k€)
2025	Olfactory AI , <i>Bio-inspired AI for accurate odor perception</i> , Bioscale (80 k€) and SATT (20 k€)
2025	MIMOSA , <i>Molecules Improving the Motility Of Sperm for Assisted reproduction</i> , Academy 4, Université Côte d'Azur (24 k€)
2025	Travel Grant , Neuromod, Université Côte d'Azur (2 k€)
2025	Draw me an odour v2 , Art & Science, Université Côte d'Azur (2 k€)
2025	Rabinbox , <i>Identification of the partners of a Rab protein repressing lipid β-oxidation</i> , Academy 4, Université Côte d'Azur (24 k€)
2026-2027	DecibOR , <i>Deciphering the Molecular Basis of Activation and Inhibition in Olfactory Receptors</i> , PHC STAR (19 k€) in collaboration with Hee-Jung Choi (Seoul University)

FUNDINGS AS CO-INVESTIGATOR

2018-2023	Optinose , <i>Odorant Binding Proteins-Based Biomimetic Optical Nose</i> , ANR-PRCE CES 42. WP leader (100 k€)
2020-2024	ChEemoSim , <i>Odorant Binding Proteins-Based Biomimetic Optical Nose</i> , ANR-PRCE CES 42. Co-coordinator (100 k€)
2024-2025	DeepGPCRLig , <i>GNN-NLP approach for predicting GPCR-ligand interactions and taking concentration effects into account</i> , GENCI
2024-2029	ARDECO , <i>Anticipating the withdrawal of phytosanitary substances through the development of a distributed, generic and high-speed Operational Chemical Ecology infrastructure</i> , FranceAgriMer, co-coordinator WP (700 k€)
2025-2029	InvORIA , <i>Reverse chemical ecology using AI-guided high-throughput screening of insect olfactory receptors</i> , France2030 – Explorae (INRAE), co-coordinator WP (400 k€)
2025-209	PHERECIA , <i>Targeting pheromone receptors by machine learning-assisted reverse chemical ecology: a promising strategy for crop pest control</i> , ANR-PRCE CE S20. Co-coordinator WP (200 k€).

Pedagogical investment / Teaching Details

My teaching activities span from the first year of the Bachelor's program (L1) to the second year of the Master's program (M2) in the *Chemistry track*, where I deliver lectures, lead tutorial sessions, and supervise practical lab work. The topics I cover include *Atomistic* (tutorials and lab sessions in L1), *Molecular Modeling* (lectures and lab sessions in L3), *Chemical Senses* (lectures in L1), *Flavor & Fragrances* (lecture and tutorial in M1), *Olfaction Science* (lectures and lab sessions in M2).

I also teach in the MSc *Management of the Flavor and Fragrance Industry*, where I deliver courses on the *Anatomy of the Chemical Senses* (lectures, M1), *Olfactory Receptor Function* (lectures and lab sessions, M2), and the *Physico-Chemical Properties of Olfactory and Gustatory Molecules* (lectures, M1). In addition, I supervise various group projects, which provide students the opportunity to work collaboratively on real-world challenges that combine chemistry, formulation, regulatory aspects, and marketing. For example, students have developed a home fragrance for the *Hotel Le Negresco* in Nice, and created the olfactory signature of Université Côte d'Azur. I also provide online instruction in *Scientific English*.

My teaching activities cover a wide range of topics in chemistry and biology. In *Atomistic* tutorials, I explore the fundamental principles of atomic structure and chemical bonding, giving students a solid foundation to understand interactions at the molecular level. In the third-year courses, I focus on molecular modeling, which, in coordination with my colleagues in spectroscopy, allows us to create links between these disciplines and explore shared concepts through complementary approaches.

At the same time, my teaching also covers biological topics, ranging from cell biology to more specific subjects such as physiology and anatomy. I provide a detailed exploration of cellular mechanisms and organ functions, helping students grasp the complexity of living systems. One area I am particularly passionate about is the study of olfaction and its anatomy, which allows me to teach this theme from L1 to M2. By analyzing the neurobiological processes involved in olfactory perception, I help students understand the molecular and physiological foundations of this essential sensory modality.

I am motivated to bring novelty and innovation to my teaching by integrating various pedagogical approaches I encountered during my 32-hour training program *on academic teaching practices* (in 2019). For example, I use digital tools such as *Moodle* to provide multiple-choice quizzes and self-paced online courses. This method allows students to work at their own pace and strengthen their understanding of key concepts.

Moreover, during lab sessions on chemical senses (L1, M1, and M2), I use the interactive platform *Wooclap* to promote student engagement and interaction. This tool enables me to ask real-time questions and gather student responses, allowing me to immediately adapt the session to revisit any misunderstood topics. This interactive approach fosters active student participation and enhances the learning experience.

TEACHING WITHIN THE GENERAL CHEMISTRY PROGRAM

In the L1 tutorials in *Atomic Structure* (2 groups of 25 students for a total of 52 hours of tutorials and 36 hours of lab work), I encourage student autonomy by involving them actively—either through oral participation or by coming to the board. I complement these sessions with real-world examples that illustrate the concepts discussed in class.

For the *Chemical Senses* course in L1 (~50 students, 8 hours of lectures and 24 hours of lab sessions), the course covers historical, chemical, and biological concepts. The objective of this elective is to demonstrate the need for an interdisciplinary approach to science. Students are assessed through lab reports (e.g., sensory analysis, forced-choice tests) and a critical review of a scientific publication. The goal is to help students develop work and writing methods. To foster interaction, students are invited to participate via the *Wooclap* app, which allows for anonymous response collection.

Pedagogical investment / Teaching Details

The L3 course in *Molecular Modeling* (~25 students, 12 hours of lectures and 12 hours of lab work) introduces various methodologies such as quantum chemistry, molecular docking, molecular modeling, and machine learning. Lab sessions are structured in pairs: the first focuses on software usage, and the second on applying it to a real-world case. Assessment is based on lab reports.

In M1 (~32 students, 9 hours of lectures and 4 hours of tutorials), I continue teaching in the *Flavor and Fragrance* field with a more molecular focus. The aim is to provide students with a solid understanding of the molecular basis of olfactory and gustatory recognition. The course is supplemented with sensory analysis exercises, during which students evaluate series of compounds. *Wooclap* is again used to facilitate responses and break down preconceptions. Students also have access to self-assessment quizzes on *Moodle* related to enzymatic processes. The final assessment consists of a multiple-choice test.

In M2 (~15 students), the *Olfaction Science* course (8 hours of lectures and 12 hours of tutorials, in English) addresses questions of chemosensory receptor specificity and affinity. I also cover the neural circuits responsible for emotion and memory. Students are assessed through a lab report and a multiple-choice test.

TEACHING WITHIN THE MSc MANAGEMENT OF THE FLAVOR & FRAGRANCE INDUSTRY

I teach twice in the M1 *Sciences track* (~15 students). During these lectures, I introduce the basics of the *Anatomy of the Chemical Senses* (8 hours of lectures) and *Physico-Chemical Properties of Olfactory and Gustatory Molecules* (14 hours of lectures). These courses, given during the first week, provide students with a general overview of knowledge in the Flavor & Fragrance field and a solid physico-chemical foundation, which is further developed by other instructors. Lectures are illustrated with videos, and some key concepts are accompanied by olfactory sessions. Students are assessed via multiple-choice quizzes on *Moodle*.

The M2 course builds on the molecular foundations of olfactory recognition, focusing on the function of olfactory and gustatory receptors (15 hours of lectures). Assessment is based on a multiple-choice quiz on *Moodle* and a lab report.

For the Marketing track, the *Chemical Analysis* course introduces fundamental scientific concepts to students who have not studied subjects like math, physics, chemistry, or biology since high school.

As program director, I also chair the juries for the tutored projects. These group projects (3 to 4 students per group) place students in industry-like conditions. They must respond to a company brief and work collaboratively to develop a project that addresses marketing, regulatory, and formulation aspects. The jury consists of faculty members and partner company representatives. Students present a prototype along with a complete plan for production and marketing.

I also lead the defense juries for the end-of-internship presentations of MSc students (~40 students). This includes students from M1, M2 Sciences, and M2 Marketing tracks. I evaluate both the written reports and oral presentations, which take place at the beginning of July.

Pedagogical investment / Teaching Details

SUMMARY OF TEACHING DUTIES

Year 2019-2020

Level	Diploma	Name	Type	Nb of students	Annual workload (in hours)
L1	Licence Sciences et Technologie	<i>Atomistic</i>	TD & TP	30	28 & 18
L1	Licence Sciences et technologie	<i>Structure and representation of matter</i>	TD & TP	30	20 & 24
M2	Master Fragrances & Fine chemistry	Natural and synthetic fragrant raw material	CM & TD	15	3 & 9
MSc1	MSc F&F	Business & foreign languages	CM	20	18
MSc2	MSc F&F	Business & foreign languages	CM	10	4
MSc2	MSc F&F	Natural Raw Materials	CM	10	10

Year 2020-2021

Level	Diploma	Name	Type	Nb of students	Annual workload (in hours)
L1	PO1 Sciences, ingénierie, technologie, environnement	<i>Atomistic</i>	TD & TP	30	28 & 18
L1	PO1 Sciences, ingénierie, technologie, environnement	<i>Structure and representation of matter</i>	TD & TP	30	24 & 18
L1	PO1 Sciences, ingénierie, technologie, environnement	<i>Chemical Senses</i>	CM & TD	50	7,5 & 27
L3	Chimie	<i>Molecular Modeling</i>	CM & TD	30	6 & 6
M1	Chimie	<i>Flavor and Fragrances</i>	CM & TD	35	9 & 3,5
M2	Fragrances & fine chemistry	Olfaction Sciences	CM	15	6 & 12
MSc1	MSc F&F	Flavor & Fragrance general knowledge	CM & TD	13	8 & 6
MSc2	MSc F&F	Flavor & Fragrance general knowledge	TD	19	8
MSc2	MSc F&F	Flavor & Fragrance chemistry	CM	19	18
MSc2	MSc F&F	Practical Work	TP	19	2
DU	Phytothérapie & Aromathérapie	<i>Physiology of the chemical senses</i>	CM	40	1,5

Pedagogical investment / Teaching Details

Year	Level	Diploma	Name	Type	Nb of students	Annual workload (in hours)
2021-2022	L1	PO1 Sciences, ingénierie, technologie, environnement	Atomistique	TD & TP	30	28 & 22
2021-2022	L1	PO1 Sciences, ingénierie, technologie, environnement	Structure et représentation de la matière	TD & TP	30	24 & 12
2021-2022	L1	PO1 Sciences, ingénierie, technologie, environnement	Sens chimique	CM & TD	50	8 & 24
2021-2022	L3	Chimie	Modélisation moléculaire	CM & TD	28	12 & 12
2021-2022	M1	Chimie	Arômes et parfums	CM & TD	32	9 & 4
2021-2022	M2	Fragrances & fine chemistry	Olfaction Sciences	CM	12	6 & 12
2021-2022	MSc1	MSc F&F	Flavor & Fragrance general knowledge	CM & TD	7	12
2021-2022	MSc1	MSc F&F	Physico chemical properties	CM	7	12
2021-2022	MSc2	MSc F&F	Professional project	TD	13	6
2021-2022	MSc2	MSc F&F	Receptomic & polymorphisms	CM	13	10
2021-2022	DU	Phytothérapie & Aromathérapie	Physiologie des sens chimiques	CM	40	1,75

Year	Level	Diploma	Name	Type	Nb of students	Annual workload (in hours)
2022-2023	L1	PO1 Sciences, ingénierie, technologie, environnement	Atomistique	TD & TP	30	28 & 22
2022-2023	L1	PO1 Sciences, ingénierie, technologie, environnement	Structure et représentation de la matière	TD & TP	30	24 & 12
2022-2023	L1	PO1 Sciences, ingénierie, technologie, environnement	Sens chimique	CM & TD	42	8 & 24
2022-2023	L3	Chimie	Modélisation moléculaire	CM & TD	27	12 & 12
2022-2023	M1	Chimie	Arômes et parfums	CM & TD	24	9 & 3,5
2022-2023	M1	Génétique & Développement	Neurophysiology of aging	CM	15	2
2022-2023	M2	Fragrances & fine chemistry	Olfaction Sciences	CM	14	6 & 12
2022-2023	MSc1	MSc F&F	Flavor & Fragrance general knowledge	CM	15	8
2022-2023	MSc1	MSc F&F	Physico chemical properties	CM	15	12
2022-2023	MSc1	MSc F&F	Languages	CM	15	6
2022-2023	MSc1	MSc F&F	Professional project	TD	15	12
2022-2023	MSc1	MSc F&F	Internship	CM	15	10
2022-2023	MSc1	MSc F&F	Tutored project	TD	15	12
2022-2023	MSc2	MSc F&F	Receptomic & polymorphisms	CM	6	10

Pedagogical investment / Teaching Details

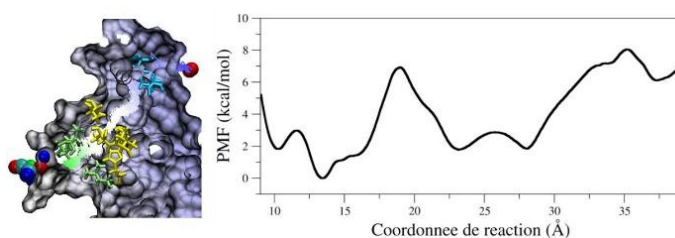
2022-2023	MSc2 marketing track	MSc F&F	Analytical science	CM	8	8
2022-2023	DU	Phytothérapie & Aromathérapie	Physiologie des sens chimiques	CM	40	1,75

2 – OVERVIEW OF RESEARCH TOPICS

2.1 RESEARCH DURING UNDERGRADUATE AND GRADUATE TIME

With an academic background rooted in physical chemistry, I initially focused on experimental biophysical techniques, particularly the development and application of mass spectrometry methods coupled with hydrogen-deuterium exchange (HDX-MS). This approach enabled the identification and characterization of protein–protein interaction interfaces at a structural level, offering valuable insights into molecular recognition mechanisms.

I subsequently specialized in theoretical and computational chemistry during my PhD and the following two postdoctoral research experiences. This phase of my career allowed me to acquire in-depth expertise in molecular modeling, quantum chemical calculations, and simulation techniques aimed at understanding biomolecular structure and reactivity.

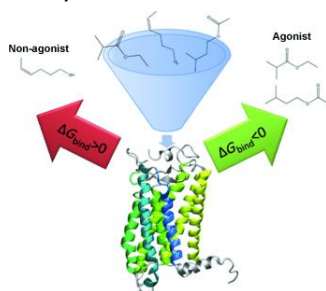


On the left, the O_2 trajectory through the Wild Type protein during an US simulation. On the right, the potential of mean force

and Umbrella Sampling) were applied to decipher the free energy profile associated to the diffusion. Based on the energy results, we have established a kinetic model which allows ranking the mutants according to their tolerance to oxygen. This model is able to qualitatively reproduce the results obtained by electrochemistry, and thus connects experimentally determined phenotypes to their microscopic behaviour.

2.2 RESEARCH AS A POSTDOCTORAL RESEARCHER

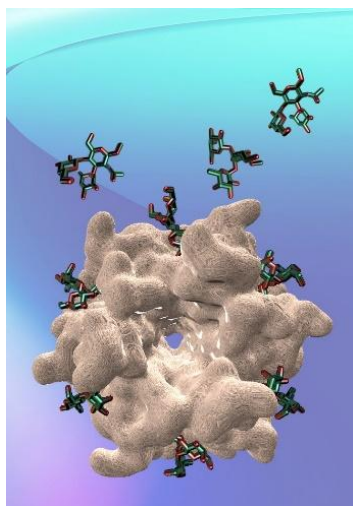
During the following year after my PhD, I was studying the behaviour of the proteins involved in the perception of smell: olfactory receptors (ORs) and odorant binding protein (OBP). Multiple Molecular Dynamics simulations were carried out to provide a detailed description of the ligand entry into the OBP.



Subsequently, an Olfactory Receptor (GPCR class A) has been built by homology modeling and also by an ab initio structure prediction. The free energy of binding of ten odorants was computed by means of molecular-dynamics simulations. This state function allows separating the eight agonists from the two non-agonists for which potency was determined by calcium-imaging assays. This study constitutes a proof-of-principle for the computational deorphanization of olfactory receptors.

My postdoctoral work at **CERMAV** (Université Josef Fourier, Grenoble, France) focused on the molecular mechanisms governing carbohydrate-protein interactions, a central process in bacterial recognition and host tissue adhesion during infection. This research addressed key questions in glycobiology, particularly the role of lectins in decoding the glycode through the specific recognition of oligosaccharides. I investigated virulence-associated lectins such as LecB from *Pseudomonas aeruginosa* and BambL from *Burkholderia ambifaria*, combining structural biology, molecular modeling, and thermodynamic approaches. I established and validated a robust computational workflow based on docking and MM-GBSA free energy calculations to predict the affinity of fucosylated oligosaccharides for lectins. The calculated binding free energies were in good

Overview of Research Topics



agreement with experimental data obtained by isothermal titration calorimetry, demonstrating the reliability of the approach. Structural predictions were further validated by comparison with available crystallographic data.

In parallel, I studied the unusual high-energy conformation of Lewis x observed in the crystal structure of the BamBL/Lewis x complex. Extensive molecular dynamics simulations were performed to investigate glycan flexibility in solution and in complex with lectins. In particular, I showed that the Lewis x trisaccharide, which is rigid in solution, undergoes rare but functionally relevant conformational openings upon binding, facilitated by distortion of the central N-acetylglucosamine ring and guided by a conserved tryptophan residue in the lectin binding site. These results highlighted the importance of oligosaccharide conformational adaptation in molecular recognition

and provided conceptual and methodological frameworks for the rational design of anti-infective glyco-compounds and for the interpretation of high-throughput glycan array data.

During my postdoctoral position at **Sanofi**, I conducted research at the interface of computational chemistry and drug discovery, with a focus on the kinetic aspects of drug–target interactions. The objective of this work was to identify design principles enabling the fine-tuning of kinetic parameters of lead compounds, such as association and dissociation rates, which are increasingly recognized as critical determinants of in vivo efficacy and clinical success.

This research addressed the growing awareness that traditional affinity-based optimization strategies are insufficient to fully capture drug performance in complex biological systems. By integrating binding kinetics into lead optimization workflows, my work contributed to a more mechanistic understanding of drug efficacy and safety, and to the development of predictive strategies aimed at reducing attrition rates in drug discovery. This experience strengthened my expertise in translating fundamental biophysical concepts into applied pharmaceutical research and reinforced my scientific autonomy in multidisciplinary environments.



In my most recent postdoctoral position in **Université Nice Sophia Antipolis**, I broadened my research scope to include human studies, developing expertise in the assessment of emotional states through physiological measurements, such as heart rate, skin conductance, and other autonomic indicators, linking chemical signals to affective responses.

I investigate how olfactory stimulation modulates physiological processes under the control of the autonomic nervous system, using measures such as heart rate, skin conductance, skin temperature, respiratory rhythm, and other autonomic indicators. These physiological readouts provide objective markers of affective states, allowing emotional responses to be studied beyond purely subjective sensory descriptors. A major outcome of this research activity has been the development of an original experimental framework combining immersive virtual reality with physiological measurements to both quantify and qualify emotional responses to sensory stimulation. This methodological breakthrough was achieved through a close collaboration with Robertet, within the framework of the CIFRE PhD project of Xhino Meleqi, and represents a significant step toward objective, reproducible, and context-aware assessment of emotional responses to chemical stimuli.

2.3 RESEARCH AS ASSOCIATE PROFESSOR

Since obtaining a permanent academic position, I have continued to diversify my methodological toolkit by engaging in synchrotron-based spectroscopy. In particular, I conduct circular dichroism (CD) experiments on both isolated molecules and complex biomolecules such as proteins, in order to investigate their conformational properties and chiroptical responses under various conditions.

In parallel, I have established multiple collaborations with key industrial stakeholders in the flavor & fragrance sector. These partnerships have led to joint research initiatives supported by national funding agencies, such as participation in two ANR projects (ANR PRCE Chemosim in collaboration with Expressions Parfumées, and ANR Optinose in collaboration with Aryballe), as well as the co-supervision of CIFRE-funded doctoral thesis (Xhino Meleqi, in partnership with Robertet and Ulrich Ambassa Menyié, in partnership with Expressions Parfumées). In addition, I have conducted contract research and service activities for several companies, including Takasago, Expressions Parfumées, and Robertet.

Today, my primary research activity focuses on numerical modeling, within which I conceive, structure, and supervise research projects based on various approaches. I am responsible for defining scientific questions, designing the computational strategies, and overseeing their implementation by doctoral students and postdocs engaged in molecular modeling and simulation. In this context, I ensure both the scientific rigor of the numerical methods and the conceptual coherence of the projects, while guiding PhD candidates toward autonomy and high-level expertise.

My research relies on the organization and development of advanced computational methodologies to investigate biophysical questions through atomic-scale simulations of molecular interactions and dynamics. By combining molecular modeling with experimental data, I aim to uncover the fundamental mechanisms governing biological processes, including protein function, ligand recognition, and allosteric regulation in complex biomolecular systems. This integrative approach enables a deeper understanding of structure–function relationships that are central to sensory biology, drug discovery, and molecular medicine.

I also coordinate interdisciplinary collaborations with multiple academic and institutional partners, acting as the scientific interface between numerical modeling, experimental biophysics, and molecular biology. This role involves harmonizing methodologies, guiding data interpretation, and ensuring the effective integration of computational results into broader collaborative research programs. To support this work, I rely on a strong network of national collaborators (Loïc Briand's group, CSGA; Christophe Moreau's group, IBS; Yanxia Hou-BROUTIN's group, CEA ; Emmanuelle Jacquin-Joly, INRAE) and international partners (Cheil Moon, DGIST, Republic of Korea; Hee-Jung Choi, SNU, Republic of Korea, Antonella Di Pizzo, TUM, Germany, Steffen Wolf, Frei Universität Freiburg, Germany). Through these activities, I have developed and consolidated a coherent and autonomous research program centered on the use of advanced computational methods to address fundamental biophysical questions.

My main topic of research is to investigate the structure-function relationships in proteins involved in chemosensory perception. Our main results include:

- i) the creation of the most complete database gathering olfactory receptor / odorant pairs
- ii) the development of a model capable of predicting the activity of any molecule on any olfactory receptor-representing a key step toward deciphering the combinatorial code of olfaction
- iii) the characterization of both the diffusion pathway and binding site in protein involved in the sense of smell

This research focus is complemented by studies on emotional responses to olfactory stimulation, using physiological measurements regulated by the autonomic nervous system. This line of research is being developed in close collaboration with companies in the aroma and fragrance industry

Overview of Research Topics

located in the Grasse region.

I am currently expanding my research to address medical questions. I am collaborating with biologists from Université Côte d'Azur (Jérôme Gilleron, C3M) on a project focused on identifying inhibitors of proteins involved in fatty liver disease (project IDEX KillRab and PhD funding for Julie Te). My expertise in this collaboration lies in numerical modeling to understand allosteric events in proteins and to help identify potential inhibitors.

I also expanded my collaboration with biologists from Université Côte d'Azur, notably Romain Barrès (IPMC), on a project aimed at enhancing the quality or fertilization potential of human sperm by targeting olfactory receptors (ORs) expressed on the surface of spermatozoa (project IDEX MIMOSA).

I am also involved in another research theme at the Institut de Chimie de Nice: the origin of biological asymmetry. In this context, I investigate the spectroscopic properties of the building blocks of life, combining synchrotron-based measurements with quantum chemical calculations. This work has led to the characterization of the spectroscopic properties of amino acids in the gas phase.

2.4– LIST OF PUBLICATIONS

Articles in international scientific journal ([‡] co-first, * corresponding author)

- [34] C. Cornut, C. Belloir, A. Karolkowski, M. Lalis, S. Chometton, S. Fiorucci, J. Topin*, L. Briand. Sweet and Umami TAS1R Receptors: From Molecular Recognition to Physiological Function. [*Chemical Senses*](#), **2026**.
- [33] A. Kamal, M. Hladiš, J. Topin, M. Plantevit, S. Fiorucci, C. Robardet. Explainability of Molecular Graph Neural Network. [*IEEE 12th International Conference on Data Science and Advanced Analytics \(DSAA\)*](#), Birmingham, United Kingdom, **2025**.
- [32] S. Fiorucci, J. Topin, J.-C. Sakdavong, B. Mesz. Editorial: Cutting-Edge Technologies for Multi-Sensory Research at the Frontier Between Disciplines. [*Frontiers in Computer Science*](#), **2025**.
- [31] C. Cornut, A. Karolkowski, M. Lalis, A. Thomas, R. Menin, J. Topin, L. Briand, C. Belloir. Glutathione as a taste modulator: molecular mechanisms of interaction with umami and sweet taste receptors. [*Food Chemistry: Molecular Sciences*](#), 100319, **2025**.
- [30] H. Benkhatar, J. Topin, I. Parrot, S. Fiorucci. Targeting Highly Expressed Olfactory Receptors to Improve Olfactory Testing and Training. [*Rhinology*](#), **2025**.
- [29] M. Lalis, L. Moitrier, M. Jäger, C. Meinert, M. Brulé, C. Belloir, N. C. Jones, S. V. Hoffmann, S. Fiorucci, S. Wolf, L. Briand, J. Topin*. How Allosteric Mutations Control Ligand Binding in Lipocalin Protein: Odorant Binding Protein as a Test Case, [*Cellular and Molecular Life Science*](#), **2025**.
- [28] A. Comte, M. Lalis, L. Brajon, R. Moracci, N. Montagné, J. Topin, E. Jacquin-Joly, S. Fiorucci. Accelerating Ligand Discovery for Insect Odorant Receptors. [*International Journal of Biological Science*](#), **2025**.
- [27] A. Lagunas, C. Belloir, M. Lalis, L. Briand, J. Topin, P. Gorostiza, J. Samitier. Ligand discrimination in hOR1A1 based on the capacitive response, [*Biosensors and Bioelectronics*](#), **2025**, vol. 271, p. 117000. DOI: 10.1016/j.bios.2024.117000
- [26] M. El Kazzy, M. Lalis, C. Hurot, J. S. Weerakkody, R. Mathey, C. Saint-Pierre, A. Buhot, T. Livache, J. Topin, L. Moitrier, C. Belloir, L. Briand, Y. Hou. Study and optimization of the selectivity of an odorant binding protein-based bioelectronic nose, [*Biosensors and Bioelectronics*](#), **2025**, p. 116879. <https://doi.org/10.1016/j.bios.2024.11687>
- [25] M. Lalis, M. Hladiš, S. A. Khalil, C. Deroo, C. Marin, M. Bensafi, N. Baldovini, L. Briand, S. Fiorucci, J. Topin*. A status report on human odorant receptors and their allocated agonists, [*Chemical Senses*](#), **2024**, Jan 1;49:bjae037. doi: 10.1093/chemse/bjae037. PMID: 39400708.
- [24] M. Lalis, M. Hladiš, S. A. Khalil, L. Briand, S. Fiorucci, J. Topin*. M2OR: a database of olfactory receptor–odorant pairs for understanding the molecular mechanisms of olfaction, [*Nucleic Acids Research*](#), **2024**, <https://doi.org/10.1093/nar/gkad886>.
- [23] J. Pacalon, G. Audic, J. Magnat, M. Philip, J. Golebiowski, C. J. Moreau, J. Topin*. Elucidation of the

Overview of Research Topics

- structural basis for ligand binding and translocation in conserved insect odorant receptor co-receptors. *Nature Communications*. 14, 8182, **2023**, <https://doi.org/10.1038/s41467-023-44058-5>.
- [22] Bogdan, R. Fabre, T. Desmidt, J. Golebiowski, J. Topin, I. Bethus, ... & R. David. Different Trajectories of Apathy and Depression Among Subjective Cognitive Impairment Individuals with or without Conversion to Dementia: Results from the Memento Cohort in France. *Journal of Alzheimer's Disease*, 1-12. **2023**
- [21] J. Bocková, N. C. Jones, J. Topin, S. V. Hoffmann, & C. Meinert. Uncovering the chiral bias of meteoritic isovaline through asymmetric photochemistry. *Nature Communications*. 14(1), 338, **2023**, <https://doi.org/10.1038/s41467-023-39177-y>.
- [20] M. Hladiš, M. Lalis, S. Fiorucci, & J. Topin*. Matching receptor to odorant with protein language and graph neural networks. *The Eleventh International Conference on Learning Representations*. **2023**
- [19] L. Moitrier, C. Belloir, M. Lalis, Y. Hou, J. Topin*, & L. Briand. Ligand Binding Properties of Odorant-Binding Protein OBP5 from *Mus musculus*. *Biology*, 12(1), 2, **2022**, <https://doi.org/10.3390/biology12010002>.
- [18] D. Garcia[‡], J. Topin[‡], J. Bocková, N. C. Jones, S. V. Hoffmann, & C. Meinert. Chiroptical activity of gas-phase propylene oxide predicting the handedness of interstellar circular polarization in the presolar nebula. *Science Advances*, 8(46), eadd4614, **2022**, [10.1126/sciadv.add4614](https://doi.org/10.1126/sciadv.add4614).
- [17] T. Shim, J. Pacalon, W. C. Kim, X. Cong, J. Topin*, J. Golebiowski, & C. Moon. The Third Extracellular Loop of Mammalian Odorant Receptors Is Involved in Ligand Binding. *International Journal of Molecular Sciences*, 23(20), 12501, **2022**, <https://doi.org/10.3390/ijms232012501>.
- [16] Y. Yu, Z. Ma, J. Pacalon, L. Xu, W. Li, C. Belloir, J. Topin, L. Briand, J. Golebiowski, X. Cong. Extracellular loop 2 of G protein-coupled olfactory receptors is critical for odorant recognition. *Journal of Biological Chemistry*, 298(9), **2022**, DOI: 10.1016/j.jbc.2022.102331.
- [15] C. Meinert, A. D. Garcia[‡], J. Topin[‡], N. C. Jones, M. Diekmann, R. Berger, L. Nahon, S. V. Hoffmann, U. J. Meierhenrich. Amino acid gas phase circular dichroism and implications for the origin of biomolecular asymmetry, *Nature Communications* 13, 502, **2022**, doi.org/10.1038/s41467-022-28184-0.
- [14] J. Topin*, C. Bouysset, J. Pacalon, Y. Kim, M. Rhyu, S. Fiorucci, J. Golebiowski. Functional Molecular Switches of Mammalian G Protein-Coupled Bitter-Taste Receptors, *Cellular and Molecular Life Sciences*, **2021**, DOI: 10.1007/s00018-021-03968-7.
- [13] S. Kuhaudomlarp, E. Siebs, E. Shanina, J. Topin, I. Joachim, P. da Silva Figueiredo Celestino Gomes, A. Varrot, D. Rognan, C. Rademacher, A. Imberty, A. Titz. Non-Carbohydrate Glycomimetics as Inhibitors of Calcium(II)-binding Lectins, *Angewandte Chemie International Edition*, **2020**, doi.org/10.1002/ange.202013217.
- [12] C. Bushdid, C. A. de March, J. Topin, M. Do, H. Matsunami, J. Golebiowski. Mammalian class I odorant receptors exhibit a conserved vestibular binding pocket, *Cellular and Molecular Life Sciences*, **2019**, DOI:10.1007/s00018-018-2996-4.
- [11] C. A. de March[‡], J. Topin[‡], E. Bruguera, G. Novikov, K. Ikegami, H. Matsunami, J. Golebiowski. Odorant Receptor 7D4 Activation Dynamics, *Angewandte Chemie International Edition*, **2018**, Apr 16;57(17): 4554-4558. DOI: 10.1002/anie.201713065.
- [10] X. Cong, J. Topin, J. Golebiowski. Class A GPCRs: Structure, function, modeling and structure-based ligand design. *Current Pharmaceutical Design*, **2017**, issn 1381-6128/1873-4286, DOI: 10.2174/1381612823666170710151255.
- [9] J. Topin[‡], M. Lelimosin[‡], J. Arnaud, A. Audfray, S. Pérez S, A. Varrot, A. Imberty. The Hidden Conformation of Lewis x, a Human Histo-Blood Group Antigen, Is a Determinant for Recognition by Pathogen Lectins. *ACS Chemical Biology*, **2016**, 11 (7), pp 2011–2020, DOI: 10.1021/acschembio.6b00333.
- [8] J. Topin, C. A. de March, L. Charlier, C. Ronin, S. Antonczak, J. Golebiowski. Discrimination between Olfactory Receptor agonists and non-agonists. *Chemistry-A European Journal*, **2014**, DOI: 10.1002/chel.201402486.
- [7] A. Novoa, T. Eierhoff, J. Topin, A. Varrot, S. Barluenga, A. Imberty, W. Römer, N. Winssinger. Novel

Overview of Research Topics

- LecA Ligand Identified from Glycan Array Inhibits Host Cell Invasion by *Pseudomonas Aeruginosa*. [*Angewandte Chemie International Edition*](#), **2014**, DOI: 0.1002/ange.201402831.
- [6] J. Topin, J. Diharce, S. Fiorucci, S. Antonczak, J. Golebiowski. O₂ Migration Rates in [NiFe] Hydrogenases. A joint approach combining free energy calculations and kinetic modeling. [*The Journal of Physical Chemistry B*](#), **2014**, DOI: 10.1021/jp4093964.
- [5] N. Le Breton, M. Martinho, K. Kabytaev, J. Topin, D. Blocquel, S. Longhi, A. Rockenbauer, J. Golebiowski, B. Guigliarelli, S. R. A. Marque & V. Belle. Diversification of EPR signatures in site directed spin labeling using a β -phosphorylated nitroxide. [*Physical Chemistry Chemical Physics*](#), **2014**, DOI: 10.1039/C3CP54816C.
- [4] J. Topin, J. Arnaud, A. Sarkar, A. Audfray, E. Gillon, D. Smith, S. Pérez, H. Jamet, A. Varrot, A. Imberty & A. Thomas. Deciphering the glycan preference of bacterial lectins by glycan array and molecular docking with validation by microcalorimetry and crystallography. *PLOS ONE*, **2013**, [*PLOS ONE*](#) 8(8): e71149. DOI: 10.1371/journal.pone.0071149.
- [3] L. Charlier, J. Topin, C. Ronin, S.K. Kim, W.A. Goddard III, R. Efremov, J. Golebiowski. Molecular features underlying broadly tuned olfactory receptors chemoreception. Human OR1G1 as a test case. [*Cellular and Molecular Life Sciences*](#), **2012**, DOI: 10.1007/s00018-012-1116-0.
- [2] J. Golebiowski, J. Topin, L. Charlier, L. Briand. Interaction between odorants and proteins involved in the perception of smell. The case of Odorant-Binding Proteins probed by molecular modeling and biophysical data. [*Flavour and Fragrance Journal*](#), **2012**, DOI: 10.1002/ffj.3121.
- [1] J. Topin, M. Rousset, S. Antonczak, J. Golebiowski. Kinetics and thermodynamics of gas diffusion in a NiFe hydrogenase. [*Proteins: Structure, Function, and Bioinformatics*](#), **2012**, DOI: 10.1002/prot.23248.

Patent (1, China, Europe, France, Japan, USA)

- [1] X. Meleqi, A. Pegard, J. Topin, Method for characterizing an olfactory stimulation. CN116135139A; EP4183327A1; FR3129072A1; JP2023074481A; US2023148941A1 **2023**.

Book chapter

- [2] L. Charlier, J. Topin, C. De March, P.C. Lai, C. J. Crasto, J. Golebiowski. Chapter 4. Molecular modeling of odorant/olfactory receptor complexes. [*Olfactory receptors: methods and protocols, methods in molecular biology*](#), vol 1003, **2013**, DOI 10.1007/978-1-62703-377-0_4.
- [1] J. Golebiowski, L. Charlier, J. Topin, S. Fiorucci, S. Antonczak. Molecular features underlying the chemoreception of odorant binding proteins and olfactory receptors. [*Flavour Science: Proceedings from XIII Weurman Flavour Research Symposium*](#), **2013**, p 520-524.

French scientific journal

- [4] J. Topin. De la molécule à l'odeur. [*Phytothérapie Européenne*](#), **2023**.
- [3] A. Garcia, J. Topin, R. Pepino, U. Meierhenrich, L. le Sergeant d'Hendecourt, C. Meinert. L'asymétrie de la vie : une origine extraterrestre ? [*L'Actualité Chimique*](#), Décembre **2019**.
- [2] C. Bushdid, J. Topin et J. Golebiowski. Quand l'atmosphère sent le soufre. Perception des pollutions odorantes. [*Pollution atmosphérique*](#), N°234 Avril - Juin **2017**.
- [1] C. Bushdid, C.A. de March, J. Topin, S. Antonczak, M. Bensafi, J. Golebiowski. Ces molécules qui nous mènent par le bout du nez. Le codage moléculaire de la perception des odeurs. [*L'Actualité Chimique*](#), Avril **2016**.

2.5– LIST OF COMMUNICATIONS

Oral communication at international/national conferences

- [12] M. Lalis, S. Fiorucci, J. Topin, M. Hladiš. A Deep Metric Learning Approach for Olfactory Mixture Discriminability. [RECOMB/ISCB Conference on Regulatory & Systems Genomics with DREAM Challenges \(RSGDREAM 2024\)](#), Madison (USA) October **2024**. (invited talk)
 - [11] J. Topin. Combining *in silico* with experimental approaches to decipher the chemical senses. [3rd edition of the Integrative Structural Biology Meeting \(BSI\)](#), Marseille (France) November **2023**. (invited talk)
 - [10] J. Topin. Déciffrer les sens chimiques à l'échelle moléculaire. [Réunion annuelle du GDR O3](#), Dijon (France) October **2023**. (invited talk)
 - [9] J. Topin. Les mécanismes de l'olfaction. [Championnat National du Chien Militaire](#), Suippes (France), October **2022**. (invited talk)
 - [8] M. Hladiš, J. Pacalon, M. Lalis, S. Fiorucci, J. Topin. Odorant binding and receptor activation deciphered at the molecular level. [European Chemoreception Research Organisation \(ECRO\)](#), Berlin (Germany), September **2022**. (invited talk)
 - [7] J. Topin. De la molécule à l'odeur. [Cynogend](#), Archamp (France), November **2021**. (invited talk)
 - [6] J. Topin. Dessine-moi une odeur. [Congrès Goût et Arômes](#), Paris (France), September **2021**. (invited talk)
 - [5] J. Topin. La relation stimulus-percept vue sous l'angle de la chimie. [Réunion annuelle du GDR O3](#), Toulouse (France), November **2019**. (invited talk)
 - [4] J. Topin. Anatomie de l'olfaction. [Congrès du Collège des Centres de Formation Universitaire en Orthophonie \(CCFUO\)](#). Nice (France), April **2019**. (invited talk)
 - [3] J. Golebiowski, C.A. de March, G. Novikov, J. Topin, M. Ma, H. Matsunami. A computational microscope focused on the sense of smell. [Centre Européen de Calcul Atomique et Moléculaire](#). Juelich (Germany), September **2015**. (invited talk)
 - [2] J. Golebiowski, L. Charlier, J. Topin, S. Fiorucci, S. Antonczak. Modeling ligand-receptor interactions at the molecular level. Molecular dynamics used as a computational microscope to decipher the sense of smell. [1er Congrès International Informatique et Sciences de l'Ingénieur : ISI](#). Meknès (Marocco), June **2011**. (invited talk)
 - [1] J. Topin, J. Golebiowski, S. Antonczak. Influence of protein structure on gas diffusion probed by Molecular Dynamic simulations. [European Materials Research Society \(E-MRS\) Spring Meeting](#). Nice (France), May **2011**. (invited talk)
-
- [32] J. Topin, M. Lalis, N. C. Jones, S. V. Hoffmann, L. Briand, C. Meinert. Spectroscopic Investigation of Protein–Odorant Complexes Toward the Development of a Bioelectronic Nose. [Complex Day](#), Nice (France), February **2026**.
 - [31] M. Lalis, M. Hladiš, S. Fiorucci, J. Topin. Code2Smell: modeling olfactory perception through a bio-inspired model, . [12e journées de la Société Française de Chemoinformatique](#), Paris (France), December **2025**.
 - [30] M. Lalis, M. Hladiš, J. Topin. Code2smell: a bio-inspired model of combinatorial code for predicting olfactory perception. [European Chemoreception Research Organisation \(ECRO\)](#), Bilbao (Spain), September **2025**.
 - [29] J. Topin, M. Hladiš, M. Lalis, S. A. Khalil, S. Fiorucci, How chirality affects the sense of smell ? A study by a combination of AI and sensory analysis. [Chirality](#), Kyoto (Japan) **2024**.
 - [28] J. Topin, J. Pacalon, M. Hladiš, C. Belloir, L. Briand, S. Fiorucci, Olfaction under the lens of a computational microscope. [International Symposium of Olfaction and Taste \(ISOT\)](#), Reykjavik (Iceland), June **2024**.

Overview of Research Topics

- [27] M. Lalis, M. Hladiš, S. A. Khalil, L. Briand, S. Fiorucci, J. Topin, Encoding Odor: Bridging Smell and Combinatorial code using M2OR database. [*International Symposium of Olfaction and Taste \(ISOT\)*](#), Reykjavik (Iceland), June **2024**.
- [26] J. Topin, J. Bocková, A. Garcia, U J. Meierhenrich, N. C. Jones, S. V. Hoffmann, C. Meinert, Combining *ab initio* Calculations and CD Spectroscopy Revealed Fundamental Properties of Molecules in Gas Phase. [*19th International Conference on Chiroptical Spectroscopy \(CD 2023\)*](#), Hiroshima (Japan), August **2023**.
- [25] J. Topin, J. Bocková, A. Garcia, U J. Meierhenrich, N. C. Jones, S. V. Hoffmann, C. Meinert, Investigating the building blocks of life by *ab initio* calculations and circular dichroism spectroscopy. [*International Society for the Study of the Origin of Life*](#), Quito (Equador), July **2023**.
- [24] M. Hladiš, J. Pacalon, M. Lalis, S. Fiorucci, J. Topin, Numerical models decipher odorant binding and receptor activation at the molecular level. [*Complex Day*](#), Nice (France), February **2023**.
- [23] J. Topin, N. C. Jones, S. V. Hoffmann, L. Briand, C. Meinert, Tracking ligand-binding effect on protein stability by CD spectroscopy. [*18th International Conference on Chiroptical Spectroscopy \(CD 2022\)*](#), New-York (USA), July **2022**.
- [22] J. Topin, A. Garcia, J. Bocková, N. C. Jones, S. V. Hoffmann, U. J. Meierhenrich, C. Meinert, CD spectroscopy and *ab initio* calculations revealed hidden conformation of amino-acids. [*Chirality*](#), Chicago (USA), July **2022**.
- [21] J. Topin, R. David, Olfactory stimulation for the treatment of apathy in neurodegenerative disorders, [*Annual Meeting of the NeuroMod Institute*](#), Antibes (France), June **2022**.
- [20] J. Topin, J. Golebiowski, Numerical modeling of odorant receptors activation dynamics. [*Pacific Chem*](#), Honolulu (USA), December **2021**.
- [19] M. Hladiš, S. Fiorucci, J. Golebiowski, J. Topin, Olfactory receptor ligand prediction via graph neural networks and representation learning. [*SophiA summit*](#), Sophia-Antipolis (France), October **2021**.
- [18] M. Hladiš, S. Fiorucci, J. Golebiowski, J. Topin, Representation learning to overcome scarce data in machine learning. Application to chemosensory receptors. [*22e congrès du Groupe Graphisme et Modélisation moléculaire \(GGMM\)*](#), Lille (France), September **2021**.
- [17] J. Pacalon, G Audic, C. Moreau, J. Topin, J. Golebiowski, Binding site identification of the highly conserved insect odorant co-receptor (ORco), [*Réunion annuelle du GDR O3*](#), France, November **2021**.
- [16] J. Topin, C.A. de March, J. Golebiowski, Numerical modeling of odorant receptors activation dynamics. Numerical modeling of odorant receptors activation dynamics. [*International Symposium of Olfactory and Taste \(ISOT\)*](#), Portland (USA), August **2020**.
- [15] J. Topin, C.A. de March, J. Golebiowski, Numerical modeling of odorant receptors activation dynamics. [*1st French congress on Integrative Structural Biology \(BSI-2019\)*](#), Toulouse (France), October **2019**.
- [14] J. Topin, C.A. de March, J. Golebiowski. Dynamique d'activation des récepteurs olfactifs. [*26ème Journée de la Chimie PACA*](#). Nice (France), April **2019**.
- [13] J. Topin, J. Golebiowski. Assessment of olfactory habituation/recovery. [*48th International Symposium on Essential Oils \(ISEO2017\)*](#). Pécs (Hungary), September **2017**.
- [12] J. Topin, J. Golebiowski. Smell and relax! [*7ème journée arômes & parfums*](#). Nice (France), June **2017**.
- [11] J. Topin, C.A. de March, J. Golebiowski. *in vitro* / *in silico* approaches reveal G Protein-Coupled Odorant Receptors activation dynamics. [*Groupe de graphisme et modélisation moléculaire 2017 \(GGMM\)*](#). Reims (France), May **2017**.
- [10] J. Topin, J. Breteau, C.A. de March, J. Golebiowski. Smell and relax! Assessment of olfactory stimulation protocols in the framework of psychophysiology measurements. [*47th International Symposium on Essential Oils \(ISEO2016\)*](#). Nice (France), September **2016**.
- [9] J. Topin, D. Kokh, R. Wade, M. Bianciotto. When is an alpha-helix less stable than a loop? An exploration of the conformational dynamics of the HSP90 N-terminal domain. [*7e journées de la Société Française de Chemoinformatique*](#), Nice (France), September **2015**.

Overview of Research Topics

- [8] [J. Topin](#), J. Arnaud, A. Sarkar, A. Audfray, E. Gillon, D. Smith, S. Pérez, H. Jamet, A. Varrot, A. Imberty, A. Thomas. Mixing experimental and theoretical approaches to decipher Glycan-Lectin interactions. [Second workshop on Computational Structural Biology: Integrative Approaches for Modeling Biomolecular Complexes](#), Nice (France), June **2013**.
- [7] [J. Topin](#), J. Arnaud, A. Sarkar, A. Audfray, E. Gillon, S. Pérez, H. Jamet, A. Imberty, A. Thomas. Study of Lectine-Saccharides interaction by means of molecular modeling (Étude des interactions Lectine saccharides par modélisation moléculaire). [XVIIIe congrès du Groupe de Graphisme et Modélisation Moléculaire](#), Saint-Pierre d'Oléron (France), May **2013**.
- [6] [J. Topin](#), J. Golebiowski, S. Antonczak. Describing ligand-receptor interactions by means of equilibrium simulations (Mise en œuvre de simulations brute force pour décrire les interactions ligand-récepteur), [13e Rencontre des Chimistes Théoriciens Francophones](#), Marseille (France), July **2012**.
- [5] [J. Topin](#), J. Golebiowski, S. Antonczak. Structural, kinetic and thermodynamics features of O₂ inhibition of a hydrogenase (Aspects structuraux, cinétiques et thermodynamiques de l'inhibition d'une hydrogénase par O₂). [GDR-Probe](#). Carry-Le-Rouet (France), December **2011**.
- [4] [J. Topin](#), J. Golebiowski, S. Antonczak. Energy study of diffusion mechanisms in a hydrogenase from *Desulfovibrio fructosovorans* and several mutants (Etude énergétique des phénomènes de diffusion au sein de l'hydrogénase de *Desulfovibrio fructosovorans* et de différents mutants). [GDR-BioH2](#), Carry-Le-Rouet (France), November **2010**.
- [3] [J. Topin](#), J. Golebiowski, S. Antonczak. Theoretical description of gas diffusion in *Desulfovibrio fructosovorans* hydrogenase: a study of the influence of protein structure. [H2ase 2010](#), Uppsala (Sweden), June-July **2010**.
- [2] [J. Topin](#), J. Golebiowski, S. Antonczak. Theoretical study of diffusion channel inside a hydrogenase from *Desulfovibrio fructosovorans* (Etude théorique des canaux de diffusion au sein de l'hydrogénase de *Desulfovibrio fructosovorans*). [GDR-BioH2](#), Lalonde les Maures (France), October **2009**.
- [1] [J. Topin](#), J. Golebiowski, S. Antonczak. Molecular modeling study of O₂ diffusion in a NiFe hydrogenase (Etude par modélisation moléculaire de la diffusion de O₂ au sein d'une hydrogénase à Ni-Fe). [Réunion du club Métalloprotéines et Modèles](#), Fréjus (France), September **2009**.

Academic and Research Institutional Seminars

- [11] The Molecular Language of Smell: Reading Odors Through Receptors, [Daegu Gyeongbuk Institute of Science and Technology](#), Daegu, Republic of Korea, March **2026**.
- [10] Polymorphism in olfaction and how to evaluate point mutation using molecular modeling methods, [Hanyang University](#), Seoul, Republic of Korea, March **2026**.
- [9] Deorphanizing the Human Olfactory Receptor Repertoire: Lessons from M2OR, [Seoul National University](#), Seoul, Republic of Korea, January **2026**.
- [8] Deorphanizing the Human Olfactory Receptor Repertoire: Lessons from M2OR, [Institute of Physics University of Freiburg](#), Germany, November **2025**.
- [7] Deciphering the Gating and Binding Events at the Origin of the Chemical Senses, [Cold Spring Harbor Laboratory](#), USA, July **2025**.
- [6] Utilisation de l'Intelligence Artificielle dans le secteur Arômes & Parfums, [Société Française des Parfumeurs](#), France, April **2025**.
- [5] From molecules to smell. Université Côte d'Azur, [Winter School Adipocible](#), France, December **2024**.
- [4] Olfaction under the lens of a computer microscope. [Hanyang International Mini Symposium](#), Hanyang University Seoul, Republic of Korea, May **2024**.
- [3] À la découverte des sens chimiques ! [Centre de Recherche Paul Pascal](#), CNRS UMR 5031, 33600 Pessac, June **2021**.
- [2] Journée des jeunes doctorants EDSFA. [Université Nice Sophia Antipolis](#), France, November **2019**.
- [1] Numerical approaches for olfaction. [Workshop MODELIFE IPMC](#), Université Nice Sophia Antipolis, France, December **2019**.

Overview of Research Topics

Poster Presentations

- [40] J. Topin, M. Lalis, C. Belloir, L. Briand, C. Meinert. Stable by design: what ligand binding reveals about the function of odorant-binding proteins in mammals? [Association of Chemoreception Sciences](#), St Pete (USA), April **2026**.
- [39] M. Hladiš, M. Lalis, M. Bronstein, J. Topin. Mimicking dose-response experiments to predict olfactory receptor activity and potency. [Association of Chemoreception Sciences](#), St Pete (USA), April **2026**.
- [38] J. Topin, M. Lalis, N. C. Jones, S. V. Hoffmann, L. Briand, C. Meinert. Toward a Bioelectronic Nose: Integrative Approaches to Protein-Based Odorant Detection and Chiral Discrimination. [Chirality](#), New-York (USA), **2025**.
- [37] J. Topin, J. Bocková, N. C. Jones, S. V. Hoffmann, C. Meinert. Investigating the Role of Polarized Light in Life's Precursors Using Ab Initio Calculations and Circular Dichroism Spectroscopy. [The second Biennial European Astrobiology Conference \(BEACON\)](#), Reykjavik (Iceland) June **2025**.
- [36] M. Lalis, M. Hladiš, L. Briand, S. Fiorucci, J. Topin. An Overview Of Human Odorant Receptors And Their Identified Agonists. [Association of Chemoreception Sciences](#), Bonita spring (USA), April **2025**.
- [35] C. Cornut, C. Belloir, M. Lalis, A. Thomas, R. Menin, J. Topin, L. Briand. Molecular Interactions Between Glutathione And Umami Taste Receptor Tas1R1/Tas1R3. [Association of Chemoreception Sciences](#), Bonita spring (USA), April **2025**.
- [34] L. Maxence, M. Hladiš, J. Pacalon, S. Fiorucci, J. Topin. Are Olfactory Receptors more than class A GPCRs? [Gordon Research Conference on Ligand Recognition and Molecular Gating](#), Ventura (USA), March **2024**.
- [33] M. Hladiš, J. Pacalon, M. Lalis, S. Fiorucci, J. Topin. Deciphering Chemical Senses at the Molecular Scale. [Gordon Research Conference on Ligand Recognition and Molecular Gating](#), Ventura (USA), March **2024**.
- [32] T. Shim, J. Pacalon, K. WonCheol, X. Cong, J. Topin, J. Golebiowski, C. Moon. The Functional Role of the Third Extracellular Loop of Mammalian Odorant Receptors in Ligand Binding. [Association of Chemoreception Sciences, Bonita spring \(USA\)](#), April **2023**.
- [31] X. Meleqi, U. Meierhenrich, A. Pegard, J. Topin. Combining physiological measurements and virtual reality to decipher emotions evoked by scents. [Neuroscience 2022](#), San Diego (USA), November **2022**.
- [30] X. Meleqi, U. Meierhenrich, A. Pegard, J. Topin. Combining virtual reality and physiological measurement to assess emotions triggered by fragrances. [32nd IFSCC Congress](#), London (UK), September **2022**.
- [29] X. Meleqi, U. Meierhenrich, A. Pegard, J. Topin. Recognizing emotions triggered by fragrances using virtual reality and physiological measurement. [European Chemoreception Research Organisation \(ECRO\)](#), Berlin (Germany), September **2022**.
- [28] L. Moitrier, C. Belloir, M. Lalis, Y. Hou, J. Topin, L. Briand. Deciphering the ligand binding properties of the mouse odorant-binding protein OBP5 from *Mus musculus*. [European Chemoreception Research Organisation \(ECRO\)](#), Berlin (Germany), September **2022**.
- [27] M. Hladiš, M. Lalis, S. Fiorucci, J. Topin. Predicting the combinatorial code for any odorant molecule: a graph neural networks approach. [European Chemoreception Research Organisation \(ECRO\)](#), Berlin (Germany), September **2022**.
- [26] J. Pacalon, C. Belloir, S. Fiorucci, L. Briand, J. Topin. TAARs response to amines are largely affected by sequence variants. [European Chemoreception Research Organisation \(ECRO\)](#), Berlin (Germany), September **2022**.
- [25] M. Lalis, M. Hladiš, Y. Hou, N. Baldovini, J. Topin. ChemOddity: Machine Learning to predict odorant detection threshold. [European Chemoreception Research Organisation \(ECRO\)](#), Berlin (Germany), September **2022**.
- [24] J. Topin, J. Pacalon, G. Audic, J. Golebiowski, C. Moreau. Identification of entry pathway and binding site of insect odorant co-receptor. [Association of Chemoreception Sciences](#), Bonita spring (USA), April **2022**.
- [23] M. Hladiš, S. Fiorucci, J. Topin. Predicting the combinatorial code of olfaction via graph neural networks and representation learning. [Association of Chemoreception Sciences](#), Bonita spring (USA),

Overview of Research Topics

April **2022**.

- [22] S. Fiorucci, C. Bouysset, Y. Kim, M. Rhyu, J. Golebiowski, J. Topin. Integrative structural modelling reveals functional molecular switches of human G protein-coupled bitter-taste receptors. [Association of Chemoreception Sciences](#), Bonita spring (USA), April **2022**.
- [21] M. Hladiš, J. Pacalon, M. Lalys, C. Bouysset, S. Fiorucci, J. Topin. Deciphering Chemical Senses at the Molecular Scale. [Gordon Research Conference on Ligand Recognition and Molecular Gating](#), Lucca (Italy), March **2022**.
- [20] J. Topin, S. Kuhaudomlarp, A. Varrot, A. Imberty, With or without sugar? A search for lectin inhibitors. [22e congrès du Groupe Graphisme et Modélisation moléculaire \(GGMM\)](#), Lille (France), October **2021**.
- [19] J. Pacalon, G Audic, C. Moreau, J. Topin, J. Golebiowski. Binding site identification of the highly conserved insect odorant co-receptor (ORco). [22e congrès du Groupe Graphisme et Modélisation moléculaire \(GGMM\)](#), Lille (France), October **2021**.
- [18] C. Bouysset, J. Topin, J. Pacalon, Y. Kim, M. Rhyu, S. Fiorucci, J. Golebiowski. Functional molecular switches of mammalian G protein-coupled bitter-taste receptors. [European Chemoreception Research Organisation \(ECRO\)](#), Cascais (Portugal), September **2021**.
- [17] J. Pacalon, G Audic, C. Moreau, J. Topin, J. Golebiowski. Binding site identification of the highly conserved insect odorant co-receptor (ORco). [European Chemoreception Research Organisation \(ECRO\)](#), Cascais (Portugal), September **2021**.
- [16] M. Hladiš, S. Fiorucci, J. Golebiowski, J. Topin. Odor prediction via graph neural networks and representation learning. [European Chemoreception Research Organisation \(ECRO\)](#), Cascais (Portugal), September **2021**.
- [15] J. Elbaum, J. Golebiowski, J. Sabatier, J. Topin, R. David. Perte de l'odorat et perte de l'autonomie : quels liens chez la personne âgée ? [Congrès Français de Psychiatrie, 11ème édition](#), Nice (France), December **2019**.
- [14] J. Topin, J. Golebiowski, R. David. Evaluation psychophysiologique et prédiction de la relaxation par les odeurs. [Congrès Français de Psychiatrie, 11ème édition](#), Nice (France), December **2019**.
- [13] J. Topin, C. A. de March, J. Golebiowski. Combining modeling approaches and *in vitro* measurements to decipher activation mechanism of olfactory receptors. [European Chemoreception Research Organisation \(ECRO\)](#), Trieste (Italy), September **2019**.
- [12] J. Pacalon, J. Topin, J. Golebiowski. Phylogénie structurale des récepteurs couplés aux protéines G. Focus sur les récepteurs aux odorants. [26e Journée de la Chimie PACA](#), Nice (France), April **2019**.
- [11] J. Pacalon, J. Topin, J. Golebiowski. Phylogénie structurale des récepteurs couplés aux protéines G. Focus sur les récepteurs aux odorants. [21e congrès du Groupe Graphisme et Modélisation moléculaire \(GGMM\)](#), Nice (France), April **2019**.
- [10] J. Topin, C. Bushdid, J. Golebiowski. Psychophysiological effects upon smelling essential oils and their compounds. [European Chemoreception Research Organisation \(ECRO\)](#), Würzburg (Germany), September **2018**.
- [9] C Bushdid, J Topin, A Claire, C Belloir, L Briand, J Golebiowski. Towards computational biomimetic nose. [17th International Symposium on Olfactory and Taste \(ISOT 2016\)](#), Yokohama (Japan), June **2016**.
- [8] J. Topin, C.A. de March, J. Golebiowski. Assessment of olfactory stimulation protocols in the framework of psychophysiology measurements. [17th International Symposium on Olfactory and Taste \(ISOT 2016\)](#), Yokohama (Japan), June **2016**.
- [7] J. Topin, H. Jamet, A. Thomas, A. Imberty. Combining theoretical and experimental methods to design inhibitors for *Pseudomonas aeruginosa* lectins LecA and LecB. [Biopolymers Gordon Research Conference](#), Newport (USA), June **2014**.
- [6] J. Topin, J. Golebiowski, S. Antonczak. Aspects structuraux, cinétiques et thermodynamiques de la perméation de O₂ dans une hydrogénase à NiFe. [13e Rencontre des Chimistes Théoriciens Francophones](#), Marseille (France), July **2012**.
- [5] J. Topin, J. Golebiowski, S. Antonczak. Description théorique de la diffusion de gaz au sein de l'hydrogénase de *Desulfovibrio Fructusovorans* : une étude sur l'influence de la structure protéique.

Overview of Research Topics

- [*12e Rencontre des Chimistes Théoriciens Francophones*](#), Namur (Belgium), July **2010**.
- [4] J. Topin, J. Golebiowski, S. Antonczak. Influence de la structure protéique sur la diffusion de gaz au sein d'une enzyme de type hydrogénase. [*4e Colloque de la Chimie Azurée*](#), Nice (France), October **2009**.
- [3] J. Topin, J. Golebiowski, S. Antonczak. Influence de la structure protéique sur la diffusion de gaz au sein d'une enzyme de type hydrogénase. [*17e congrès du Groupe Graphisme et Modélisation moléculaire \(GGMM\)*](#), Mittelwihr (France), May **2009**.
- [2] J. Topin, J. Golebiowski, S. Antonczak. Influence de la structure protéique sur la diffusion de gaz au sein d'une enzyme de type hydrogénase. [*Journée de la chimie 2009 \(SCF PACA\)*](#), Marseille (France), April **2009**.
- [1] J. Topin, J. Golebiowski, S. Antonczak. Influence of the protein structure upon gas diffusion in hydrogenase type enzymes. [*Workshop Flexibility and Biological Recognition: from Biophysics to Data Models*](#), Sophia Antipolis (France), March **2009**.

Public Outreach

- [12] [*Colloque UNESCO Les savoir-faire liés au parfum en Pays de Grasse face aux défis du XXIe siècle*](#): Les parfums de demain : big data et intelligence artificielle, Musée International de la Parfumerie Grasse, November **2025**.
- [11] [*Pint of Science*](#): À la découverte des sens chimiques ! Beer District, Nice, May **2024**.
- [10] [*Neuroplanète*](#): Les recettes d'un repas qui excite tous nos sens. Centre Universitaire Méditerranéen, Nice, April, **2024**. <https://www.youtube.com/watch?v=XBIClOzIRak>
- [9] [*Café art/science regards scientifiques*](#): la transmission de savoir-faire olfactifs. Le Hublot Nice, December **2022**.
- [8] [*Forum scientifique - Café rencontre*](#): L'IA au service de la création parfum, mythe ou réalité ? Musée International de la Parfumerie Grasse, October **2022**.
- [7] [*42eme forum territorial culture science PACA*](#), Musée International de la Parfumerie Grasse, October **2021**.
- [6] [*Cité des sens*](#), Cité des sciences Paris, September **2021**.
- [5] [*Inauguration of the Neuromod institute*](#) (Université Côte d'Azur), Nice, April **2021**. <https://www.youtube.com/watch?v=py6TqbjAyll&t=987s>
- [4] [*Cité des sens*](#), Cité des sciences Paris, September **2020**.
- [3] [*Semaine du cerveau opening conference*](#), Nice, March **2020**.
- [2] [*Semaine du cerveau opening conference*](#), Nice, March **2019**.
- [1] [*Semaine du cerveau*](#), Musée International de la Parfumerie Grasse, April **2017,2018**.

Medias

- [3] [*Podcast \(Fr\)*](#): France Culture, Avec Sciences **2024**. [lien](#)
- [2] [*Podcast \(Fr\)*](#): RTS, CQFD **2023**.
- [1] [*Video \(Fr\)*](#): La science des odeurs et du parfum – CNRS Le journal **2017**. [lien](#)

3. DISSEMINATION OF (HUMANIST) CULTURE

Knowledge dissemination to high schools: APERLA Project: Discovering the Chemical Senses

I have developed a public outreach project aimed at raising awareness of the chemical senses : smell and taste. I lead this initiative in partnership with the *Musée International de la Parfumerie* de Grasse, as well as with industrial partners from the company *Expressions Parfumées*. The project is funded by the regional council (APERLA 2021 and 2023, 3,000 € per year). The combination of academic and industrial stakeholders provides an opportunity to engage the general public with themes related to olfaction and taste, across the fields of scientific research, the arts, and industry. The project is designed to promote the dissemination of scientific and industrial knowledge through engaging activities that spark curiosity and transmit knowledge to secondary school students.

The various sessions offer an overview of several topics, including: the biological mechanisms underlying the chemical senses, chemical communication between humans, and the development of next-generation ingredients. These activities illustrate the multidisciplinary approach required to understand the complex pathways from molecule to perception, and from perception to emotion. The sessions aim to highlight the knowledge needed to understand the chemical senses, placing particular emphasis on multidisciplinary and the diversity of expertise.

The involvement of industrial actors also helps raise student awareness of careers in the flavor & fragrance sector. This project also offers doctoral students in my team the opportunity to engage in science outreach and educational activities—for example, through a practical workshop in high schools entitled “*La Fraise sans Fraise!* ”, where students reconstruct an aroma from individual components.

From Molecule to smell: study of mammalian olfaction

Quand, les deux yeux fermés, en un soir chaud d'automne,

Je respire l'odeur de ton sein chaleureux,

Je vois se dérouler des rivages heureux

Qu'éblouissent les feux d'un soleil monotone.

(Charles Baudelaire, *Parfums exotiques*)



In poetry, olfaction is used for its powerful capacity to evoke memories, emotions and sensuality. Poets describe fragrance as an invisible thread linking the world to our most intimate feelings. Charles Baudelaire gave smell a privileged place in his poetic universe, making it not only a source of sensual pleasure but also a gateway to deeper metaphysical truths. In the poem *Correspondances*, he explores the idea of synesthesia, where the senses blend and the material world reflects spiritual realities. He writes: *"Les parfums, les couleurs et les sons se répondent"*, asserting that scents—like sounds and colors—form part of a symbolic language that transcends rational perception. In *Parfums Exotiques*, the speaker, immersed in the scent of his lover, is transported to a distant, imagined paradise: *"Quand, les deux yeux fermés, en un soir chaud d'automne, / Je respire l'odeur de ton sein chaleureux"*. Scent here acts as a vehicle for escape and reverie, evoking an exotic and timeless elsewhere. Arthur Rimbaud, building on Baudelaire's legacy, also uses olfaction to convey the fragility of life and the finality of death. In *Le Dormeur du Val*, he writes with stark poignancy: *"Les parfums ne font pas frissonner sa narine"*, using the absence of olfactory sensation to underline the irreversible stillness of a young soldier's corpse. From Baudelaire's structured sensuality to Rimbaud's haunting silences, olfaction in poetry becomes a language of the ineffable, capable of evoking inner states where reason gives way to instinct, memory, and the sublime.

Abstract

Our sense of smell plays a crucial role in our social interactions and emotional experiences. This perception relies on the interaction between odorant molecules and a class of proteins called olfactory receptors (ORs). These receptors, expressed on the surface of olfactory sensory neurons, decode a vast array of distinct odors through a sophisticated activation pattern. It appears that odors may be, at least partially, encoded within the chemical properties of the perceived compounds.

Introduction

Olfaction is the sense dedicated to analyzing our volatile chemical environment. It continuously monitors the air we inhale, allowing us to identify the molecules present around us. Olfactory information provides vital cues, such as identifying a food source, detecting danger, or recognizing a potential mate. In animals, these stimuli often trigger automatic avoidance or attraction responses. In humans, however, olfactory stimulation primarily influences the emotional sphere, affecting thoughts, mood, and even food preferences. This relationship between the molecular world and our behaviors is highly personal and varies greatly between individuals. Establishing a reliable correlation between the structure of an odorant molecule and its perceived smell remains a major challenge for the scientific community, as it likely requires interdisciplinary approaches combining chemistry, biology, behavioral sciences, and humanities.

Olfaction in History

For a long time, olfaction was an understudied and misunderstood sense, largely due to its association with animal instincts. In his classification of the senses, Emmanuel Kant placed olfaction at the very bottom, even advising against training this sense, which he saw as too closely linked to animal nature. According to him, smelling involved perceiving particles that altered the original purity of the air. It is interesting to note that in his discussion on smell and taste, Kant highlighted the subjective nature of these senses and their reliance on the chemical reception of fine particles.

The rehabilitation of olfaction began with the Sensualist philosophical movement in the 18th century. In his famous *parable of the statue*, Abbé Étienne Bonnot de Condillac drew an analogy between the human soul and a marble statue gradually coming to life through sensory experiences, particularly smells¹. According to his concept, olfaction acts as a stimulator, provoking pleasure or melancholy and transitioning the statue from an inert object to a living being.

In the early 19th century, Dr. Cabanis emphasized the interconnection between the five senses, considering olfaction essential: *"Odors strongly affect the entire nervous system; they predispose it to all sensations of pleasure and impart that slight degree of turmoil that seems inseparable from such experiences."*² Finally, at the turn of the 20th century, Nietzsche famously declared: *"All my genius resides in my nostrils,"*³ thus becoming a spokesperson for the renewed importance of olfaction. From these various philosophical perspectives, we can infer that the sense of smell symbolizes clairvoyance, discernment, and intelligence, enabling foresight and anticipation.

¹ Traité des sensations, 1754.

² Rapports du physique et du moral de l'homme, 1805, tome 1.

³ Ecce homo : Wie man wird, was man ist, 1908.

Scientific Research

For a long time, scientists neglected olfaction. Only recently has research been conducted to decipher this sense and understand its influence on memory, language, and behavior. Researchers have demonstrated that odors trigger sensations with a high likelihood of being remembered due to their strong connection with the amygdala, a brain region associated with emotions¹. Since olfactory sensations bypass the neocortex, they are more challenging to describe because they exist prior to language. Unlike physical senses such as hearing, vision, and touch—where stimuli are characterized by measurable physical properties—an odor is the result of the olfactory system's interpretation of an odorant's physicochemical properties. Consequently, most odors cannot be directly measured, and the olfactory properties of an odorant must be characterized by human subjects².

Major breakthroughs in understanding olfaction began in 1985 with the discovery of olfactory receptors (ORs) in vertebrates. This progress continued in 1989 with the identification of G-protein involvement in olfaction mechanisms. In 1991, Buck and Axel identified the genes encoding olfactory receptors³, a discovery that earned them the Nobel Prize in Medicine in 2004. The total number of OR genes has been estimated at approximately 400 in humans and can reach nearly 2,000 in elephants, making ORs the largest gene family!

Odorant Molecules

Odorant molecules exhibit great diversity in function, conformation, and properties. They can be classified based on their physicochemical properties, their ability to disperse in the air, and their solubility in water. Odorant compounds may be:

- Volatile but insoluble in water
- Non-volatile and insoluble in water
- Volatile and soluble in water
- Non-volatile but soluble in water

Recent research by Mayhew and colleagues has attempted to define the chemical space of odorant molecules using machine learning techniques⁴. Their models, trained on previously identified odorant molecules, suggest that transport in the environment is the most crucial characteristic a molecule must possess to trigger an olfactory perception. By analyzing a database of 166 billion chemical compounds, their models predicted that at least 40 billion potential odorant compounds exist.

Furthermore, odorant molecules can also be classified based on their functional roles:

- Environmental odorants: these include smells associated with food, water, nesting sites, or dangers such as fire.
- Allelochemicals: these are odors emitted by other species or organisms, conveying information about prey presence, nest location, symbiotic relationships, territorial marking, predator avoidance, metamorphosis, growth, or pollination.
- Pheromones: these are odors emitted by individuals of the same species to communicate information related to sexual attraction, individual identification, or social group belonging.

In many cases, the scents responsible for olfactory perception are composed of complex mixtures of volatile molecules. Some of these substances are detectable by the human nose even in extremely small amounts. For example, the *wine lactone*, can be perceived at trace levels ($\sim 10^{-14}$ g·L⁻¹). This

molecule has a sweet, woody aroma reminiscent of coconut and plays a role in the fragrance of Gewurztraminer wines and the aromatic profiles of orange and grapefruit juices⁵. Its olfactory potency is so strong that dissolving just 20 mg in an Olympic-sized swimming pool would be enough to detect its scent!

Finally, the perception of an odor can vary depending on its concentration in a mixture. For example, indole exhibits a floral note at very low concentrations but takes on a fecal-like odor at higher concentrations!

Structure-Odor Relationships

In the fragrance and flavor industry, discovering new compounds is largely driven by serendipity. Establishing a relationship between a molecule's structure and its chemosensory properties has been a longstanding challenge for chemists. Despite extensive research, predicting the smell of a novel molecule based solely on its structure remains a formidable task.

Efforts to correlate chemical structure with odor have primarily focused on specific fragrance families of industrial interest, such as woody, green, floral, or musky scents. Although no absolute rules exist, certain chemical functional groups seem to be associated with specific olfactory characteristics.

In general, the extreme complexity of structure-odor relationships makes the rational design of odorants challenging. Even slight structural modifications—such as shifting a double bond, altering chirality, or adding a single methyl group—can lead to significant changes in odor quality and intensity or even complete loss of aroma (Figure 1).

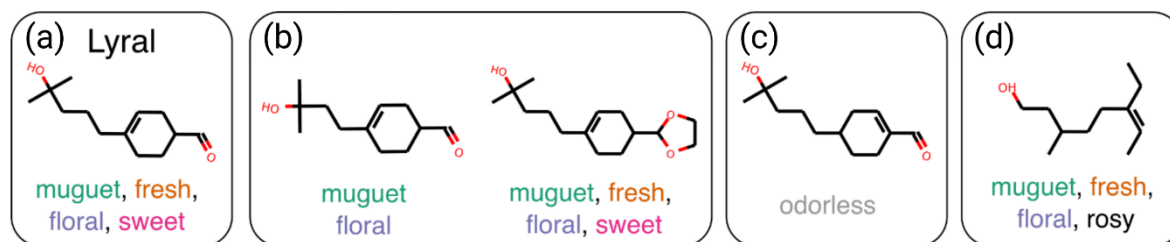


Figure 1. Molecules with similar structures do not necessarily have similar odors. (a) Lyral. (b) Molecules with similar structures can have similar descriptors. (c) However, some molecules with minimal variations have different semantic descriptors. (d) Conversely, very different molecules can share multiple semantic descriptors. Adapted from⁶.

The Diversity of Odors

Odorant molecules span a chemically complex and vast space. The number of potential odorants is estimated to be in the billions, as all volatile molecules have the potential to be odorants⁴. New odorants are regularly discovered and utilized by the perfume and cosmetics industries.

About a decade ago, it was proposed that the human sense of smell could discriminate more than a trillion odors⁷. This astonishing finding, labelled as a "major breakthrough of 2014", sparked significant controversy within the scientific community⁸. The debates surrounding this figure highlight the great challenge of quantifying the capacity of the human olfactory system. Unravelling the biological mechanisms that convert a molecule into an odor will help provide less contentious answers to these questions of olfactory acuity.

Physiology of olfaction

The olfactory system has several distinctive characteristics that have profoundly influenced its organisation and functioning. First, it is continuously exposed to toxic and oxidising substances present in the ambient air, making it the most vulnerable of the sensory systems. This constant exposure has led to the development of protective mechanisms, including an exceptional ability to regenerate olfactory sensory neurons throughout life.

Olfactory perception is a highly sophisticated process that begins when volatile molecules interact with specialised receptors on the surface of sensory neurons (**Figure 2**). The olfactory system is centred in the nasal cavity, where the olfactory epithelium houses the olfactory sensory neurons (OSN). These neurons deploy cilia in the layer of mucus that protects them and in which odorous substances dissolve. Their axons cross the cribriform plate to the olfactory bulb, which then transmits signals to higher brain structures, including the olfactory cortex, amygdala and hippocampus.

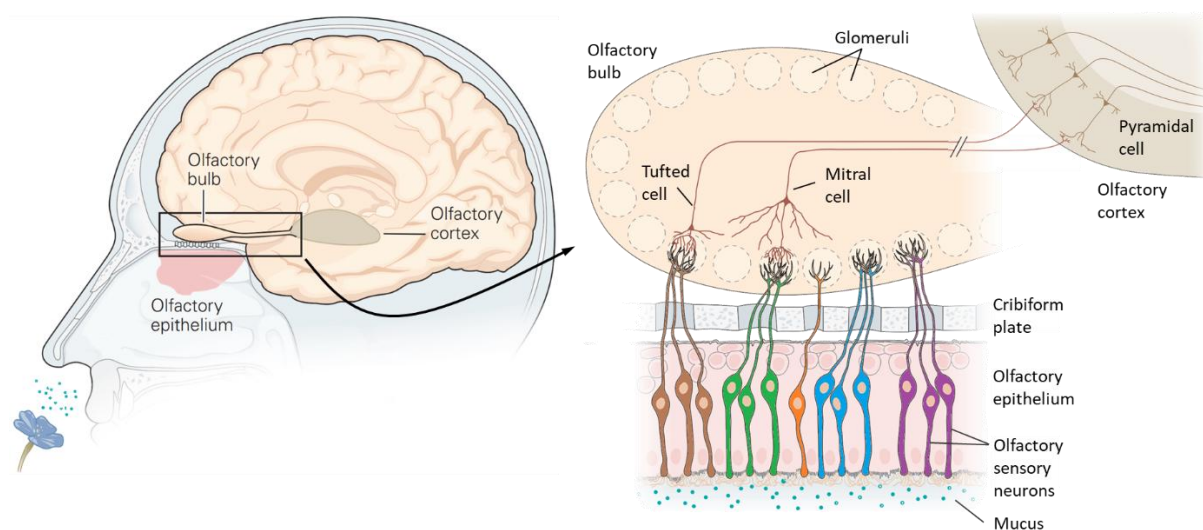


Figure 2. Overview of the Human Olfactory System adapted from ref. 4. Odorant molecules enter the nasal cavity and reach the olfactory epithelium, where they are detected by olfactory sensory neurons expressing specific olfactory receptors. Olfactory sensory neurons are colored according to the olfactory receptor gene they express. Upon activation, these neurons transmit the signal via their axons, which pass through the cribriform plate to converge onto discrete glomeruli within the olfactory bulb. From there, the olfactory information is relayed to higher-order olfactory cortical regions, where it is integrated and processed to ultimately give rise to the perceptual experience of odor.

OSN axons converge and form synapses with projection neurons (mitral and tufted cells) as well as various interneurons, including peri- and juxtaglomerular cells. In human, each OSN expresses only one of approximately 400 odorant receptors (ORs). Importantly, each OSN sends a single axon to one glomerulus, and all OSNs that innervate a given glomerulus express the same OR. As a result, the molecular identity defined by OR expression in OSNs is preserved in the organization of the glomeruli (**Figure 2**).

Mammalian olfactory receptors (ORs) belong to the class A, or rhodopsin-like, G protein-coupled receptor (GPCR) family and constitute the largest GPCR subfamily in vertebrate genomes⁹. They share the canonical GPCR architecture composed of seven transmembrane α -helices connected by extracellular and intracellular loops, forming a compact fold typically encoded by approximately 300

⁴ Doty et al., Handbook on Olfaction and Gustation, 2015

amino acids. Despite their relatively small size, ORs retain a core set of conserved GPCR motifs essential for receptor activation and signal transduction, including the GN motif in transmembrane helix (TM) 1, the DRY motif in TM3, the KA motif in TM6, and the NPxxY motif in TM7, which collectively participate in conformational rearrangements associated with G protein coupling (Figure 3).

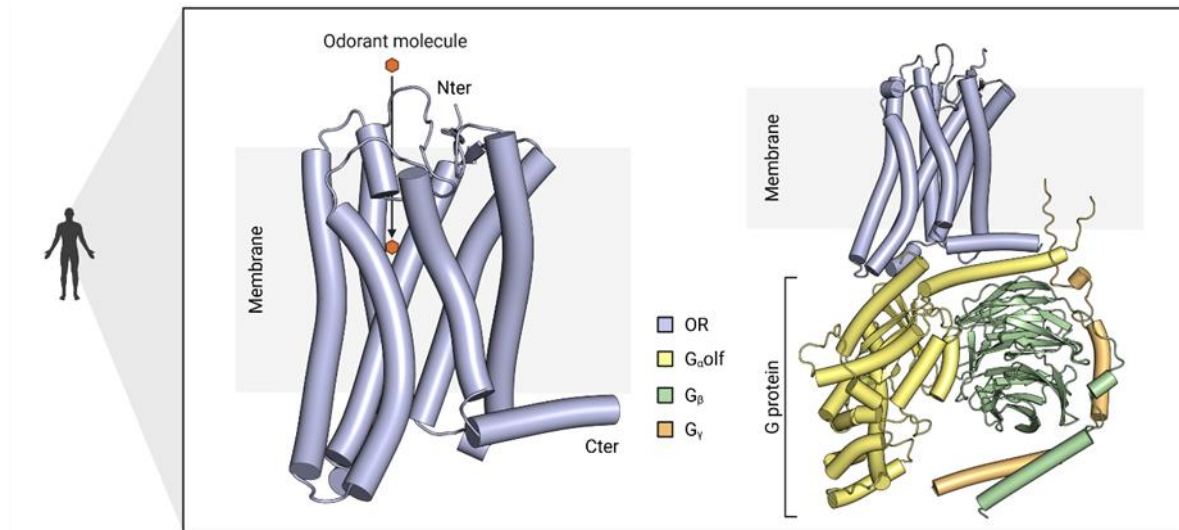


Figure 3. Structure of an Olfactory Receptors (adapted from¹⁰) and in complex with the G protein. ORs are GPCRs consisting of 7 transmembrane helices, with the N-terminus on the extracellular side and the C-terminus on the cytosolic side. A scent molecule binds in a cleft buried in the receptor and, through an allosteric process, allows the G protein complex to bind to the cytosolic side of the receptor. OR in complex with its G protein composed of $G_{\alpha\text{olf}}$, G_{β} and G_{γ} (yellow, green, and orange respectively).

In addition to these canonical motifs, mammalian ORs exhibit a series of OR-specific conserved sequences that distinguish them from other class A GPCRs and are thought to fine-tune odorant recognition and receptor activation. These include the PMYxFL motif in TM2, the extended MAYDRYVAIC sequence in TM3, conserved serine and tyrosine residues in TM5, and characteristic motifs in TM6 such as RxKAxTCxSH and FY, which are likely involved in ligand interaction and receptor dynamics. The presence of both generic GPCR motifs and OR-specific signatures reflects the dual requirement for conserved signal transduction mechanisms and highly diversified ligand recognition.

A key feature of mammalian ORs is the extreme variability of their orthosteric binding cavity, formed primarily by residues from TMs 3, 5, 6, and 7. This structural diversity underlies the remarkable breadth of chemical space detected by the olfactory system, enabling the recognition of thousands of structurally diverse odorants with varying affinities and selectivities. Subtle sequence variations within the binding pocket translate into distinct odorant response profiles, supporting a combinatorial coding strategy in which each odorant activates a unique ensemble of ORs. This balance between structural conservation and binding-site diversity is central to the functional versatility of the vertebrate olfactory receptor family¹¹.

The signal transduction in mammalian olfaction follow a metabotropic mechanism (Figure 4). At the molecular level, ligand binding to OR 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. It results in an opening of a cleft between the intracellular parts of TM3 and TM6¹⁰. The binding of $G_{\alpha\text{olf}}$ to the activated OR promotes the exchange of GDP for GTP in the α subunit. This leads to G protein dissociation. The $G_{\alpha\text{olf}}$ subunit subsequently activates adenylyl cyclase III (ACIII), a key enzyme localized

in the membrane. ACIII catalyzes the conversion of ATP into cyclic AMP (cAMP), resulting in a rapid increase in intracellular cAMP concentration.

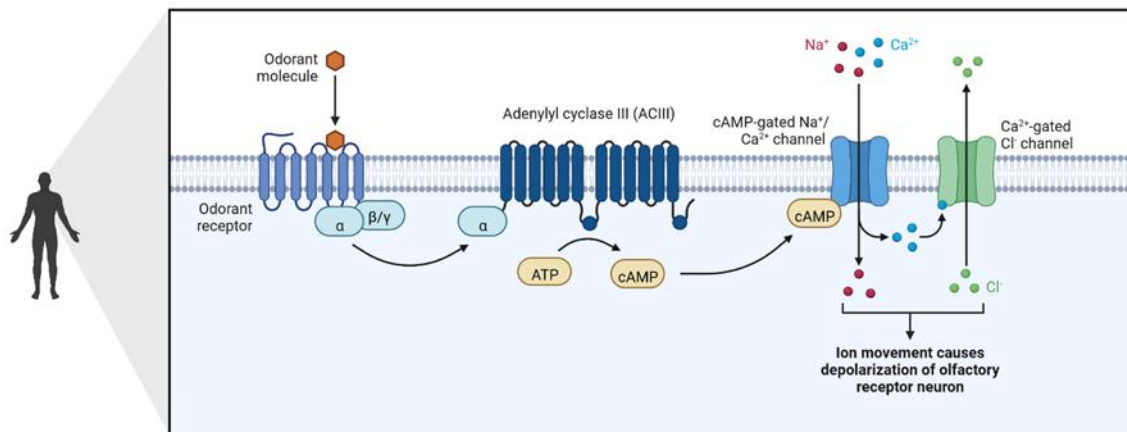


Figure 4. The different steps of olfactory transduction in mammals. Metabotropic transduction in mammals. An odorant activates the OR, in turn activating the G protein. The G protein then activates adenylyl cyclase, initiating the production of the second messenger cAMP. The cAMP opens the Na⁺/Ca²⁺ channels managed by cAMP. The cytosolic increase in Ca²⁺ concentration in turn opens Cl⁻ channels, which further depolarizes the neuron.

Elevated cAMP directly gates cyclic nucleotide-gated (CNG) cation channels, allowing the influx of Na⁺ and Ca²⁺ ions into the neuron. This cation entry produces an initial depolarization of the olfactory sensory neuron. The rise in intracellular Ca²⁺ further amplifies the signal by activating Ca²⁺-gated Cl⁻ channels. Given the unusually high intracellular chloride concentration in olfactory neurons, the opening of these channels leads to Cl⁻ efflux, which further depolarizes the membrane. If this depolarization reaches threshold, it triggers action potential firing, thereby transmitting the olfactory signal along the axon of the sensory neuron toward the olfactory bulb. This transduction cascade ensures high sensitivity and signal amplification, allowing mammals to detect low concentrations of odorants while preserving specificity through receptor-ligand interactions.

These receptors work on a combinatorial principle to encode the identity of the odour. Each odorant can activate several ORs, and each OR can respond to several odorants, forming a combinatorial coding system that enables discrimination of an myriad of distinct odours⁷. This system is based on a pattern recognition method rather than the isolated identification of a specific molecule, as the vast majority of olfactory receptors are not dedicated to the recognition of a particular molecule or chemical function¹². During inhalation, an odorant interacts with a subset of olfactory receptors. Differential activation of this repertoire is then interpreted by the central nervous system as an odour.

Despite significant advances in our understanding of the sense of smell, the fundamental principles underlying odorant-receptor interactions remain largely unknown. One of the greatest challenges in this field is the complex combinatorial nature of olfactory coding. In humans, around 400 functional OR genes encode receptors that collectively recognise and discriminate between a wide range of odour molecules¹². Unlike other sensory systems where the stimulus-response relationship is more linear, olfaction is based on a multiple recognition model: a single odorant can activate several receptors, and a single receptor can respond to several odorants¹³. This complex coding mechanism, although highly efficient, makes it extremely difficult to establish clear rules governing molecular recognition. Moreover, recent studies have suggested that the coding would be even more complex as antagonism could be a peripheral mechanism that would be as important as agonism¹⁴ (Figure 5 Figure 4).

Mammalian Olfaction

Recognition of odorants by ORs is mainly guided by hydrophobic and van der Waals interactions, resulting in low molecular affinity¹⁵. With the exception of a few conserved amino acids, the OR sequences show little overall identity. In particular, the ligand-binding pocket contains hypervariable residues¹¹, which remain relatively conserved between orthologs. This feature allows ORs to interact with a diverse range of molecules of varying structure, size and chemical properties.

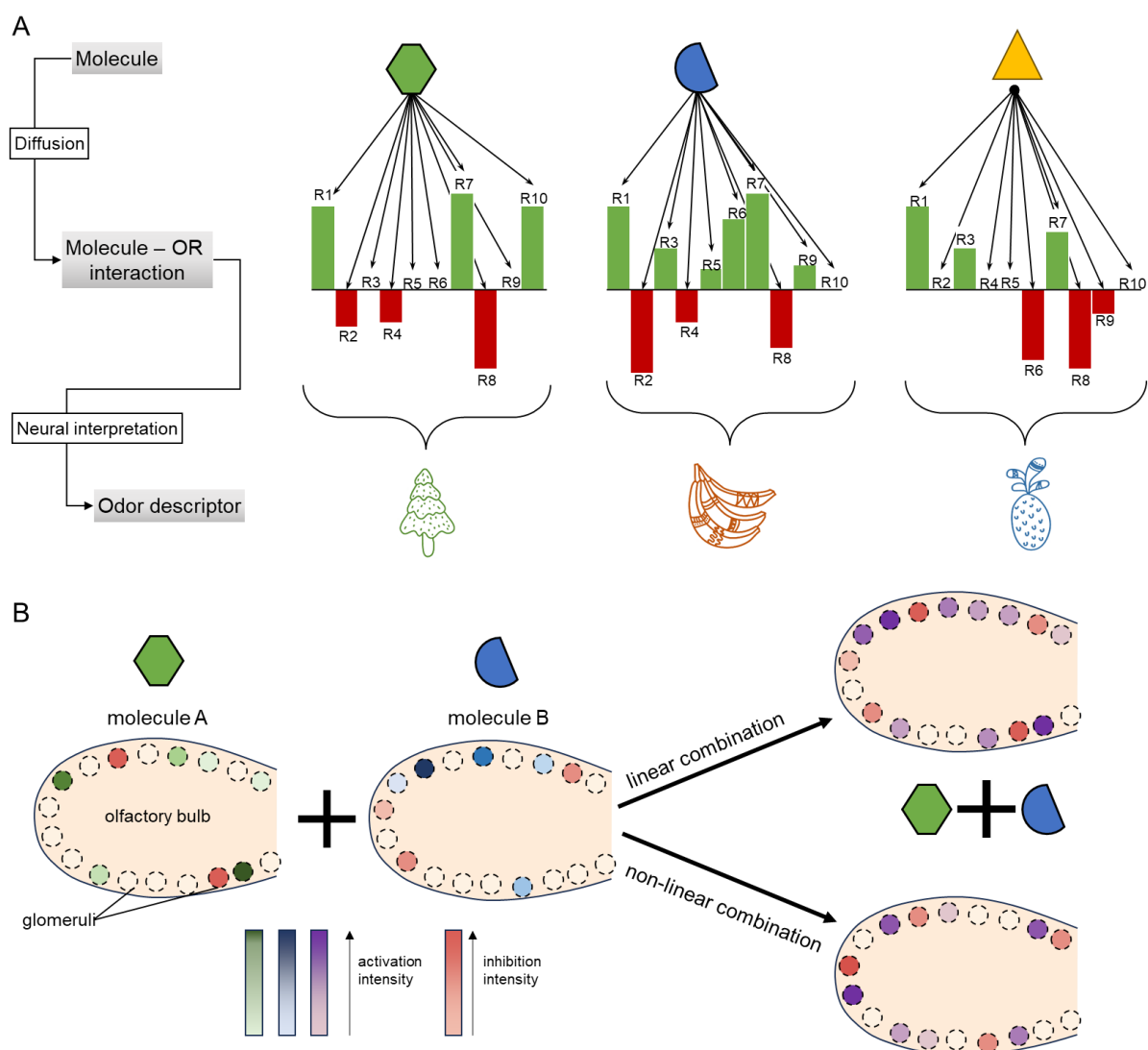


Figure 5. Scheme of the combinatorial code of olfaction. Molecules with different chemical functions are represented as geometric shapes. The response from Olfactory Receptors for each molecules appears as histograms with the height of the bar accounting for the strength of the response. Green and red bars represent the activation and inhibition respectively.

In the recent years, advances in deep learning, such as the AlphaFold2 model¹⁶, and the availability of experimental structures of ORs in apo form or bound to agonist had allow to understand functional role of several amino acids into the agonist recognition^{10, 17} (Fig. 3). Overall, these studies provides valuable insights into the activation mechanism and odorant specificity of the OR52 family, and paves the way for further experimental and computational studies to unravel the mechanisms underlying OR function.

Perireceptor events

Most odorants in the air are hydrophobic molecules, while the olfactory cilia are bathed in an aqueous mucus. In this protective fluid surrounding the neurons are enzymes and buffers that promote olfactory perception and are retained in most animals (Figure 6). The interactions between odorants and these biomolecules constitute a set of processes called peri-receptors. Studies have shown that this mucus would have a major role in the protection of the constituent tissues of the olfactory epithelium by harmful volatile molecules or pathogenic microorganisms. Perireceptor events can modify the concentration or the nature of odorant molecules present in the mucus, thereby altering the chemical information that is transformed into an electrical response at the level of olfactory sensory neurons.

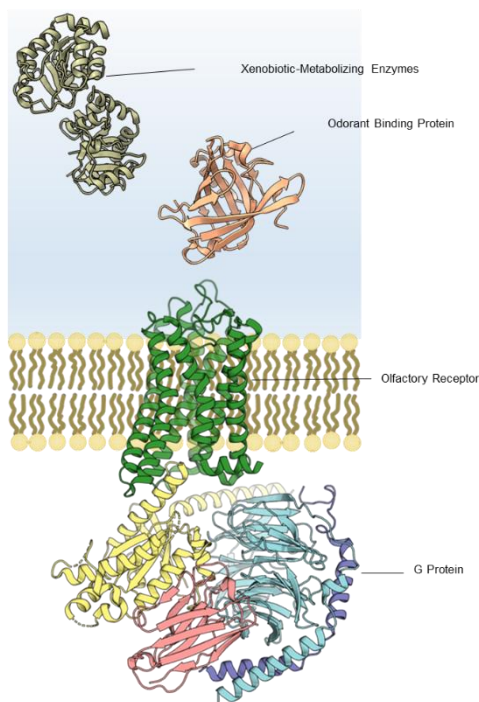


Figure 6. The different families of proteins expressed in the olfactory epithelium and involved perireceptor events. Odorant molecules present in inhaled air are solubilized in the mucus, which contains different types of soluble proteins, such as odorant-binding proteins (OBPs) and xenobiotic-metabolizing enzymes (XMEs). Odorant recognition occurs at the neuronal membrane, where olfactory receptors (ORs) are embedded. ORs belong to a family of proteins known to be involved in a wide range of cellular signaling processes. This family, the G protein-coupled receptors (GPCRs), also includes proteins responsible for taste perception (certain taste receptors) and light perception (opsins). These are transmembrane proteins composed of seven domains embedded in the cell membrane. At the intracellular level, the signal is transmitted by the G protein

Before interacting with olfactory receptors expressed on the surface of olfactory neurons, odorants pass through a layer of olfactory mucus. It is therefore assumed that the first event of peripheral olfactory detection involves at least one additional partner: an odorant-binding protein (OBP). Vertebrate OBPs, with a molecular weight of approximately 20 kDa, belong to the lipocalin family¹⁸, and are present at high levels in mucus (100 IM-1mM)¹⁹. OBPs presumably bind and solubilize volatile odorants, facilitating their diffusion across the mucus barrier to the ORs²⁰. Several other physiological functions have been suggested for OBPs. For example, they can attenuate odorant impulses by trapping excessive levels of odorants to prevent receptor inactivation and thus preserve a functional response. They can act as scavengers of excess odorants and eliminate stimuli. Recently, scavenging capabilities of OBPs towards several signaling molecules secreted by different pathogenic microorganisms have been demonstrated, suggesting an antimicrobial activity for vertebrate OBPs²¹.

In addition, binding of a porcine OBP to a human OR has been demonstrated, suggesting that a specific OR-OBP association can occur in the absence of any odorant. This provides a clue to another functional role for OBPs in perireceptor events of odor detection, but this must be validated for an OR with a known odorant ligand to infer physiological relevance of the OR-OBP complex. Specific interactions have also been reported between other members of the lipocalin family and membrane receptors. A model for tripartite OR-OBP-odorant interactions has also been proposed²².

Uncategorized References

1. Ehrlichman, H.; Bastone, L., Olfaction and emotion. In *Science of olfaction*, Springer: 1992; pp 410-438.
2. Doty, R. L., *Handbook of olfaction and gustation*. John Wiley & Sons: 2015.
3. Buck, L.; Axel, R., A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **1991**, *65* (1), 175-187.
4. Mayhew, E. J.; Arayata, C. J.; Gerkin, R. C.; Lee, B. K.; Magill, J. M.; Snyder, L. L.; Little, K. A.; Yu, C. W.; Mainland, J. D., Transport features predict if a molecule is odorous. *Proceedings of the National Academy of Sciences* **2022**, *119* (15), e2116576119.
5. Guth, H., Identification of character impact odorants of different white wine varieties. *Journal of Agricultural and Food Chemistry* **1997**, *45* (8), 3022-3026.
6. Sanchez-Lengeling, B.; Wei, J. N.; Lee, B. K.; Gerkin, R. C.; Aspuru-Guzik, A.; Wiltschko, A. B., Machine learning for scent: Learning generalizable perceptual representations of small molecules. *arXiv preprint arXiv:1910.10685* **2019**.
7. Bushdid, C.; Magnasco, M. O.; Vosshall, L. B.; Keller, A., Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **2014**, *343* (6177), 1370-1372.
8. Gerkin, R. C.; Castro, J. B., The number of olfactory stimuli that humans can discriminate is still unknown. *Elife* **2015**, *4*, e08127.
9. Pace, U.; Hanski, E.; Salomon, Y.; Lancet, D., Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature* **1985**, *316* (6025), 255-258.
10. Choi, C.; Bae, J.; Kim, S.; Lee, S.; Kang, H.; Kim, J.; Bang, I.; Kim, K.; Huh, W.-K.; Seok, C., Understanding the molecular mechanisms of odorant binding and activation of the human OR52 family. *Nature Communications* **2023**, *14* (1), 8105.
11. Pilpel, Y.; Lancet, D., The variable and conserved interfaces of modeled olfactory receptor proteins. *Protein Science* **1999**, *8* (5), 969-977.
12. Lalis, M.; Hladiš, M.; Abi Khalil, S.; Deroo, C.; Marin, C.; Bensafi, M.; Baldovini, N.; Briand, L.; Fiorucci, S.; Topin, J., A status report on human odorant receptors and their allocated agonists. *Chemical Senses* **2024**, *49*, bjae037.
13. Malnic, B.; Hirono, J.; Sato, T.; Buck, L. B., Combinatorial receptor codes for odors. *Cell* **1999**, *96* (5), 713-723.
14. Pfister, P.; Smith, B. C.; Evans, B. J.; Brann, J. H.; Trimmer, C.; Sheikh, M.; Arroyave, R.; Reddy, G.; Jeong, H.-Y.; Raps, D. A., Odorant receptor inhibition is fundamental to odor encoding. *Current Biology* **2020**, *30* (13), 2574-2587. e6.
15. Katada, S.; Hirokawa, T.; Oka, Y.; Suwa, M.; Touhara, K., Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. *Journal of Neuroscience* **2005**, *25* (7), 1806-1815.
16. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A., Highly accurate protein structure prediction with AlphaFold. *nature* **2021**, *596* (7873), 583-589.
17. Billesbølle, C. B.; de March, C. A.; van der Velden, W. J.; Ma, N.; Tewari, J.; Del Torrent, C. L.; Li, L.; Faust, B.; Vaidehi, N.; Matsunami, H., Structural basis of odorant recognition by a human odorant receptor. *Nature* **2023**, *615* (7953), 742-749.
18. Flower, D. R., Beyond the superfamily: the lipocalin receptors. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2000**, *1482* (1-2), 327-336.
19. Vincent, F.; Spinelli, S.; Ramoni, R.; Grolli, S.; Pelosi, P.; Cambillau, C.; Tegoni, M., Complexes of porcine odorant binding protein with odorant molecules belonging to different chemical classes. *Journal of Molecular Biology* **2000**, *300* (1), 127-139.

20. Briand, L.; Eloit, C.; Nespoulous, C.; Bézirard, V.; Huet, J.-C.; Henry, C.; Blon, F.; Trotier, D.; Pernollet, J.-C., Evidence of an odorant-binding protein in the human olfactory mucus: location, structural characterization, and odorant-binding properties. *Biochemistry* **2002**, *41* (23), 7241-7252.
21. Bianchi, F.; Flisi, S.; Careri, M.; Riboni, N.; Resimini, S.; Sala, A.; Conti, V.; Mattarozzi, M.; Taddei, S.; Spadini, C., Vertebrate odorant binding proteins as antimicrobial humoral components of innate immunity for pathogenic microorganisms. *PLoS One* **2019**, *14* (3), e0213545.
22. Vidic, J.; Grosclaude, J.; Monnerie, R.; Persuy, M.-A.; Badonnel, K.; Baly, C.; Caillol, M.; Briand, L.; Salesse, R.; Pajot-Augy, E., On a chip demonstration of a functional role for odorant binding protein in the preservation of olfactory receptor activity at high odorant concentration. *Lab on a Chip* **2008**, *8* (5), 678-688.

3.1 – MAMMALIAN OLFACTORY RECEPTOR

- Chapitre 1.2. Chimie des substances odorantes. X. Fernandez, S. Fiorucci, J. Topin.

This publication offers an overview of the mechanisms at the origin of smell detection. As it is written in French, it is a dedication to my family who is not fluent in English, and more specifically to my mother, a French teacher, so that she can catch a glimpse of this topic.

- C. A. de March[‡], J. Topin[‡], E. Bruguera, G. Novikov, K. Ikegami, H. Matsunami, J. Golebiowski. Odorant Receptor 7D4 Activation Dynamics, [*Angewandte Chemie International Edition*](#), **2018**.

This publication is a landmark in the prediction of ligand-induced receptor activation. It demonstrates that it is possible to evaluate the effect of a ligand or a point mutation on the receptor response. Understanding how odours activate olfactory receptors (ORs) and how specific amino acid modifications influence this activation is essential to deciphering human olfaction. This study was based on an interdisciplinary approach combining site-directed mutagenesis, functional testing and computer modelling. We have revealed the activation mechanism of OR7D4, a receptor whose genetic variants influence the perception of androstenone. Molecular simulations (0.12 ms in total) identified an activation pathway linking the ligand binding pocket to the G protein interface, which corresponds to the mechanisms described for GPCRs whose structure is known. The model accurately reflected the experimental data and predicted, in particular, that the WM variant of OR7D4 remained inactive. Activation converged on a sequence motif specific to OR. Overall, this work highlights that reliable homology modelling can effectively capture the dynamics of ORs underlying odour perception.

- J. Topin*, C. Bouysset, J. Pacalon, Y. Kim, M. Rhyu, S. Fiorucci, J. Golebiowski. Functional Molecular Switches of Mammalian G Protein-Coupled Bitter-Taste Receptors, [*Cellular and Molecular Life Sciences*](#), **2021**.

Leading an international consortium of computational and experimental scientists, I coordinated efforts to validate the TAS2R16 model through site-directed mutagenesis and functional assays. This work identifies key molecular switches involved in agonist sensing and signaling, enhancing our understanding of TAS2R function. This integrative approach is a key step to improve the modeling of bitter taste receptors (TAS2Rs), a poorly understood subgroup of GPCRs.

- M. Hladiš, M. Lalis, S. Fiorucci, & J. Topin*. Matching receptor to odorant with protein language and graph neural networks, [*The Eleventh International Conference on Learning Representation*](#), **2023**.

This publication, for which I am the corresponding author, focuses on the development of a predictive model capable of distinguishing agonists from non-agonists across the entire set of olfactory receptors (more than 400 receptors and over 6,000 molecules evaluated). This paper paves the way toward understanding the combinatorial code of olfaction. This work was advertised in the Le Nez newsletter (<https://mag.bynez.com/themes/jeremie-topin-2/>).

- M. Lalis, M. Hladiš, S. A. Khalil, L. Briand, S. Fiorucci, J. Topin*. M2OR: a database of olfactory receptor–odorant pairs for understanding the molecular mechanisms of olfaction, [*Nucleic Acids Research*](#), **2024**.

The M2OR database represents a pivotal resource in olfactory research, offering the most comprehensive compilation of odorant–odorant receptor (OR) pairs to date. By aggregating over 50,000 experimentally validated interactions from published studies and public databases, M2OR enables researchers to explore the molecular basis of smell with unprecedented depth. It facilitates the identification of receptor specificity, supports predictive modeling of odor perception, and accelerates the design of novel compounds for applications in flavor, fragrance, and biomedical fields. As such, M2OR is an essential tool for advancing our understanding of olfactory coding and receptor-ligand recognition.

Chapitre 1.2. Chimie des substances odorantes

Xavier Fernandez, Sébastien Fiorucci, Jérémie Topin

Institut de Chimie de Nice, UMR 7272, Université Côte d'Azur, Parc Valrose, 06108 Nice cedex 2, France.

Comme d'autres esprits voguent sur la musique,

Le mien, ô mon amour ! Nage sur ton parfum.

Charles Baudelaire, La chevelure.

1. Introduction

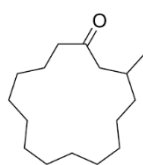
Les huiles essentielles sont des mélanges particulièrement complexes de composés organiques d'origine naturelle. A l'exception notable de celles obtenues par expression mécanique (agrumes), elles sont majoritairement constituées de molécules volatiles à l'origine de leur odeur particulière. L'analyse de ces composés présente un intérêt majeur, tant pour la compréhension des arômes naturels que pour leurs applications dans les domaines de la parfumerie, de la cosmétique, de l'agroalimentaire ou encore de la pharmacologie.

L'objectif de ce chapitre est double. Dans un premier temps, il s'agira de passer en revue la diversité chimique et structurale qui caractérise les molécules odorantes présentes dans les huiles essentielles. Dans un second temps, seront abordés les mécanismes moléculaires et physiologiques à l'origine de l'odorat. Bien que ce sens soit quotidiennement sollicité, les phénomènes moléculaires de la reconnaissance des odorants ne sont encore que partiellement élucidés et continuent de faire l'objet de recherches intensives.

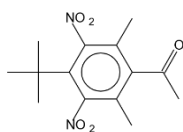
2. Structures des composés odorants

De nouvelles substances odorantes sont continuellement identifiées et exploitées par l'industrie de la parfumerie et des cosmétiques. Les molécules odorantes présentent une grande diversité en termes de fonction chimique ou de structure. Des recherches récentes menées par Mayhew et ses collègues ont tenté de définir l'espace chimique des molécules odorantes à l'aide de techniques d'apprentissage automatique¹. Leur modèle, entraîné sur des molécules odorantes précédemment identifiées, suggère que le transport dans l'environnement est la caractéristique la plus importante qu'une molécule doit posséder pour déclencher une perception olfactive. En analysant une base de données de 166 milliards de composés chimiques, leurs modèles ont prédit qu'il existe au moins 40 milliards de composés odorants potentiels !

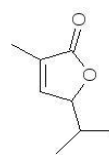
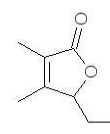
A titre d'exemple, le tableau 1 **Erreur ! Source du renvoi introuvable.** illustre l'observation que des structures chimiques très différentes peuvent conduire à des perceptions quasi-identiques, alors que des structures très similaires déclenchent des sensations olfactives très éloignées. La création de structures chimiques nouvelles, à l'origine d'odeurs présentant un certain intérêt est souvent une question de sérendipité (découverte par chance).

Tableau 1 : structures chimiques et leur odeur perçue.


Adrostenol

musc


Musk ketone

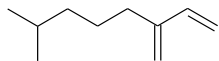
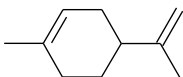
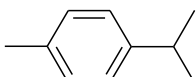
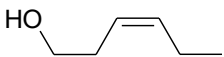
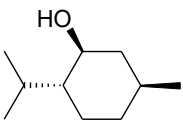
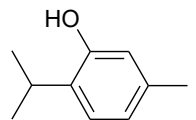
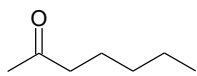
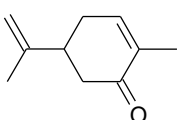
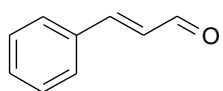
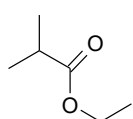
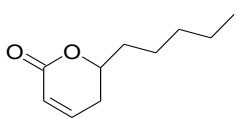
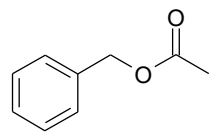
musc

isopropyl-
methylfuranone
herbacée


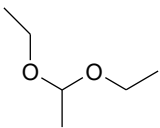
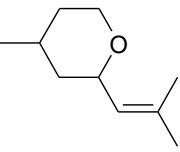
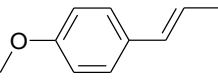
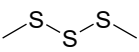
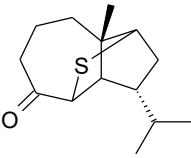
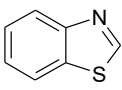
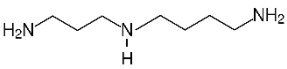
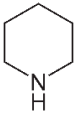
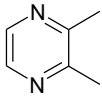
ethyl-dimethylfuranone

épicée

Les constituants des huiles essentielles possèdent un squelette hydrocarboné qui peut être linéaire, cyclique ou encore aromatique. Ils peuvent posséder toutes les grandes fonctions de la chimie organique ² : alcools, composés carbonylés (principalement aldéhydes et cétones), esters, phénols, et dans une moindre mesure dérivés azotés et soufrés (Tableau 2). Néanmoins les terpènes (hydrocarbures en C₁₀ ou C₁₅) et terpénoïdes (terpènes fonctionnalisés) sont de loin les plus abondants.

Tableau 2 : principales familles chimiques des odorants. Les perceptions olfactives des molécules apparaissent en italique.

Fonction	Linéaire	Cyclique	Aromatique
Hydrocarbure	 myrcène <i>musqué, herbacé</i>	 limonène <i>agrumes</i>	 <i>p</i> -cymène <i>origan, thym</i>
Alcool	 (Z)-hex-3-en-3-ol <i>herbe coupée</i>	 menthol <i>menthe</i>	 thymol <i>thym, origan</i>
Carbonyle	 heptan-2-one <i>fruitée, épicée</i>	 carvone <i>graine de carvi, menthe</i>	 cinnamaldehyde <i>cannelle</i>
Ester / lactone	 isobutyrate d'éthyle <i>sucrée, fruitée</i>	 massoïa lactone <i>noix de coco, crémeux</i>	 acétate de benzyle <i>jasmin</i>

Ether	 1,1-diéthoxyéthane <i>éthérée, herbe</i>	 oxyde de rose <i>rose</i>	 anéthol <i>anis</i>
Soufré	 trisulfure de diméthyle <i>ail, oignon</i>	 mint sulfide <i>menthe, sauge</i>	 benzothiazole <i>caoutchouteux, soufré</i>
Azoté	 spermidine <i>sperme, pourri</i>	 pipéridine <i>poivre, ammoniacque</i>	 2,4-diméthyl pyrazine <i>chocolat, noisette</i>

3. Les mécanismes de l'olfaction

Les mécanismes biologiques dictant la reconnaissance des odeurs n'ont été que très récemment étudiés. Les grands progrès dans la compréhension de l'olfaction ont débuté en 1985 par la découverte des récepteurs olfactifs (ROs) chez les vertébrés ³ et ont continué en 1989 par la mise en évidence de l'intervention de protéines G dans le mécanisme de l'olfaction ⁴. En 1991, Buck et Axel ont identifié les gènes codant pour les ROs et ont par suite été récompensés en 2004 par un prix Nobel de médecine et physiologie⁵. Le nombre total de gènes ROs a été estimé à environ 400 chez l'homme et peut atteindre près de 2000 chez l'éléphant, ce qui fait des ROs la plus grande famille génétique ! La multitude d'interactions entre les divers protagonistes de la perception d'une odeur conduit à la production d'un signal de nature combinatoire ⁶, capable de discriminer un nombre virtuellement illimité d'odorants ⁷. Dans la suite, nous passerons en revue la transformation d'un signal chimique en un signal électrique cheminant le long des neurones situés dans notre bulbe olfactif.

3.1. Le bulbe olfactif, l'organe de la perception des odeurs

Situé à l'interface entre la cavité nasale et les régions corticales supérieures, le bulbe olfactif est une structure neuronale majeure du système nerveux central, intégrée au cortex cérébral antérieur. Il est le siège de la première étape de traitement des informations olfactives.

Le bulbe olfactif se trouve à la croisée de deux voies qui permettent aux molécules odorantes de l'atteindre: la voie directe (ou ortho-nasale) est celle que l'on connaît le plus. Cette voie rejoint directement le bulbe olfactif à travers nos fosses nasales. Nous l'utilisons à chaque fois que nous respirons par le nez, comme lors de l'évaluation sensorielle d'un parfum par exemple (Figure 1).

La voie indirecte (ou rétro-nasale) permet quant à elle d'atteindre l'épithélium olfactif depuis le fond de notre palais. C'est elle qui rend possible la perception des molécules odorantes libérées lors de l'ingestion et de la mastication des aliments. Ce qui est communément désigné comme goût est en réalité la combinaison de la perception de molécules sapides qui confèrent une grande partie de la

richesse gustative des aliments ⁸. On ne distingue en effet que 5 grands types de molécules sapides : sucré, salé, amer, acide et umami. Le reste de la perception du goût est donc lié à la perception retro-nasale de molécules odorantes contenues dans nos aliments (ainsi qu'à la perception de la texture de l'aliment, mais qui se rapproche plutôt du sens du toucher, *via* l'activation des nerfs trijumeaux).

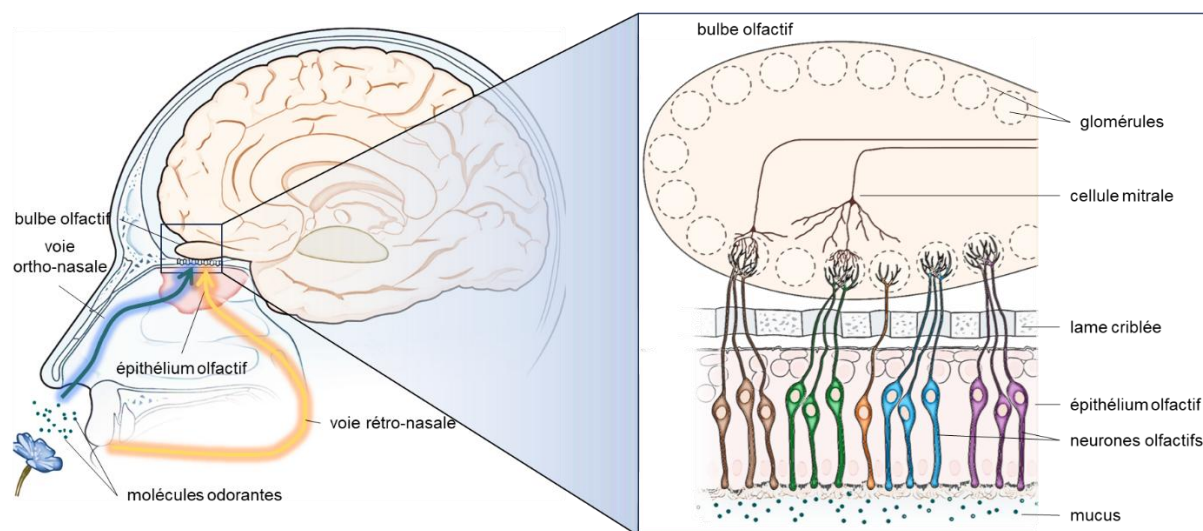


Figure 7. Coupe sagittale d'un crâne humain. Les parfums que nous respirons par le nez empruntent la voie dite directe (ou ortho-nasale) afin d'atteindre le bulbe olfactif qui contient les neurones olfactifs. Lors de la mastication, les molécules odorantes libérées dans la bouche rejoignent ce même bulbe olfactif à travers la voie rétro-nasale (en jaune).

Le bulbe olfactif est le seul endroit où notre cerveau est en contact quasi-immédiat avec l'environnement extérieur. En effet, les neurones olfactifs dont l'extrémité baigne dans le mucus, traversent la paroi osseuse de notre crâne (la lame criblée) et atteignent le bulbe olfactif (Figure 1).

Dans la partie intracrânienne, on distingue les glomérules, qui sont des centres intégrateurs de l'information envoyée par les neurones olfactifs. Ces glomérules, dans lesquels convergent les neurones olfactifs, vont moduler la réponse de certains. Ils permettent ainsi de réguler la formidable masse d'informations qui sera ensuite envoyée, *via* les cellules mitrales, dans le reste du cortex cérébral⁹. Les extrémités des neurones olfactifs forment des cils qui baignent dans le mucus olfactif, un fluide physiologique qui évite le dessèchement de la muqueuse olfactive, elle-même constituée de différents types de cellules (Figure 1). C'est dans ce mucus qu'ont lieu l'ensemble des phénomènes moléculaires qui vont initier la transformation d'un signal chimique (portée par la molécule odorante) en un signal neuronal qui sera transformé en une perception sensorielle au sein de notre cerveau.

Ce signal est transmis au bulbe olfactif puis se propage ensuite à d'autres aires cérébrales, telles que le cortex olfactif primaire, qui permet de reconnaître l'odeur et d'en apprécier la valence, l'hippocampe, dont le rôle consiste à mémoriser l'odorant, ou l'hypothalamus qui réalise une liaison entre l'odeur perçue et le vécu, permettant éventuellement de déclencher une émotion ¹⁰.

3.2. La perception des odeurs d'un point de vue moléculaire

Lors de l'inhalation, les molécules odorantes présentes dans l'air entrent en contact avec le mucus qui couvre l'épithélium olfactif comme schématisé dans la Figure 2. Le mucus contient différents types de protéines, dont des protéines de liaison aux odorants, ainsi que des protéines de dégradation. Ces deux types de protéines sont à l'origine de ce que l'on nomme les phénomènes péri-récepteurs, c'est-à-dire qui interviennent avant la reconnaissance des odorants par les ROs ¹¹.

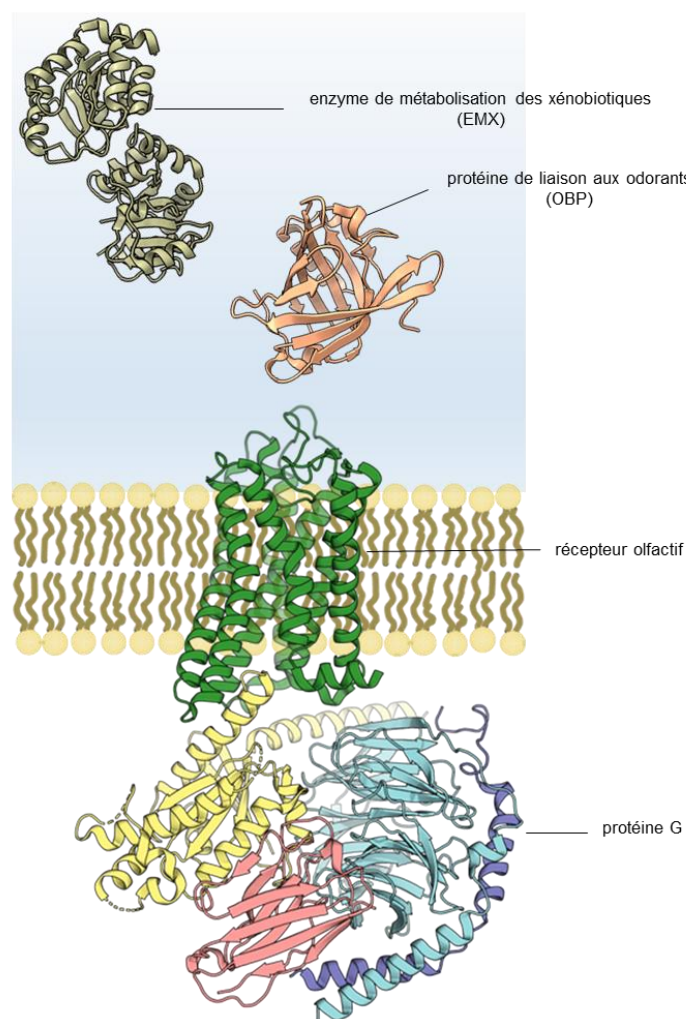


Figure 2. Les différentes familles de protéines exprimées au niveau de l'épithélium olfactif et impliquées dans la reconnaissance des odorants. Les molécules odorantes présentes dans l'air inhalé sont solubilisées dans le mucus qui contient différents types de protéines solubles : les protéines de liaison aux odorants (OBP) ou des enzymes de métabolisation des xénobiotiques (EMX). La reconnaissance des odorants se fait au niveau de la membrane du neurone dans laquelle sont enchâssés les Récepteurs Olfactifs (RO). Les RO appartiennent à une famille de protéines connues pour être impliquée dans une multitude de processus de signalisation cellulaire. Cette famille, les Récepteurs Couplés à une Protéine G (RCPG) englobe aussi les protéines qui nous permettent de percevoir le goût (certains récepteurs gustatifs) et celles qui nous permettent de percevoir la lumière (les opsines). Il s'agit de protéines transmembranaires, constituées de 7 domaines enchâssés dans la membrane d'une cellule. Au niveau intracellulaire, le signal est transmis par la protéine G.

3.3. Les phénomènes péri-récepteurs

3.3.1. Les protéines de liaison des odorants : OBP

Les protéines de liaison aux odorants (Odorant Binding Proteins ou OBP) sont des protéines solubles de la famille des lipocalines sécrétées en abondance dans le mucus nasal d'une grande variété d'espèces animales, y compris chez l'être humain¹². Les OBP sont sécrétées par l'épithélium olfactif dans le mucus nasal. Différents sous-types d'OBP ont été mis en évidence chez une même espèce animale. Par exemple, deux OBP ont été décrites chez le porc, quatre chez la souris ou trois chez le lapin. L'expression d'une seule OBP est également observée chez l'Homme. Elle est limitée à la plus haute région de la fosse nasale, où les molécules odorantes sont détectées par les neurones porteurs des récepteurs olfactifs.

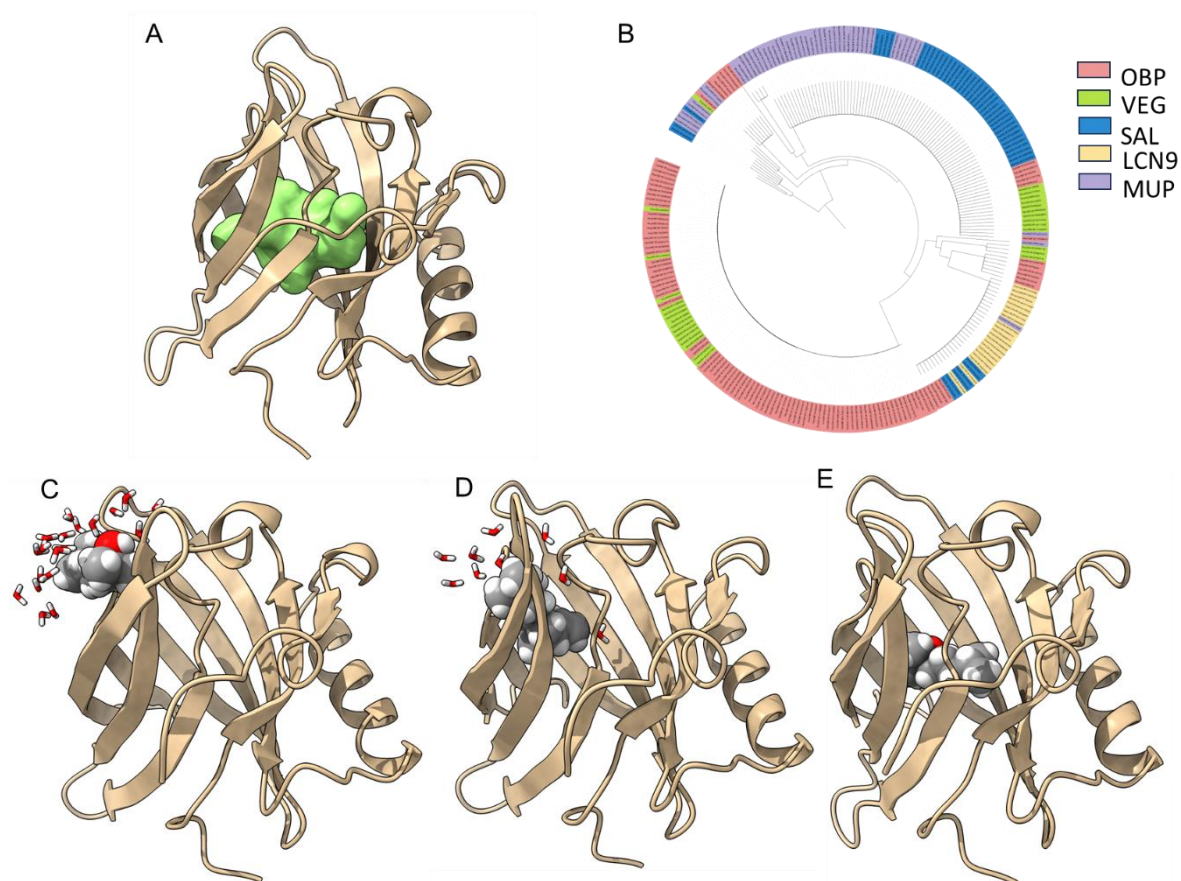


Figure 3 : Structure et mécanisme de reconnaissance des odorants par les OBP. A Illustration de la cavité de liaison des OBP en vert au centre de la structure. B arbre phylogénétique des lipocalines (OBP : odorant binding protein, VEG : von Ebner's glands, SAL : Salivary lipocalin, LCN9 : lipocalin genes 9, MUP : Major Urinary Proteins). C,D,E Illustration du chemin de liaison d'un odorant dans une OBP. Durant la simulation, l'odorant rejoint le cœur de la protéine selon un chemin conservé. Ce processus s'accompagne d'une désolvatation du ligand.

La résolution des structures d'OBP a révélé que cette famille de protéines partage un repliement très conservé caractérisé par une vaste cavité hydrophobe centrale (Figure 4A). Elles présentent un tonneau β constitué de huit feuillettes β anti-parallèles reliés par sept boucles et coiffé par une hélice α (Figure 4A). Ce tonneau β délimite une cavité hydrophobe, appelée calice, qui permet la liaison de molécules lipophiles, ce qui leur a valu le nom de lipo-calines. Cette architecture est particulièrement adaptée au transport des molécules odorantes à travers le mucus vers les neurones olfactifs. Par ailleurs, cette structure se caractérise par une stabilité remarquable face à la température, aux solvants organiques et à la digestion protéolytique, ce qui en fait des candidats intéressants pour le développement de biocapteurs robustes¹³. Les OBP ont la propriété de pouvoir lier des odorants variés en terme de taille et de fonctionnalités¹⁴. Le processus de liaison se fait selon un chemin de translocation conservé (Figure 4CDE), durant lequel, on observe une désolvatation du ligand¹⁵.

Le rôle essentiel des OBP dans le déclenchement de comportements et du codage de l'odeur a été démontré chez les insectes, mais il reste encore à démontrer chez les vertébrés et les humains en particulier. Ces protéines font l'objet de nombreuses recherches dans cette optique.

3.3.2. Les Enzymes de Métabolisation de Xénobiotiques : EMX

Un xénobiotique est une molécule d'origine exogène. Ces xénobiotiques, de manière générale, sont métabolisés par un réseau d'enzymes qui favorisent leur élimination par diverses voies d'excrétion.

Lorsque ce métabolisme a lieu dans le foie (l'organe métabolique majeur), la voie d'excrétion est l'urine ou la bile.

Les odorants que nous respirons sont donc des xénobiotiques. Dans le mucus nasal, différentes enzymes joueraient un rôle de métabolisation de ces odorants, afin d'éviter la saturation des récepteurs olfactifs. Elles nous permettent de percevoir un gradient de molécules par flairages successifs. Sans elles, le mucus nasal se retrouverait très rapidement saturé en molécules odorantes, nous empêchant ainsi de percevoir de nouvelles odeurs dans des échelles de temps très rapides. Différentes observations confortent l'hypothèse selon laquelle la biotransformation des molécules odorantes aurait un impact notable sur la sensibilité du système olfactif. La perception olfactive résulterait de la détection à la fois des molécules odorantes mais aussi des métabolites formés au niveau du mucus et/ou de la muqueuse olfactive¹⁶.

Les enzymes du métabolisme des xénobiotiques (EMX) occupent une place essentielle dans la biotransformation et l'élimination des composés exogènes. Fonctionnant au sein d'un réseau intégré de réactions enzymatiques, elles catalysent divers processus visant à rendre ces molécules plus hydrophiles et ainsi plus facilement éliminables par l'organisme. On rencontre ainsi dans le mucus nasal humain une alcool déshydrogénase ainsi qu'une aldéhyde déshydrogénase et au moins deux Uridine-di-Phosphate-GlucuronosylTransférases (UGT)¹⁷. Par exemple, la glucurono-conjugaison de molécules odorantes, catalysée par les UGT, rend les molécules odorantes inactives d'un point de vue olfactif. Cette réaction interviendrait ainsi dans l'arrêt du signal olfactif¹⁸.

Lors de l'inhalation, des molécules odorantes diffusent par la cavité nasale jusqu'à l'épithélium olfactif. À ce niveau, elles interagissent avec les ROs exprimés à la surface des neurones olfactifs. La formation du complexe odorant-récepteur induit une cascade de réaction biochimiques qui transforme un stimulus chimique (odorant) en un influx neuronal interprétable par le cerveau.

Au-delà des phénomènes péri-récepteurs, les odorants vont jouer leur rôle fondamental : déclencher la cascade de détection du signal olfactif. Dans le fond du mucus nasal, se trouvent les cils des neurones olfactifs. Dans la membrane de ces cils, sont enchâssés les ROs, ces protéines qui sont à l'origine de la cascade d'événements qui déclencheront la perception olfactive.

3.4. La détection du signal : Les récepteurs olfactifs

Contrairement à leur nom, les récepteurs olfactifs (ROs) ne sont pas spécifiques des neurones sensoriels olfactifs, où ils sont exprimés à la surface des cils et participent à la détection des odeurs. En effet, leur expression ectopique a été largement démontrée dans de nombreux tissus non olfactifs, tels que les testicules, le foie, les poumons, les intestins, la peau, le cœur ou le sang. Certains ROs y présentent une expression spécifique à un seul type cellulaire, tandis que d'autres sont largement distribués dans l'organisme¹⁹.

Dans ces tissus, les ROs rempliraient diverses fonctions cellulaires de signalisation et d'orientation, comme la chimiotaxie observée dans les cellules spermatiques. Ainsi une étude a montré qu'il existerait un lien entre anosmie spécifique et infertilité²⁰! Ce rôle de guidage est également envisagé au niveau de l'épithélium olfactif, dont les neurones se régénèrent continuellement grâce à des cellules souches basales²¹. Les nouveaux neurones olfactifs semblent retrouver leur glomérule cible dans le cortex grâce à l'action de leurs propres ROs, qui interviendraient ainsi avant même la perception sensorielle, en orientant la croissance neuronale. Ainsi, loin d'être confinés au nez, les récepteurs olfactifs apparaissent comme des capteurs polyvalents jouant un rôle clé dans divers processus physiologiques et pathologiques, et pourraient constituer à l'avenir des cibles diagnostiques et thérapeutiques prometteuses.

3.4.1. D'un signal chimique vers un signal électrique

Le système olfactif suit une stratégie de reconnaissance extrêmement sophistiquée reposant sur un répertoire d'environ 400 ROs présents à la surface de l'épithélium olfactif⁵. Ces récepteurs sont au commencement d'une cascade de processus biologiques menant à la transformation d'un signal chimique en potentiels électriques interprétables par notre cerveau (Figure 4). L'activation du RO va déclencher une voie de signalisation faisant intervenir des protéines intracellulaires. Cette voie est à l'origine de la modification du gradient de concentration d'ions entre l'intérieur de la cellule (le cytoplasme) et l'extérieur de la cellule, ici le mucus olfactif). Ce changement de concentration va induire une dépolarisation de la membrane du neurone. Ce signal électrique va se propager le long de l'axone du neurone, jusqu'à atteindre le glomérule.

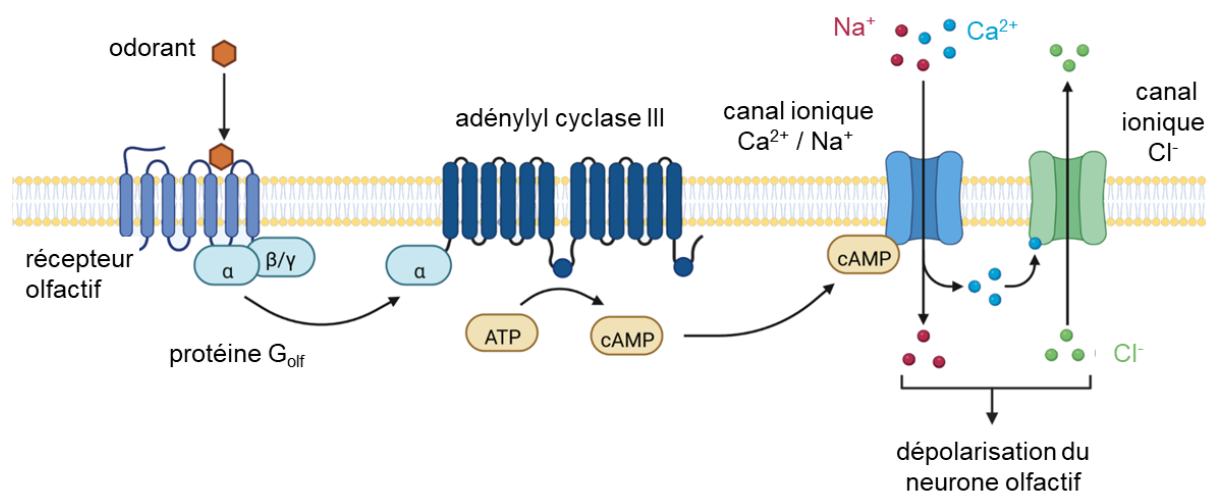


Figure 4. Illustration du mécanisme de transduction de l'information chimique en signal électrique. La formation du complexe odorant-récepteur induit une cascade de réaction biologique dont la première étape est l'activation de la protéine Golf. Le message est ensuite transmis à l'adénylyl cyclase qui convertit l'adénosine triphosphate (ATP) en adénosine 3,5-monophosphate cyclique (cAMP). L'augmentation de la concentration en cAMP entraîne à son tour l'ouverture d'un canal ionique qui se traduit par l'entrée de cations Ca^{2+} et Na^{+} . Cette augmentation de concentration à son tour induit l'ouverture de canaux ioniques spécifiques aux anions Cl^{-} . Cette accumulation d'ions de part et d'autre de la membrane crée une dépolarisation du neurone olfactif qui se propage en potentiel d'action jusqu'au bulbe olfactif.

Les neurones olfactifs ont la particularité d'exprimer majoritairement qu'un seul type de récepteur olfactif²². Par conséquent, l'activation d'un type de récepteur par un odorant va alors entraîner l'activation des neurones qui expriment ce récepteur. Observer l'activation d'un neurone est alors équivalent à observer l'activation d'un RO au niveau moléculaire.

3.4.2. Le code combinatoire de l'olfaction

A l'échelle moléculaire, la reconnaissance des odorants est basée sur la reconnaissance de formes plutôt que sur l'identification isolée d'une molécule spécifique, la grande majorité des récepteurs olfactifs répondent à différentes molécules²³. Lors de l'inhalation, un odorant interagit avec un sous-ensemble de récepteurs olfactifs (Figure 5). L'activation différentielle de ce répertoire est ensuite interprétée par le système nerveux central comme une odeur. Dans ce schéma, une odeur est définie par un code combinatoire d'activation de récepteurs olfactifs : une molécule peut activer un sous-ensemble de récepteurs, et chaque récepteur peut répondre à plusieurs molécules différentes⁶. Ce concept est illustré dans la figure 5.

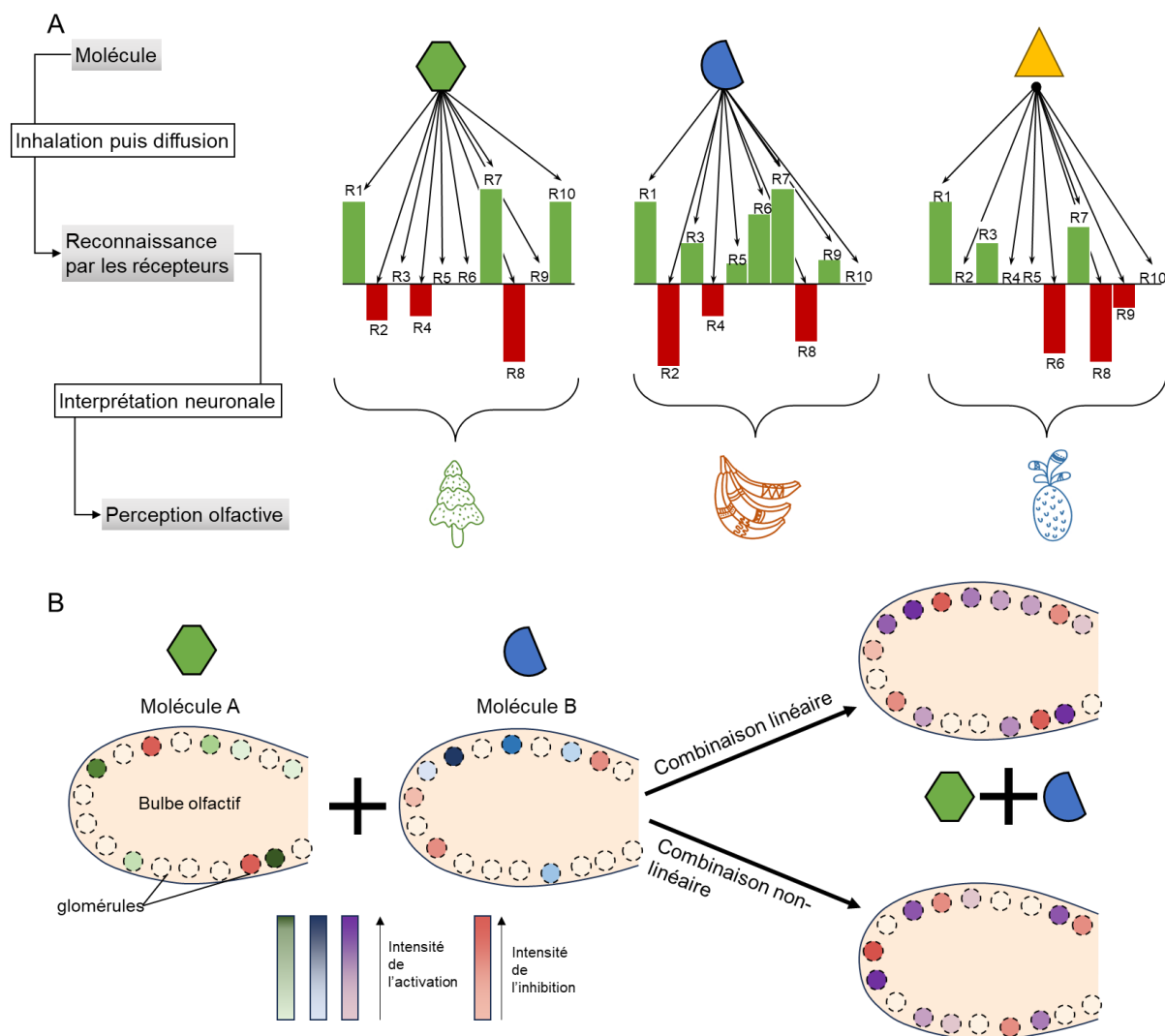


Figure 5 : Illustration de l'interprétation de la molécule par les récepteurs olfactifs. A. L'interaction d'une molécule odorante avec l'ensemble des récepteurs olfactifs portés par des neurones indépendants induit une activation et inhibition qui dépendra de l'affinité de cet odorant avec les ROs. Chaque odorant va générer une combinaison caractéristique, en fait associée aux glomérules qui sont activés via les neurones olfactifs. Cette réponse, a priori unique pour chaque odorant sera interprétée comme une odeur par le sujet. La réponse des récepteurs olfactifs pour chaque molécule apparaît sous forme d'histogrammes dont la hauteur des barres correspond à l'intensité de la réponse. Les barres vertes et rouges représentent respectivement l'activation et l'inhibition. B. Les images olfactives. Au niveau du bulbe olfactif, les glomérules seront activés ou inhibés différemment selon les molécules inhalées. Le mélange des deux odorants, A+B peut éventuellement être la somme des deux premiers signaux, bien que cela ne soit pas systématique.

Comprendre le mécanisme par lequel un nombre relativement faible de récepteurs, typiquement quelques centaines²⁴, permet la discrimination d'une multitude de molécules odorantes reste un défi majeur. De nouvelles recherches indiquent que même des altérations mineures de la fonctionnalité d'un seul récepteur peuvent avoir des conséquences perceptives notables²⁵. Bien que nous sachions que la famille des gènes de ROs présente une diversité importante, la nature complexe du code olfactif et nos connaissances limitées sur les paires OR-odorant spécifiques posent des problèmes pour traduire cette variabilité en distinctions perceptives²³.

4. Vers la numérisation de l'olfaction

La perception sensorielle induite par les odorants est la résultante d'un processus d'une formidable complexité. Durant de nombreuses années il a été admis qu'il était extrêmement difficile d'obtenir une méthode universelle de prédiction de l'odeur d'une molécule (et à plus forte raison d'un mélange) sur la base des structures moléculaires ou d'une ou plusieurs propriétés chimiques intrinsèques des odorants ²⁶. Avant même d'évoquer l'aspect purement chimique de ces phénomènes, il convient de souligner un aspect important du défi de la prédiction des odeurs. La réponse physiologique liée à la perception d'une molécule odorante n'est pas une propriété physique ou une activité biologique mesurable, mais une perception qui ne peut être que décrite qualitativement par un sujet. Ces approches sont donc asservies à la richesse de notre vocabulaire, à notre vécu ainsi qu'à notre entraînement à reconnaître les odeurs.

4.1. Déorphanisation numérique des récepteurs olfactifs

Les progrès de la biologie moléculaire ont ouvert la voie vers l'identification d'odorants pour un récepteur dont on ne connaît pas les agonistes. On parle dans ce cas de récepteurs orphelins. Ces approches, dites de déorphanisation restent néanmoins coûteuses. Les données regroupant les résultats des tests d'associations de plus de 75 000 paires odorants-récepteurs ont été regroupées et sont désormais en libre accès à travers la base de données M2OR ²⁷. Néanmoins, le code combinatoire complet n'est connu que pour une poignée d'odorants et ne permet pas encore de déchiffrer les règles de reconnaissance des odorants pour les ROs ²³.

En s'appuyant sur les données de M2OR ainsi que sur les avancées en apprentissage profond (ici, des réseaux de neurones à graphes), un modèle de prédiction des interactions odorants-récepteurs a permis d'établir l'ensemble du code combinatoire (400 récepteurs) pour 6000 molécules connues ²⁸.

4.2. Prédiction de l'odeur d'une molécule

Le développement de l'intelligence artificielle couplé à l'ouverture de bases de données sur les odorants ont donné lieu à des avancées majeures dans la prédiction de l'odeur d'une molécule ²⁹. Cette équipe a abandonné les descripteurs chimiques classiques, jugés inadaptés aux discontinuités entre espace chimique et espace perceptuel et développé un réseau neuronal à passage de messages (MPNN) représentant chaque odorant sous forme de graphe. Leur modèle atteint désormais des performances prédictives comparables à celles d'un panel d'individus non entraînés !

5. Conclusion

La compréhension du système olfactif repose aujourd'hui sur des bases neuro-anatomiques solides : l'architecture neuronale impliquée dans la perception des odeurs est bien décrite, depuis l'épithélium olfactif jusqu'au bulbe olfactif, véritable plateforme d'intégration des signaux sensoriels olfactifs. La nature intrinsèquement combinatoire de l'olfaction, où chaque odeur résulte d'un ensemble complexe d'interactions entre les molécules odorantes, les récepteurs olfactifs et les protéines du mucus, confère à ce sens une dimension fascinante.

Les molécules odorantes constituent une classe extraordinairement diverse, dont seule une fraction infime a été identifiée à ce jour. Cette richesse, encore largement inexplorée, représente une formidable opportunité d'innovation pour le secteur des arômes et des parfums, où la découverte de nouvelles matières premières reste un moteur essentiel de créativité.

Parallèlement, les mécanismes de reconnaissance moléculaire entre odorants et récepteurs commencent à être élucidés grâce à deux avancées majeures : d'une part, la résolution expérimentale des structures actives et inactives de RO ³⁰; d'autre part, l'essor de l'intelligence artificielle, qui permet désormais d'aborder de manière systématique le décryptage du code combinatoire de l'olfaction. Ces

outils ouvrent la voie non seulement à la prédiction des interactions ligand–récepteur, mais également à la prédiction de l'odeur perçue à partir de la structure chimique d'une molécule.

Ainsi, l'objectif de pouvoir prédire l'odeur d'un mélange complexe, comme celui d'une huile essentielle, longtemps considéré comme utopique, serait en passe de devenir techniquement envisageable dans un futur proche.

Références bibliographiques

Uncategorized References

1. Mayhew, E. J.; Arayata, C. J.; Gerkin, R. C.; Lee, B. K.; Magill, J. M.; Snyder, L. L.; Little, K. A.; Yu, C. W.; Mainland, J. D., Transport features predict if a molecule is odorous. *Proceedings of the National Academy of Sciences* **2022**, 119 (15), e2116576119.
2. Fernandez, X.; Kerverdo, S.; Duñach, E.; Lizzani-Cuvelier, L., Les hétérocycles dans la chimie des arômes. *Actualité Chimique* **2002**, (4), 4-14.
3. Pace, U.; Hanski, E.; Salomon, Y.; Lancet, D., Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature* **1985**, 316 (6025), 255-258.
4. Jones, D. T.; Reed, R. R., G (olf): an olfactory neuron specific-G protein involved in odorant signal transduction. *Science* **1989**, 244 (4906), 790-796.
5. Buck, L.; Axel, R., A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **1991**, 65 (1), 175-187.
6. Malnic, B.; Hirono, J.; Sato, T.; Buck, L. B., Combinatorial receptor codes for odors. *Cell* **1999**, 96 (5), 713-723.
7. Bushdid, C.; Magnasco, M. O.; Vosshall, L. B.; Keller, A., Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **2014**, 343 (6177), 1370-1372.
8. Small, D. M.; Prescott, J., Odor/taste integration and the perception of flavor. *Experimental brain research* **2005**, 166 (3), 345-357.
9. Nagayama, S.; Homma, R.; Imamura, F., Neuronal organization of olfactory bulb circuits. *Frontiers in neural circuits* **2014**, 8, 98.
10. Kontaris, I.; East, B. S.; Wilson, D. A., Behavioral and neurobiological convergence of odor, mood and emotion: A review. *Frontiers in Behavioral Neuroscience* **2020**, 14, 35.
11. Heydel, J. M.; Coelho, A.; Thiebaud, N.; Legendre, A.; Bon, A. M. L.; Faure, P.; Neiers, F.; Artur, Y.; Golebiowski, J.; Briand, L., Odorant-binding proteins and xenobiotic metabolizing enzymes: implications in olfactory perireceptor events. *The Anatomical Record* **2013**, 296 (9), 1333-1345.
12. Pelosi, P.; Knoll, W., Odorant-binding proteins of mammals. *Biol Rev Camb Philos Soc* **2022**, 97 (1), 20-44.
13. El Kazzy, M.; Lalis, M.; Hurot, C.; Weerakkody, J. S.; Mathey, R.; Saint-Pierre, C.; Buhot, A.; Livache, T.; Topin, J.; Moitrier, L., Study and optimization of the selectivity of an odorant binding protein-based bioelectronic nose. *Biosensors and Bioelectronics* **2025**, 268, 116879.
14. Moitrier, L.; Belloir, C.; Lalis, M.; Hou, Y.; Topin, J.; Briand, L., Ligand binding properties of Odorant-Binding protein OBP5 from *Mus musculus*. *Biology* **2022**, 12 (1), 2.
15. Lalis, M.; Moitrier, L.; Jäger, M.; Meinert, C.; Brulé, M.; Belloir, C.; Jones, N. C.; Hoffmann, S. V.; Fiorucci, S.; Wolf, S., How allosteric mutations control ligand binding in Lipocalin protein: odorant binding protein as a test case. *Cellular and Molecular Life Sciences* **2025**, 82 (1), 250.
16. Kida, H.; Fukutani, Y.; Mainland, J. D.; de March, C. A.; Vihani, A.; Li, Y. R.; Chi, Q.; Toyama, A.; Liu, L.; Kameda, M., Vapor detection and discrimination with a panel of odorant receptors. *Nature communications* **2018**, 9 (1), 4556.
17. Débat, 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. *Journal of proteome research* **2007**, 6 (5), 1985-1996.
18. Lazard, D.; Zupko, K.; Poria, Y.; Net, P.; Lazarovits, J.; Horn, S.; Khen, M.; Lancet, D., Odorant signal termination by olfactory UDP glucuronosyl transferase. *Nature* **1991**, 349 (6312), 790-793.

19. Maßberg, D.; Hatt, H., Human olfactory receptors: novel cellular functions outside of the nose. *Physiological reviews* **2018**.
20. Sinding, C.; Kemper, E.; Spornraft-Ragaller, P.; Hummel, T., Decreased perception of bourgeonal may be linked to male idiopathic infertility. *Chemical senses* **2013**, 38 (5), 439-445.
21. Lledo, P.-M.; Valley, M., Adult olfactory bulb neurogenesis. *Cold Spring Harbor perspectives in biology* **2016**, 8 (8), a018945.
22. Barnes, I. H.; Ibarra-Soria, X.; Fitzgerald, S.; Gonzalez, J. M.; Davidson, C.; Hardy, M. P.; Manthravadi, D.; Van Gerven, L.; Jorissen, M.; Zeng, Z., Expert curation of the human and mouse olfactory receptor gene repertoires identifies conserved coding regions split across two exons. *BMC genomics* **2020**, 21 (1), 196.
23. Lalis, M.; Hladiš, M.; Abi Khalil, S.; Deroo, C.; Marin, C.; Bensafi, M.; Baldovini, N.; Briand, L.; Fiorucci, S.; Topin, J., A status report on human odorant receptors and their allocated agonists. *Chemical Senses* **2024**, 49, bjae037.
24. Kim, M.; Lee, J.; Park, I.; Kim, J.; Lee, K.; So, J.; Choi, J.-W.; Jang, J. E.; Kwon, H.-J.; Moon, C., Stable olfactory receptor activation across odor complexity. *iScience* **2025**, 28 (11).
25. Ignatieva, E. V.; Levitsky, V. G.; Yudin, N. S.; Moshkin, M. P.; Kolchanov, N. A., Genetic basis of olfactory cognition: extremely high level of DNA sequence polymorphism in promoter regions of the human olfactory receptor genes revealed using the 1000 Genomes Project dataset. *Frontiers in psychology* **2014**, 5, 247.
26. Sell, C. S., On the unpredictability of odor. *Angewandte Chemie International Edition* **2006**, 45 (38), 6254-6261.
27. Lalis, M.; Hladiš, M.; Khalil, S. A.; Briand, L.; Fiorucci, S.; Topin, J., M2OR: a database of olfactory receptor–odorant pairs for understanding the molecular mechanisms of olfaction. *Nucleic Acids Research* **2023**.
28. Hladiš, M.; Lalis, M.; Fiorucci, S.; Topin, J. In *Matching receptor to odorant with protein language and graph neural networks*, The Eleventh International Conference on Learning Representations, 2022.
29. Lee, B. K.; Mayhew, E. J.; Sanchez-Lengeling, B.; Wei, J. N.; Qian, W. W.; Little, K. A.; Andres, M.; Nguyen, B. B.; Moloy, T.; Yasonik, J., A principal odor map unifies diverse tasks in olfactory perception. *Science* **2023**, 381 (6661), 999-1006.
30. Choi, C.; Bae, J.; Kim, S.; Lee, S.; Kang, H.; Kim, J.; Bang, I.; Kim, K.; Huh, W.-K.; Seok, C., Understanding the molecular mechanisms of odorant binding and activation of the human OR52 family. *Nature Communications* **2023**, 14 (1), 8105.

Molecular Dynamics

International Edition: DOI: 10.1002/anie.201713065

German Edition: DOI: 10.1002/ange.201713065

Odorant Receptor 7D4 Activation Dynamics

Claire A. de March, Jérémie Topin, Elise Bruguera, Gleb Novikov, Kentaro Ikegami, Hiroaki Matsunami,* and Jérôme Golebiowski*

Abstract: Deciphering how an odorant activates an odorant receptor (OR) and how changes in specific OR residues affect its responsiveness are central to understanding our sense of smell. A joint approach combining site-directed mutagenesis and functional assays with computational modeling has been used to explore the signaling mechanics of OR7D4. In this OR, a genetic polymorphism affects our perception of androstenone. Molecular simulations totaling 0.12 ms predicted that, similarly to observations for other G-protein-coupled receptors with known experimental structures, an activation pathway connects the ligand and the G-protein binding site. The 3D model activation mechanism correlates with in vitro data and notably predicts that the OR7D4 WM variant is not activated. Upon activation, an OR-specific sequence motif is the convergence point of the mechanism. Our study suggests that robust homology modeling can serve as a powerful tool to capture OR dynamics related to smell perception.

Our sense of smell is triggered by the activation of odorant receptors (ORs) expressed on the surface of olfactory sensory neurons. Smelling relies on a so-called combinatorial code of OR activations.^[1] Our approximately 400 ORs endow us with spectacular discriminatory power. ORs represent more than 3% of our protein-coding genes and belong to the class A G-

protein-coupled receptor (GPCR) family. Structurally, class A GPCRs are made up of seven transmembrane helices (TMs), named TM1 to TM7, which form a bundle within the cell membrane. Upon activation, an allosteric signal is transmitted from the ligand to the G-protein binding site over a distance of more than 20 Å (ca. 5 helix turns).^[2] The intracellular part of TM6 then tilts outward with respect to its position in the inactive state.^[3]

Mammalian OR sequences show a highly conserved and OR-specific RxKAFSTCxSH motif in the intracellular part of TM6, suggesting a crucial role in the control of OR activation. In non-olfactory class A GPCRs, only the second lysine residue is conserved.^[4] Odorant-receptor-activation dynamics still remain to be uncovered, but a significant body of work has focused on the molecular details of OR activation, especially concerning the interaction between ORs and ligands.^[5] The strength of molecular-modeling approaches on GPCRs was demonstrated by a groundbreaking study based on homology modeling and virtual screening, which deorphanized GPR68 and GPR65.^[6] By applying this approach to ORs, we recently demonstrated how certain mutations alter receptor function or dynamics.^[4,7]

Human OR7D4, which is narrowly tuned to detect the testosterone metabolites androstadienone and androstenone, is a spectacular example of how natural sequence variations can affect activation and subsequent odor perception. People with two copies of the most common genetic variant (OR7D4-RT or wild type, wt) tend to describe the androstenone smell as sickening, putrid, foul, or urine-like. However, people who possess the unresponsive R88W/T133M variant (OR7D4-WM) tend to report the smell to be more pleasant.^[8] This polymorphism correlates with a differential preference for pork containing varying amounts of androstenone,^[9] which has a consequent effect on the percentage of castrated pigs in European countries.^[10]

To evaluate how an OR sequence affects its activation dynamics, we have used a joint approach combining functional assays with molecular dynamics (MD) simulations. Mouse and human genomes account for approximately 1500 ORs, making the OR family suited to the building of robust homology models. In this study, a 3D model of OR7D4 was built following a previously published protocol.^[4]

The sequence alignment was performed by comparing residue conservation between the ORs and crystallized class A GPCRs. Manual adjustments were made to recover constraints imposed by the 141 in vitro data points that were tested by site-directed mutagenesis in various ORs, including OR7D4.^[4,7,11] In summary, they represent 96 distinct positions within the sequence of OR7D4 (31% of the 312 residues; see the Supporting Information). They

[*] Dr. C. A. de March,^[**] Dr. J. Topin,^[**] Dr. G. Novikov, Prof. J. Golebiowski

Institute of Chemistry—Nice, UMR 7272 CNRS

Université Côte d'Azur

06108 Nice cedex (France)

E-mail: jerome.golebiowski@unice.fr

Dr. C. A. de March,^[**] E. Bruguera, K. Ikegami, Prof. H. Matsunami

Department of Molecular Genetics and Microbiology

Duke University Medical Center

Durham, NC 27710 (USA)

E-mail: hiroaki.matsunami@duke.edu

K. Ikegami

Department of Biotechnology and Life Science

Tokyo University of Agriculture and Technology Tokyo (Japan)

Prof. H. Matsunami

Department of Neurobiology, Duke Institute for Brain Sciences

Duke University Medical Center, Durham, NC 27710 (USA)

Prof. J. Golebiowski

Department of Brain and Cognitive Sciences, DGIST

333, Techno JungAng Daero, HyeongPoong Myeon Daegu, 711-873

(Republic of Korea)

[**] Joint first authors

Supporting information, including experimental details of conservation analysis, molecular modeling, and functional assays, and the ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/anie.201713065>.

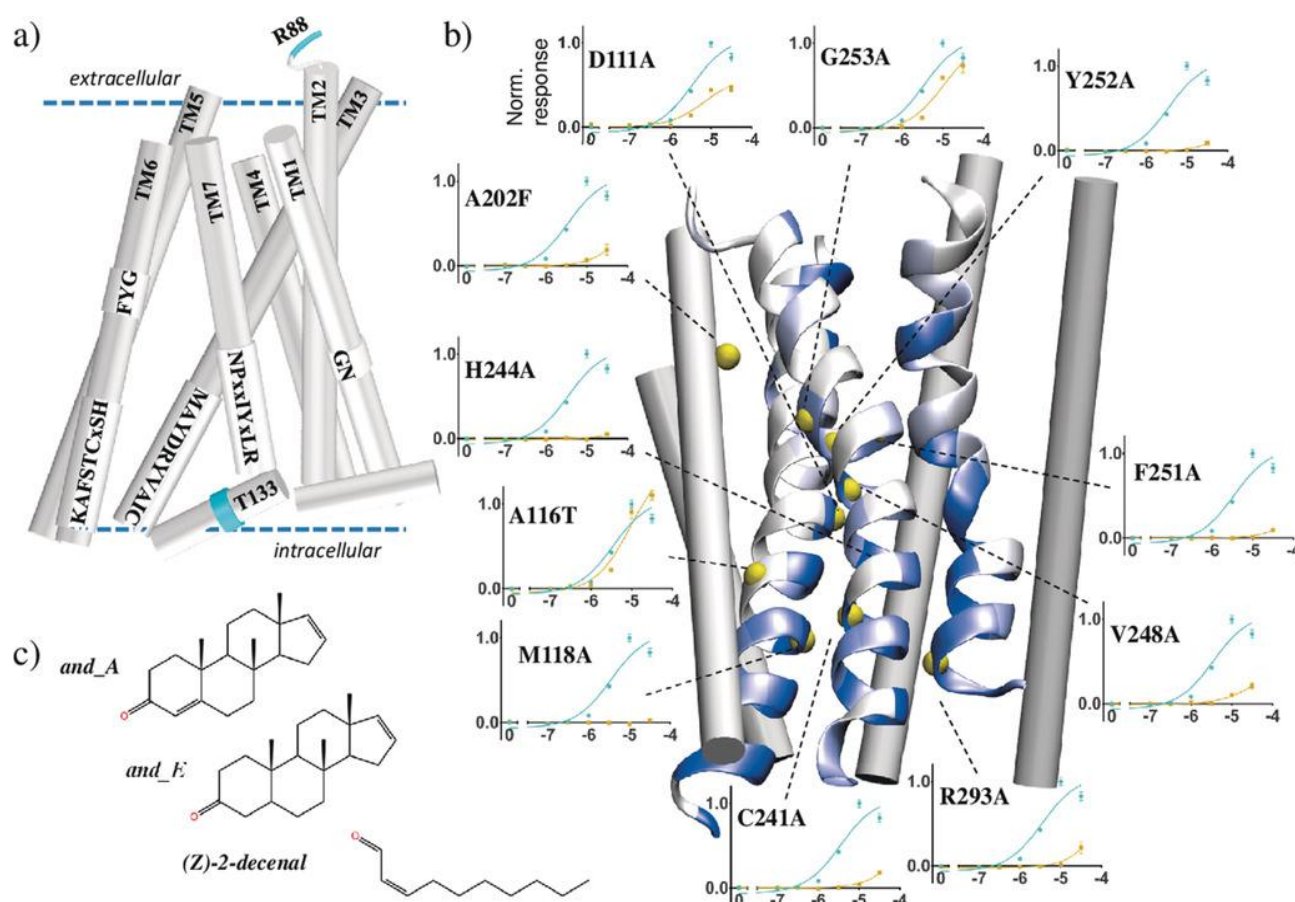


Figure 8. OR7D4 3D model with the loops omitted for clarity. a) Location of the conserved sequence motifs as predicted by the model. Residues R88 and T133 are highlighted in cyan. b) 3D model of OR7D4 with TM3, TM6, and TM7 shown as helices with residues colored according to conservation (from white (low) to blue (high)). Yellow spheres indicate the position of residues for which androstadienone dose–response curves were obtained. For each position, the mutant OR (yellow curve) is compared to the wt (blue curve). Normalized responses are shown as means and standard error of the mean (s.e.m.; $n=3$). Stimulation with androstenone induce receptor responses similar to that of androstadienone (see the Supporting Information). c) Chemical structures of androstadienone (and_A), androstenone (and_E), and (Z)-2-decenal (decenal).

cover the seven TMs as well as the orthosteric binding site and unambiguously confirmed our sequence alignment.

More specifically, in this study we performed 27 site directed mutations on 16 sites to assess the final 3D model of OR7D4. Figure 1 highlights the positions of conserved motifs and alanine-scanning functional assays. Nine mutations interrogated the binding cavity, motifs putatively involved in the transmission of the signal, and segments in contact with the G protein. In mammalian ORs, higher sequence-specific conservation suggests the existence of points of contact on the faces of the TM3, TM6, and TM7 helices. Our model demonstrated that most of these conserved residues are facing each other. Consistently, the responses of the corresponding mutant ORs are strongly affected (Figure 1b). The two control A116T^{3,44} and G253A^{6,49} mutant ORs (see Ref. [12] for numbering), whose side chains are not involved in any TM contact, showed an unaffected response. Neither the wt nor the mutant ORs showed a luciferase response upon decenal stimulation (see the Supporting Information).

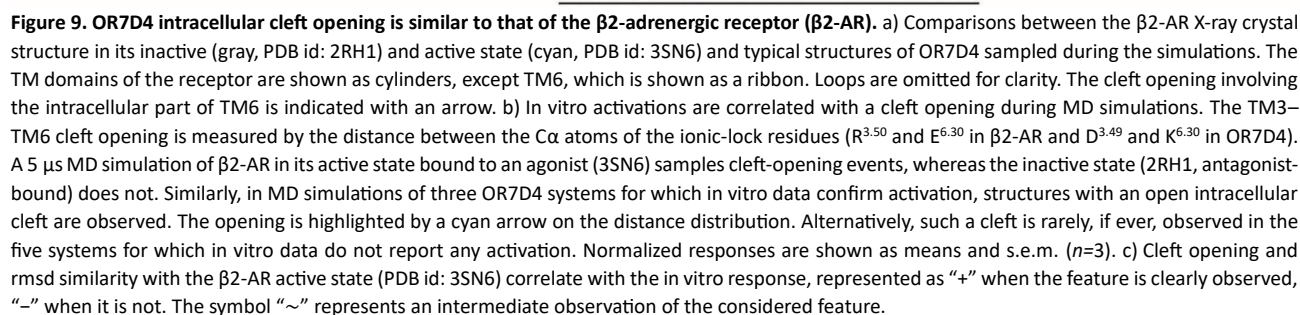
The molecules and_A, and_E, and decenal were docked to the apo receptor prior to microsecond-timescale MD simulations. As expected, the model predicted a much larger binding energy of the receptor for the two agonists than for the control (see the Supporting Information), consistent with the notion that agonists must reach a certain affinity to trigger OR activation.^[13]

From a structural point of view, the binding cavity is made up of residues belonging to TM3, TM5, TM6, and TM7, consistent with those identified in GPCRs with a known experimental structure.^[14] Considering the ORs specifically, the key residues forming the binding cavity are consistent with those identified by previous modeling approaches.^[4,5,7] In the OR7D4 model, the A202^{5,42} methyl side chain points towards the ligand and allows for the binding of the bulky agonists. Accordingly, the A202^{5,42}F mutant OR is unresponsive to agonists in vitro.

The model also predicts that A279^{7,42} points towards the binding cavity. Its mutation to a bulkier and hydrophilic residue (A279D) was reported to abolish the in vitro response of the OR upon agonist stimulation.^[8]

Starting from the same initial model structure of OR7D4, three unconstrained MD simulations of approximately 5 ms were performed for each of these eight systems. In total, OR7D4 dynamics were investigated for 0.12 ms and compared to MD simulations of the β 2-adrenergic receptor (β 2AR) bound to an agonist (PDB id: 3SN6) or an antagonist (PDB id: 2RH1). The models provided mechanistic hypotheses that were further tested in vitro.

Interestingly, for experimentally activated systems, back and forth movements involving TM6 and the intracellular loop 3 (ICL3) opened a cleft in the intracellular domain of the receptor.



From a mechanistic point of view, the model recovers the typical features of the activation of classA GPCRs (Figure 3a). Upon activation, residue Y^{6.48} shifts towards P^{5.50}, consistent with data reported on the muscarinic 2 or the cannabinoid 1

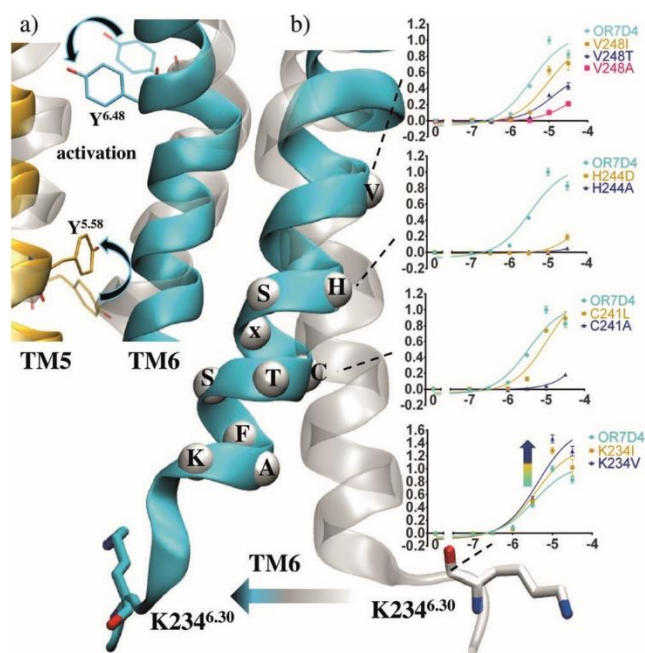


Figure 3. a) The mechanism of activation involves reorganization of TM5 and TM6 residues Y6.48 and Y5.58. b) and_A dose-response curves involving the RxKAFSTCxSHTM6 motif residues. Normalized responses are shown as means and s.e.m. ($n=3$). Comparison between the open (cyan) and closed conformations (transparent) of OR7D4.

receptors.^[16] Also, opening of the TM6 cleft is accompanied by an inward shift of Y^{5.58}, as typically reported for β_2 -AR.^[17] More generally, a series of complex allosteric movements propagate the activation signal through a dynamic network involving the whole protein structure comprising the loops (see the Supporting Information).

A comparison between inactive and active structures revealed differential receptor dynamics involving the OR-specific RxKAFSTCxSH^{TM6} motif. As emphasized in Figure 3, this motif forms a rigid helical segment, whose conformational change is associated with activation. Similar structural behavior is observed in β_2 -AR and more generally in class A GPCRs, which show a conserved lysine residue (K^{6.30}) that is involved in the canonical ionic lock at the junction between TM6 and ICL3.^[16a,18] In luciferase assays probing OR7D4 mutants with a modified intracellular TM6, mutations introducing a hydrophobic residue at position K234^{6.30} are likely to prevent the formation of the ionic lock between TM6 and TM3.^[7] Consistently, the K234I or V mutant ORs are more responsive than the wt (Figure 3b).

Moving towards the extracellular end of this conserved motif, H244A and H244D or C241A mutations abolish the response to agonists, which is consistent with a conservation rate of 98% in human OR sequences. Away from the RxKAFSTCxSH motif, mutations are less deleterious to the receptor function. Mutations at position 248^{6.44} differentially modulate the responsiveness of the receptor, thus suggesting a role in communicating the signal from the binding pocket to the convergence point represented by this TM6 motif.

In this study, a robust homology model of OR7D4 was analyzed using MD simulations of its apo form or bound form with androstenedione and androstadienone, two known agonists, and also in the presence of (Z)-2-decenal as a negative control. The

intriguing WM variant associated with a differential perception of boar taint was also studied together with two additional control mutants. The model of OR7D4 is consistent with conservation analyses and in vitro functional assays on 27 mutant ORs obtained by site-directed mutagenesis at 16 different sites. In total, the dynamics of OR7D4 variants were explored for approximately 0.12 ms. The model more frequently explored structures showing the typical features of an active (open) state, when in vitro data assessed receptor activation.

Conversely, when considered in its apo form, bound to a non-agonist or mutated at deleterious sites (WM, C241A, H244A), the receptor was mostly modeled in a closed (inactive) state. Thus, starting from this single model of the receptor, we were able to discriminate agonists of the human OR7D4 from non-agonists by monitoring the allosteric movement of the TM3–TM6 interhelical cleft, the metric most representative of a GPCR preorganized for recruiting the G protein. A total of 24 independent MD simulations were run, 18 of which were consistent with in vitro data, thus leading to a predictive power of 75%. For each system studied, at worst, only one MD result out of three led to either a false positive (apo-wt, C241A, and WM) or a false negative (and_A, and_E, and D111A), which highlights the importance of enhanced sampling through multiple MDs to prevent erroneous conclusions on the nature of the bound ligand.

Shifts between open and closed conformations involve rearrangements of the mammalian OR-specific motif within TM6, with respect to TM3. This RxKAFSTCxSH motif thus plays a pivotal role in activation, consistent with its high conservation across mammalian ORs.^[4] The position of this highly conserved motif emphasizes its importance in the control of both OR activation and specificity for the olfactory G protein (G_{olf}).

In the WM variant bound to androstadienone, although mutations are far removed from the orthosteric binding site and the G-protein coupling area, the receptor does not show any intracellular cleft opening, thus highlighting the crucial role of residues that do not belong to the binding cavity in the dynamics of GPCRs in general. Similarly, genetic analysis of “super smellers” of androstenedione suggested a link with a single nucleotide polymorphism in TM2 of OR7D4 (S84N). This higher detection threshold is related to the mutant OR, which is more responsive to the agonists,^[8] and emphasizes the extreme complexity of GPCR signaling mechanisms. Subtle sequence modifications affect OR activation and, in this case, can spectacularly impact smell perception and food preferences.

Acknowledgements

We thank Dr. T. Abaffy and Dr. X. Cong for critical reading of the manuscript and M. J. Ni for expert technical assistance. This research was supported by grants from the National Institute on Deafness and Other Communication Disorders, National Institute of Health grants DC014423 and DC012095, National Science Foundation (NSF) grants 1515801 (to H.M.), the Agence Nationale de la Recherche as part of NSF/NIH/ANR Collaborative Research in Computational Neuroscience, a UCA IDEX grant (to J.G.), and a grant from the city of Nice (to J.T.).

Conflict of interest

The authors declare no conflict of interest.

Keywords: activation · functional assays · molecular dynamics · molecular modeling · olfactory receptors

How to cite: Angew. Chem. Int. Ed. 2018, 57, 4554–4558
Angew. Chem. 2018, 130, 4644–4648

-
- [1] a) L. Buck, R. Axel, *Cell* 1991, 65, 175–187; b) B. Malnic, P. A. Godfrey, L. B. Buck, *Proc. Natl. Acad. Sci. USA* 2004, 101, 2584–2589.
- [2] D. M. Rosenbaum, et al., *Science* 2007, 318, 1266–1273. [3] a) U. Gether, B. K. Kobilka, *J. Biol. Chem.* 1998, 273, 17979–17982; b) C. Altenbach, A. K. Kusnetzow, O. P. Ernst, K. P. Hofmann, W. L. Hubbell, *Proc. Natl. Acad. Sci. USA* 2008, 105, 7439–7444; c) A. Manglik, et al., *Cell* 2015, 161, 1101–1111.
- [4] C. A. de March, S.-K. Kim, S. Antonczak, W. A. Goddard, J. Golebiowski, *Protein Sci.* 2015, 24, 1543–1548.
- [5] a) C. Geithe, J. Protze, F. Kreuchwig, G. Krause, D. Krautwurst, *Cell. Mol. Life Sci.* 2017, 74, 4209–4229; b) S. Li, L. Ahmed, R. Zhang, Y. Pan, H. Matsunami, J. L. Burger, E. Block, V. S. Batista, H. Zhuang, *J. Am. Chem. Soc.* 2016, 138, 13281–13288; c) S. Wolf, N. Jovancevic, L. Gelis, S. Pietsch, H. Hatt, K. Gerwert, *Sci. Rep.* 2017, 7, 16007; d) O. Baud, S. Yuan, L. Veya, S. Filipek, H. Vogel, H. Pick, *Sci. Rep.* 2015, 5, 14948.
- [6] X. P. Huang, et al., *Nature* 2015, 527, 477–483.
- [7] C. A. de March, Y. Yu, M. J. Ni, K. A. Adipietro, H. Matsunami, M. Ma, J. Golebiowski, *J. Am. Chem. Soc.* 2015, 137, 8611–8616.
- [8] A. Keller, H. Zhuang, Q. Chi, L. B. Vosshall, H. Matsunami, *Nature* 2007, 449, 468–472.
- [9] K. Lunde, B. Egelanddsdal, E. Skuterud, J. D. Mainland, T. Lea, M. Hersleth, H. Matsunami, *PLoS ONE* 2012, 7, e35259.
- [10] M. Blanch, N. Panella-Riera, P. Chevillon, M. Furnols, M. Gil, J. Gil, Z. Kallas, M. Oliver, *Meat Sci.* 2012, 90, 572–578.
- [11] Y. Yu, C. A. de March, M. J. Ni, K. A. Adipietro, J. Golebiowski, H. Matsunami, M. Ma, *Proc. Natl. Acad. Sci. USA* 2015, 112, 14966–14971.
- [12] J. A. Ballesteros, H. Weinstein in *Methods in Neurosciences*, Vol. 25 (Ed.: C. S. Stuart), Academic Press, New York, 1995, pp. 366–428.
- [13] J. Topin, C. A. de March, L. Charlier, C. Ronin, S. Antonczak, J. Golebiowski, *Chem. Eur. J.* 2014, 20, 10227–10230.
- [14] a) A. J. Venkatakrishnan, et al., *Nature* 2016, 536, 484–487; b) N. R. Latorraca, A. J. Venkatakrishnan, R. O. Dror, *Chem. Rev.* 2017, 117, 139–155.
- [15] X. Deupi, J. Standfuss, *Curr. Opin. Struct. Biol.* 2011, 21, 541–551.
- [16] a) Y. Miao, S. E. Nichols, P. M. Gasper, V. T. Metzger, J. A. McCammon, *Proc. Natl. Acad. Sci. USA* 2013, 110, 10982–10987; b) T. Hua, et al., *Cell* 2016, 167, 750–762, e714.
- [17] a) R. O. Dror, D. H. Arlow, P. Maragakis, T. J. Mildorf, A. C. Pan, H. Xu, D. W. Borhani, D. E. Shaw, *Proc. Natl. Acad. Sci. USA* 2011, 108, 18684–18689; b) R. Nygaard, et al., *Cell* 2013, 152, 532–542.
- [18] J. Shan, G. Khelashvili, S. Mondal, E. L. Mehler, H. Weinstein, *PLoS Comput. Biol.* 2012, 8, e1002473.

Manuscript received: December 19, 2017

Accepted manuscript online: February 20, 2018

Version of record online: March 24, 2018

Experimental details

Site-Directed Mutagenesis

The coding sequence of human OR7D4 was cloned into the pCI vector (Promega) with an N-terminal tag of the first 20 amino acids of rhodopsin. Site-directed mutants were constructed using overlap extension PCR. The sequences of all plasmid constructions were verified by both forward and reverse sequencing using the BigDye Terminator Sequencing Kit (Applied Biosystems) and a 3130 Genetic Analyzer (Applied Biosystems).

Evaluation of OR surface expression

Fluorescent Activated Cell Sorting (FACS) was conducted to evaluate cell surface expression of OR mutants as previously described.^[1] Hana3A were seeded in a 35mm dish (Corning) in Minimum Essential Medium containing 10% FBS (M10). Lipofectamine2000 (Invitrogen) was used for transfection of plasmid DNA. At the time of transfection, green fluorescent protein (GFP) expression vector was co-transfected to monitor the transfection efficiency. About 24 hrs post-transfection, cells were incubated with PBS containing anti Rho-tag antibody 4D2 (gift from R. Molday), 15mM NaN₃, and 2% FBS and then washed and incubated with phycoerythrin (PE)-conjugated donkey anti-mouse IgG (Jackson Immunologicals). 7-amino-actinomycin D (7-AAD; Calbiochem), a fluorescent, cell-impermeant DNA binding agent, was added before flow cytometry to eliminate dead cells from analysis. 7-AAD selectively stains dead cells. The intensity of PE signal among the GFP-positive population was measured and plotted (Figure S1).

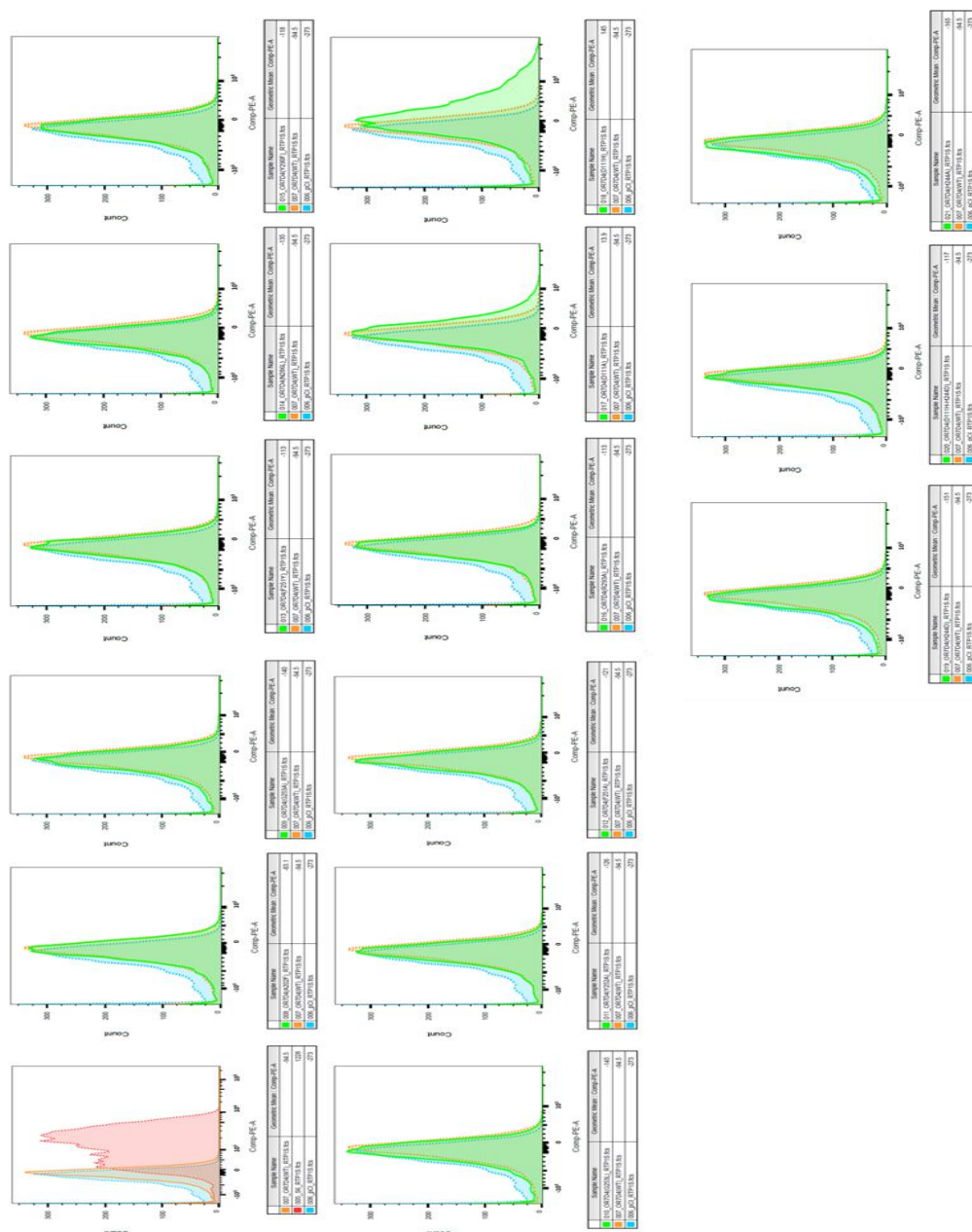


Figure S1. FACS analysis of expression levels of wt and mutants of OR7D4. The negative control for the comparison of the expression level is pCI (blue).

Luciferase assay in Hana3A cells

The Dual-Glo Luciferase Assay (Promega) was used to determine the activities of firefly and Renilla luciferase in Hana3A cells as previously described.^[2] Firefly luciferase, driven by a cAMP response element promoter (CRE-Luc; Stratagene), was used to determine OR activation levels. Renilla luciferase, driven by a constitutively active SV40 promoter (pRLSV40; Promega), functioned as an internal control for transfection efficiency and cell viability. Hana3A cells stably expressing RTP1L, RTP2, REEP1, and Golf were plated on poly-D-lysine-coated 96-well plates (Nalge Nunc) and incubated overnight in minimum essential medium eagle (Sigma) with 10% FBS at 37°C and 5% CO₂. The following day, cells were transfected using Lipofectamine 2000 (Invitrogen). For each well of a 96- well plate, 5 ng SV40-RL, 10 ng CRE-Luc, 5 ng mouse RTP1s,^[3] 2.5 ng M3 receptor^[4] and 5 ng of Rho- tagged receptor plasmid DNA were transfected. After transfection (24 h), medium was replaced with 25 µl of odorant solution diluted in CD293 chemically defined medium (Invitrogen), and cells were further incubated for 3.5 h at 37°C and 5% CO₂. The manufacturer's protocols were followed to measure firefly luciferase and Renilla luciferase activities. A POLARstar OPTIMA plate reader (BMG Labtech) was used to measure luminescence. Data were analyzed using Microsoft Excel and GraphPad Prism. Normalized activity for each well was further calculated as (Luc-400)/(RLuc-400) where Luc = luminescence of firefly luciferase and RLuc = Renilla luminescence. To facilitate comparison between OR responses from multiple plates, the Rho-tag empty vector and *wt* OR7D4 were always included as negative and positive control, respectively. The basal activity of an OR was averaged from six wells in the absence of odorants and further corrected by subtracting that of the control empty vector. An odorant-induced activity was averaged from at least three wells and further corrected by subtracting the basal activity of that receptor. Odorant-induced responses were normalized to that of *wt*.

Molecular modeling methods

sequence alignment

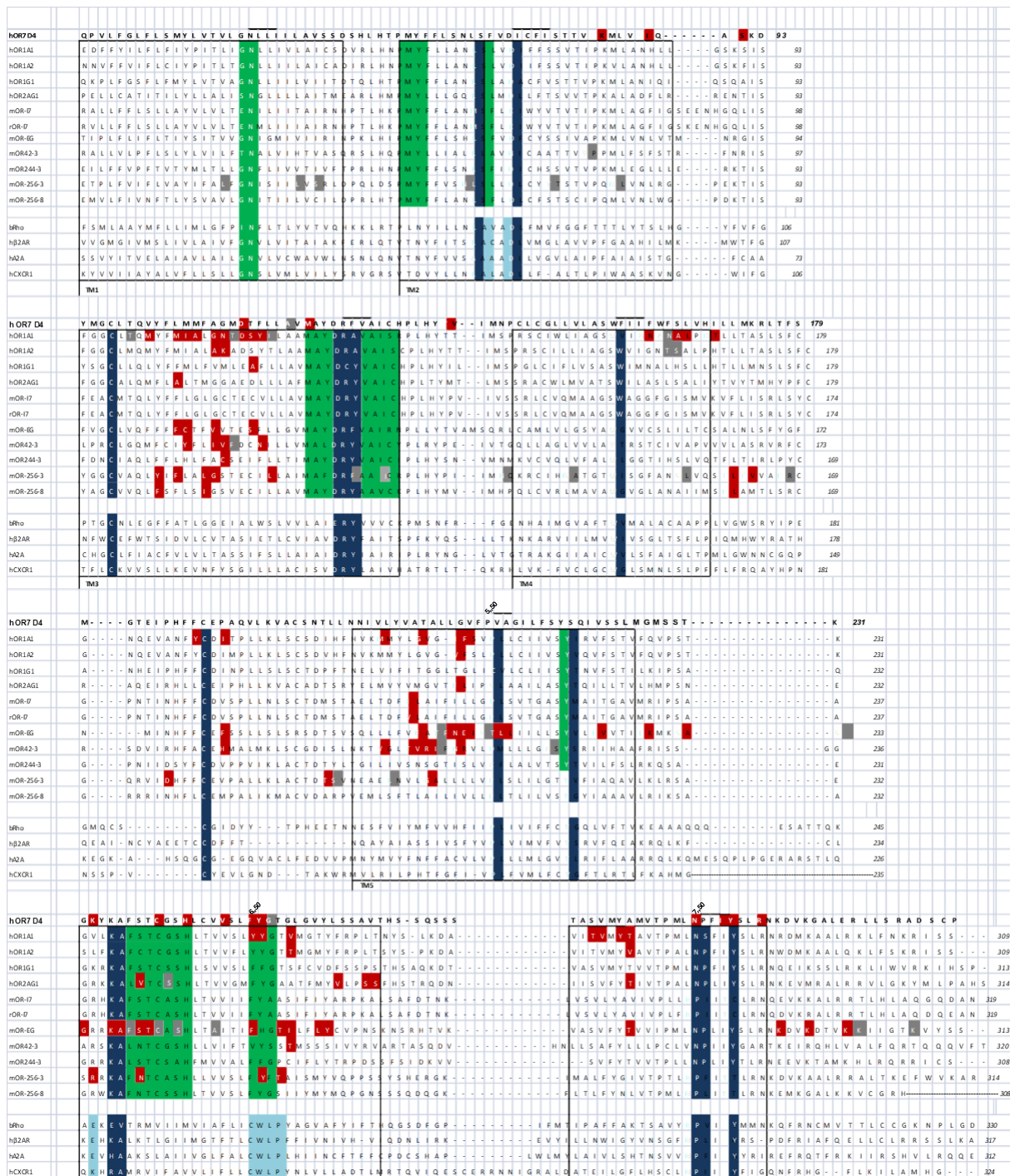


Figure S2. Sequence alignment between human ORs and four GPCRs for which X-ray structures are available. Only ORs for which site-directed mutagenesis data are documented are shown.^[4, 5] Residues for which in vitro site-directed mutagenesis data are available are shown in red when the mutation affects receptor responsiveness. Otherwise, the residue is highlighted in gray. Sequence motifs conserved amongst ORs are shown in green, those conserved amongst all class A GPCRs are highlighted in dark blue. Residues not found in ORs but in non-olfactory GPCRs are shown in light blue. TMs are boxed and central TM residues according to the Ballesteros-Weinstein notation are shown.

Model building

Human olfactory receptor 7D4

The protocol follows a previously published method with an optimal alignment which fulfills constraints provided by more than 100 mutants from the literature and from us (figure S2).^[6] Four experimental GPCR structures (1U19, 3ODU, 2YDV and 2LNL) are selected as templates to build OR7D4 by homology modeling using Modeller.^[7]

The percentage of identity for the seven transmembranes domain of hOR7D4 with respect to the four templates are as follows: 17.3% with Rhodospin (1U19), 17.8% with CxCR4 (3ODU), 21% with A2a (2YDV), and 21% with CxCR1 (2LNL).

The N-terminal structure is omitted to avoid perturbing the modeling protocol. Five models are obtained and we kept the one that was consistent with several additional structural constraints (no large folded structure in extra-cellular loops should be observed, all TMs and H8 folded as α -helices, a short α -helix structure between TM3 and TM4). Finally, only 6 amino acids show backbone dihedral angle out of the energetically allowed region ([Figure S3](#)).

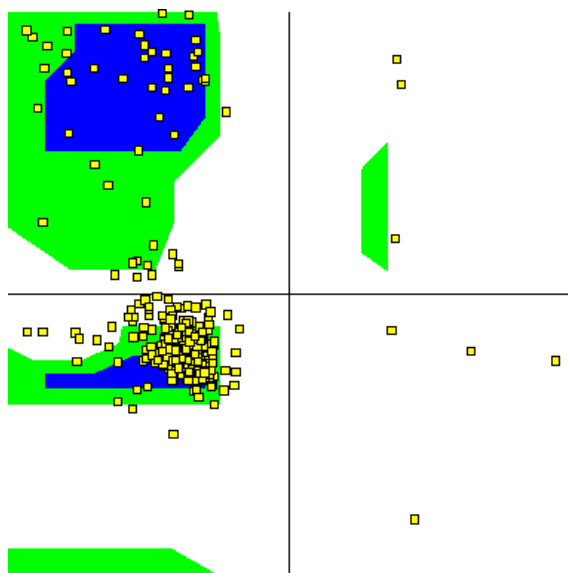


Figure S3. Ramachandran plot of the best model, generated using VMD.

Androstadienone, androstenone and (Z)-2-decenal are prepared with the antechamber module of AMBER with AM1-BCC charges. They are docked into the receptor cavity with Autodock Vina with an exhaustiveness parameter of 50. An exhaustiveness of 15 produced the same resulting poses.^[8]

All agonists poses were similar and we considered the one with the lowest binding free energy for simulations. For (Z)-2-decenal, more variations in the ligand orientation were observed and only the best pose was considered. The initial structures of the three complexes are presented in [Figure S4](#).

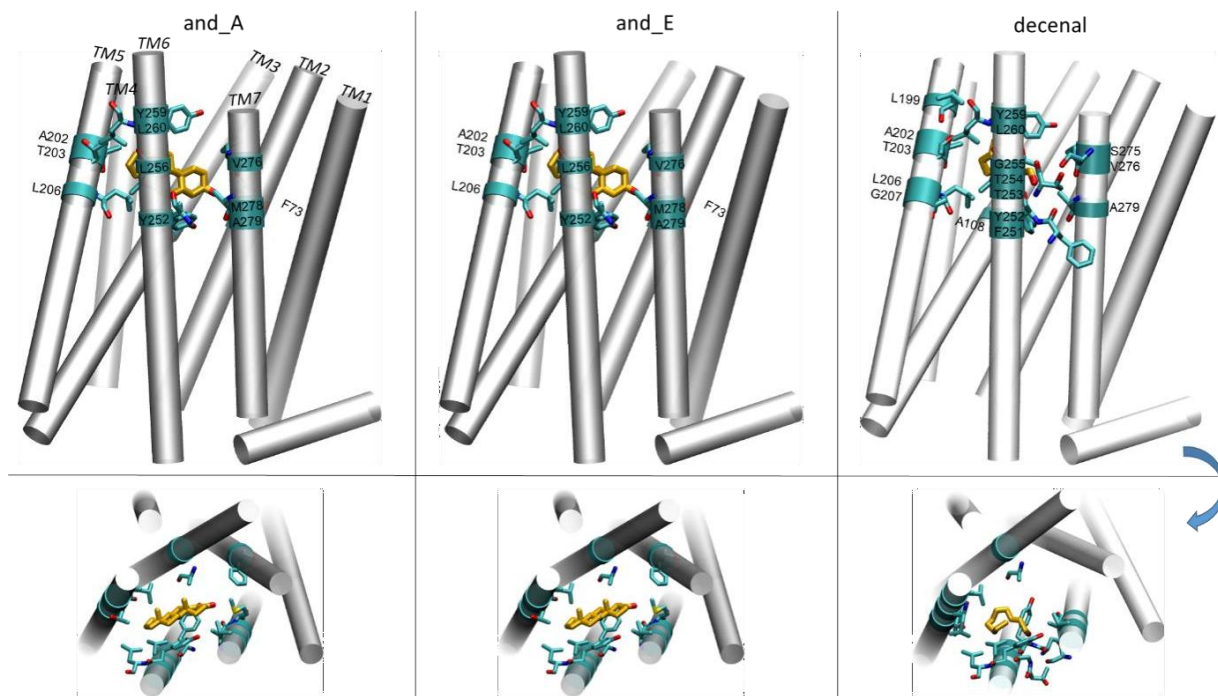


Figure S4. Comparison of the initial docking pose obtained from Autodock VINA. The ligands' carbon atoms are colored in yellow while those of the amino acids are colored in cyan. During the MD, decenal exhibited large reorientations within the cavity, consistent with a low affinity.

β2 adrenergic receptor

The systems were built using PDB ids 2RH1^[9] for the inactive state and 3SN6^[10] for the active state. All residues not belonging to the receptor sequence were removed. The protonation state at pH 7 of each amino acid is optimized using both H⁺ server and ProPka.^[11] Accordingly, all histidine residues are neutral and protonated at the Ne atom, except residue HIS172 which is considered doubly protonated. Aspartic and glutamic acid residues are modeled in their charged state except for residue GLU122 which is neutral.^[12] N-term and C-term of the receptor were capped. Ligands geometry is optimized with Gaussian G03 (B3LYP/6-31G*) and electrostatics is calculated at the HF/6-31G* level. The charges were derived from RESP calculations using the Antechamber module of AMBER.

Molecular dynamics protocol details

The orientation of all systems is optimized using OPM server.^[13] They are then embedded in a model membrane made-up of POPC lipids solvated by TIP3P water molecules using Maestro.^[14] The total system is made up of ~48,650 atoms in a periodic box of 91*89*98 Å³ for hOR7D4. It is made up of ~36,000 atoms in a periodic box of 70*65*86 Å³ for the β 2 adrenergic receptor.

For all systems in this article, molecular dynamics simulations are performed with sander and pmemd.cuda modules of AMBER with the ff03 force-field for the protein the gaff.lipid force-field for the membrane, and the gaff force-field for the ligands.^[15] Bonds involving hydrogen atom are constrained using SHAKE algorithm. Long-range electrostatics interactions are handled with the Particle Mesh Ewald (PME) method. The cutoff for non-bonded interactions is set to 8 Å. Temperature is maintained constant with a Langevin thermostat with a collision frequency of 2 ps⁻¹. In addition, a weak coupling anisotropic algorithm with a relaxation time of 1 ps⁻¹ is applied to keep a constant pressure. Snapshots are saved every 20 ps. The detailed workflow of the simulations is detailed in [Figure S5](#).

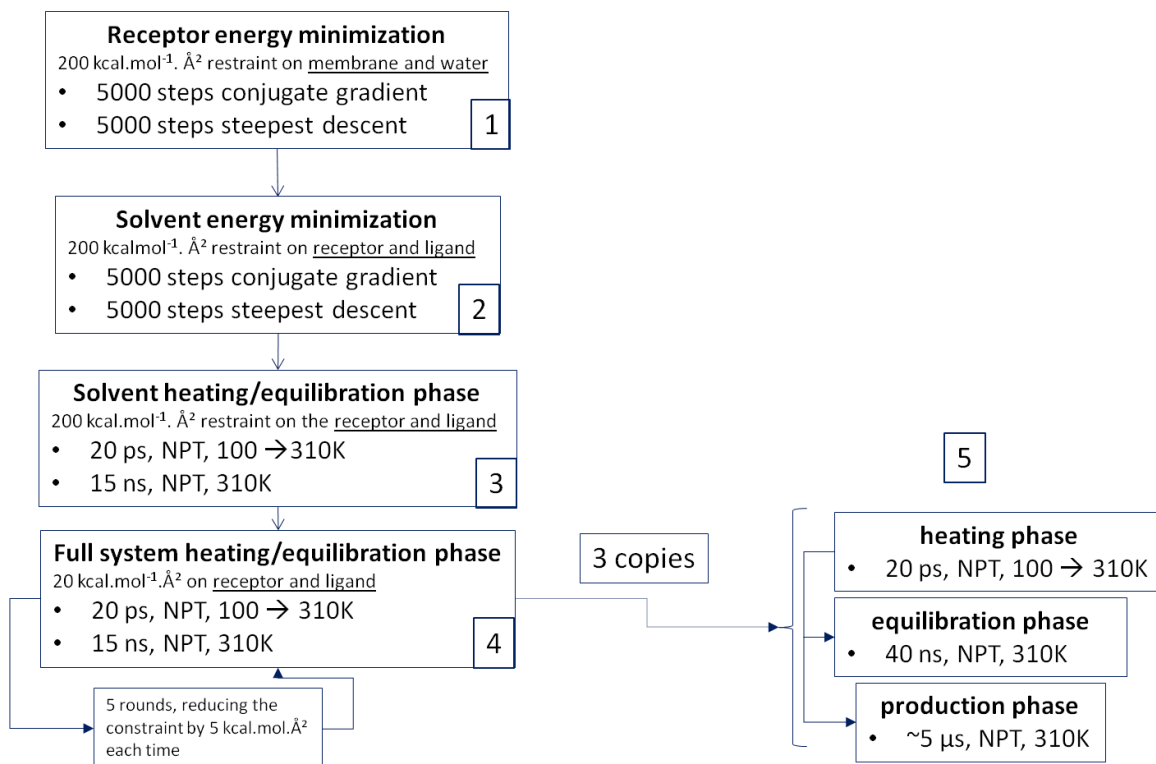


Figure S5. Molecular dynamics workflow.

Dynamic Network Analysis

A protocol similar to that described by Sethi *et al.* is used to build the community network.^[16] All analyses are performed with the bio3D package.^[17]

The protein segments with correlated atomic motion are identified by calculating the cross-correlation matrix for the displacement of all C α atom pairs. Atoms with a C α -C α distance smaller than 7.5 Å for at least 65% of the time are used to define the contact nodes. The community network is built taking into account the cross-correlation matrix (with a cut-off set to 0.35) and the contact nodes. The full network is aggregated into communities of highly intra-connected but loosely inter-connected nodes using the Girvan-Newman betweenness clustering method. For each systems, the modularity value is found to be ≈ 0.7 (which is in the optimal range for modularity score).

Supporting data

Monitoring receptor activation by functional assays in heterologous cells

Responses to (Z)-2-decenal stimulation

The responses of wild type receptor and all the constructed mutants are gathered in [Figure S6](#). In no case the receptor is responding to (Z)-2-decenal stimulation.

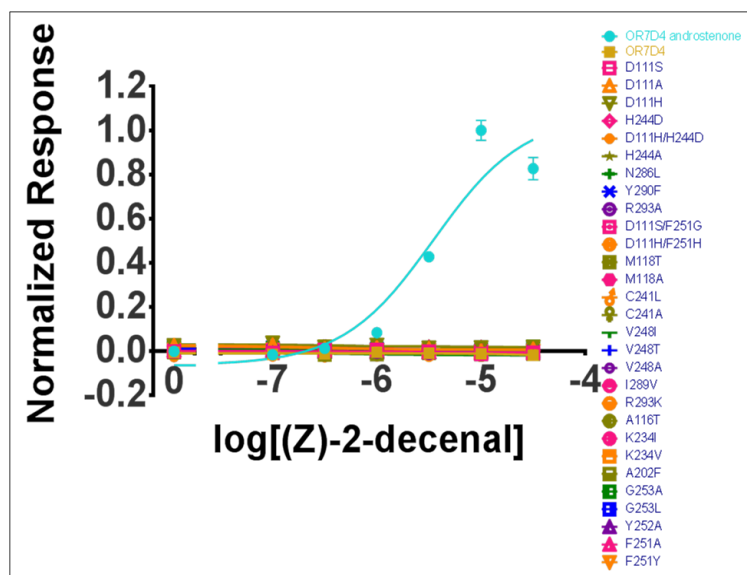


Figure S6. Dose-response of wt and mutants of OR7D4 to (Z)-2-decenal.

Responses to androstadienone stimulation

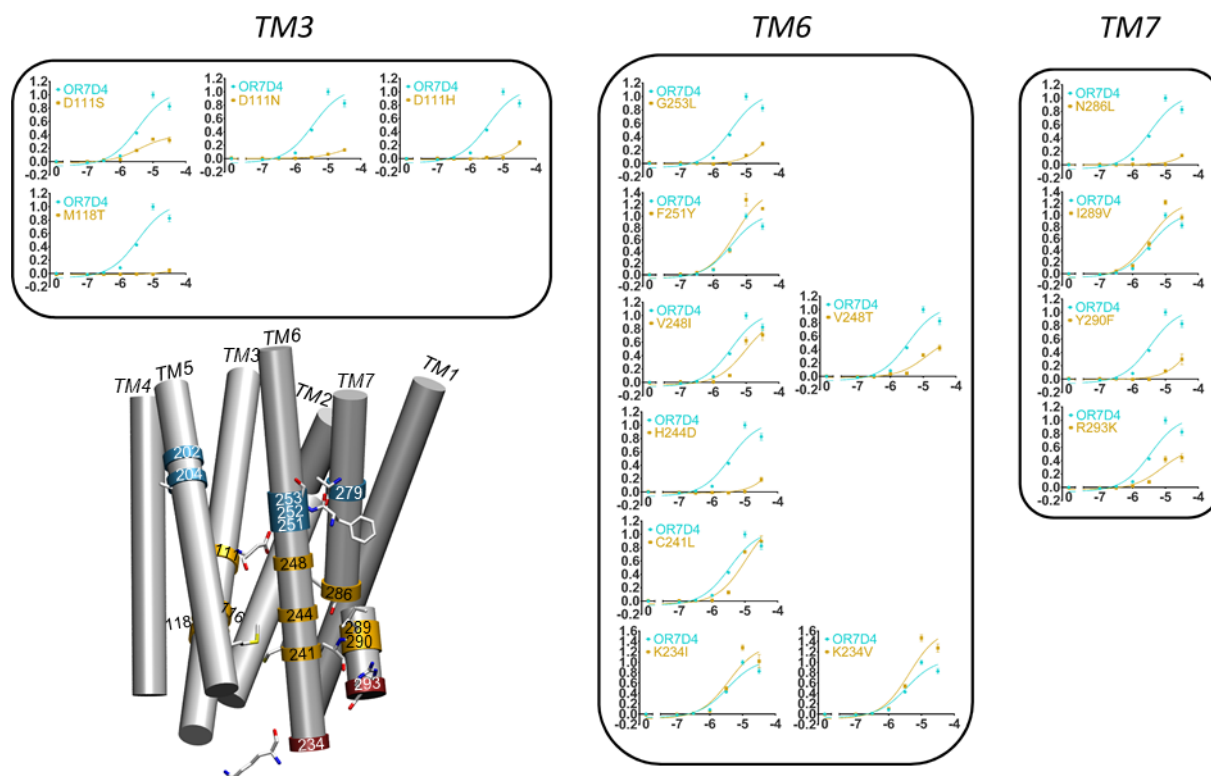


Figure S7. Dose response curves for several mutants from conserved motifs found in TM3, TM6 and TM7 upon androstadienone stimulation.

Responses to androstenone stimulation

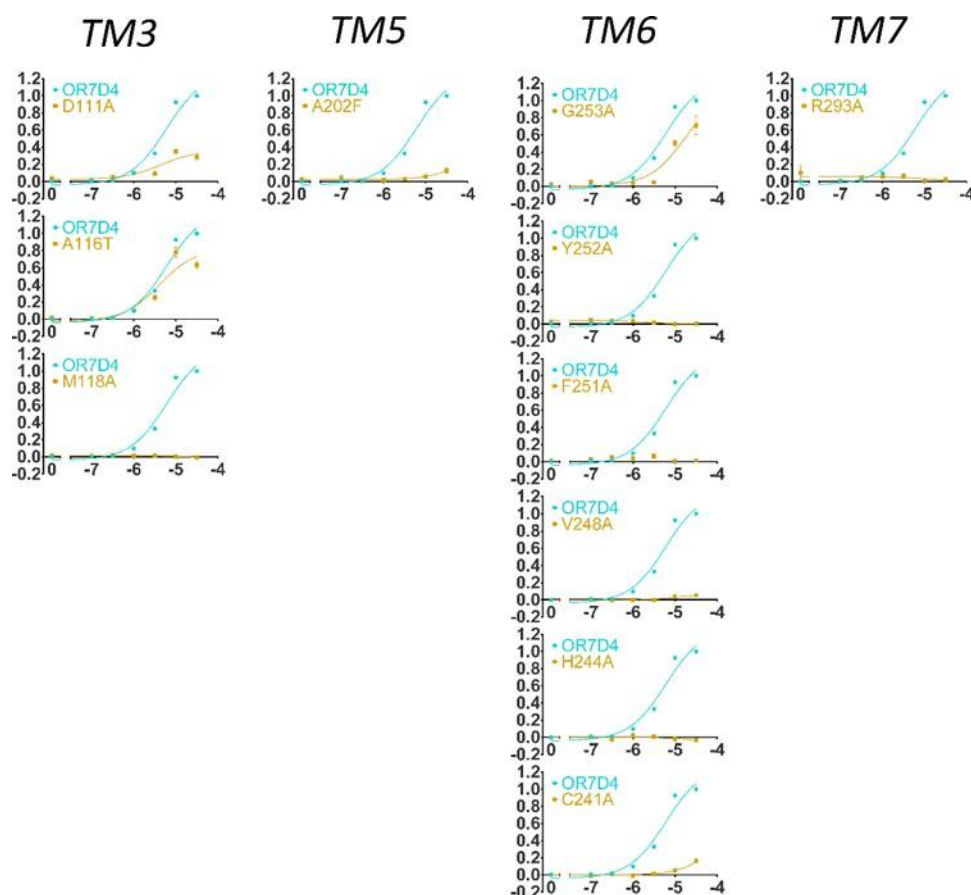


Figure S8. Dose-response of *wt* and ala scanning mutants of OR7D4 to Androstenone. The 11 mutation sites cover the binding cavity, residues putatively involved in the transmission of the signal and in the contact with the G protein. The control A116T^{3,44} and G253A^{6,49} mutant ORs are unaffected since their side chains do not point towards the other TMs.

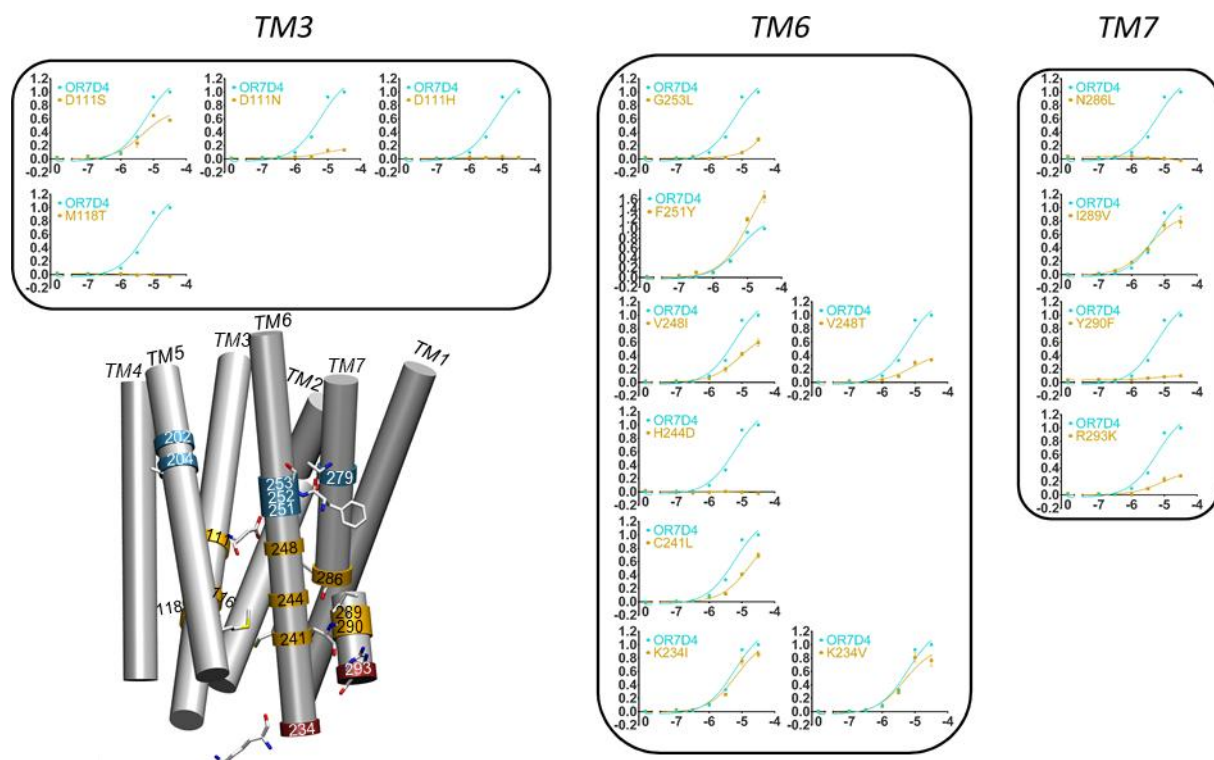


Figure S9. Dose response curves for several mutants from conserved motifs found in TM3, TM6 and TM7 upon androstenone stimulation.

Molecular modeling simulations analyses

Stability of the system

The system stability is evaluated from the root mean square deviation (rmsd) evolution of the backbone of transmembrane helices. For each system, the rmsd is computed using cpptraj module from AMBER package. The first frame of the production phase is taken as reference unless specified.

Figure S10 gathers the rmsd evolution for the two β 2-AR systems. In both cases, the transmembrane part of the receptor undergoes small fluctuations (rmsd < 3 Å). This small deviation shows that the two systems remain correctly folded during microsecond simulations.

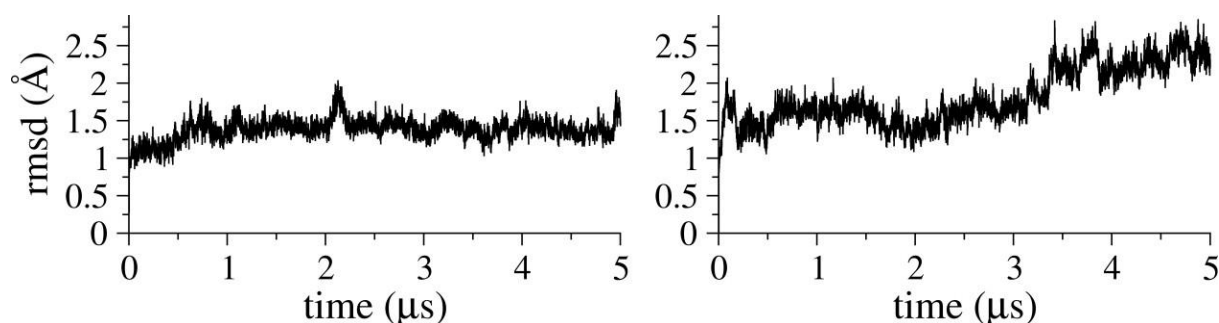


Figure S10. β 2-AR backbone rmsd (in Å) for the agonist bound system (left) and the antagonist bound (right). The rmsd is computed for the receptor bundle backbone (CA, C, N) with respect to the first frame of the MD production phase.

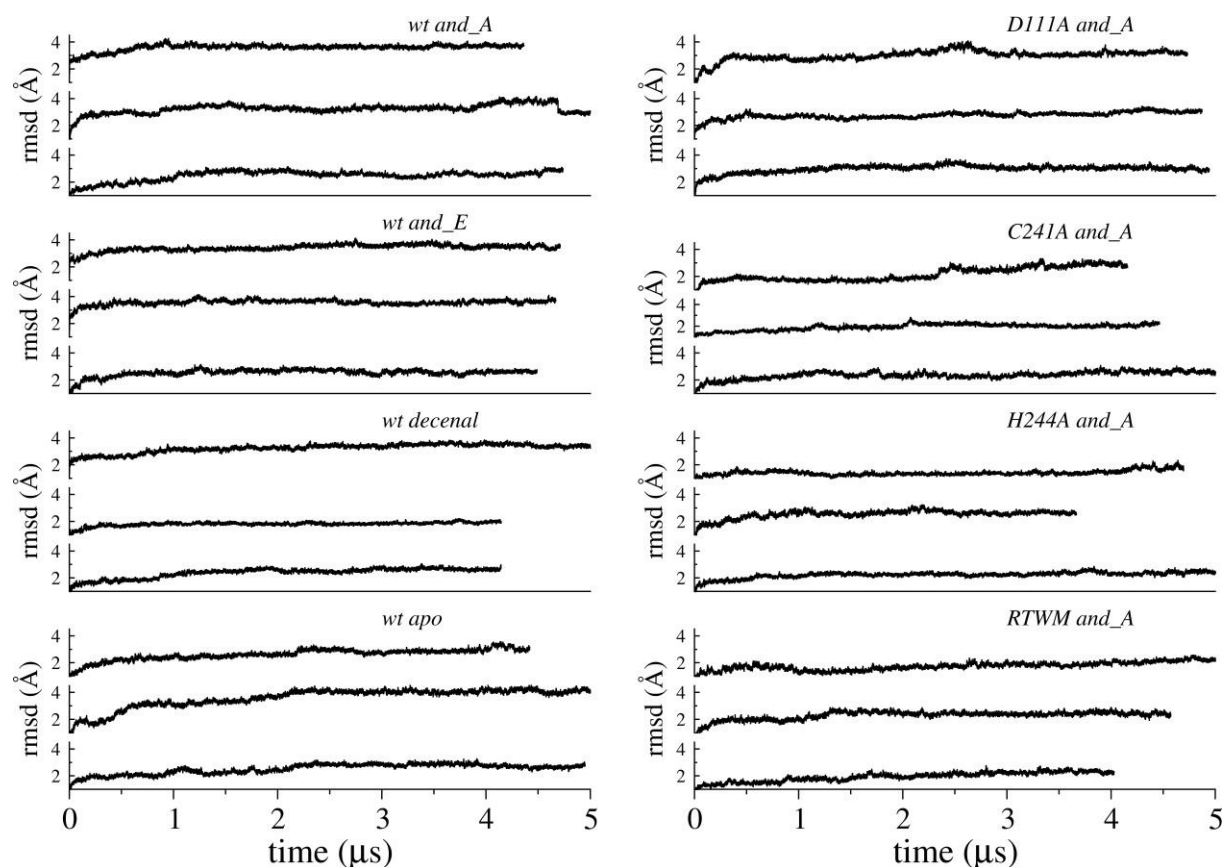


Figure S11. Rmsd of the eight systems studied with respect to time. The rmsd is computed for the receptor bundle backbone (CA, C, N) with respect to the first step of the MD production phase.

Figure S11 gathers the rmsd evolution for the eight ORs systems. For all system, one can observe that the rmsd fluctuates between 2.5 and 4 Å. This range of fluctuation highlights the stability of the transmembrane part of the systems during microsecond MD simulations.

Root Mean Square Deviation to active and inactive structures

All simulations are imaged and rms fit to the first step of molecular dynamic simulations. rmsd to active and inactive state are calculated using PLUMED software^[18] and taking $\beta 2$ adrenergic X-ray structures (PDB id: 2RH1 and 3SN6) as references. The frames alignment is made on the CA atoms from the trans-membrane segments without the RxKAFSTCASH motif from the TM6. The rmsd is then calculated for the CA atoms from this motif.

During the MD simulation on $\beta 2$ -AR bound to agonist two transitions from the inactive to active conformation are observed ([Figure S12](#)). When the system is bound to an antagonist the system remains in an inactive conformation.

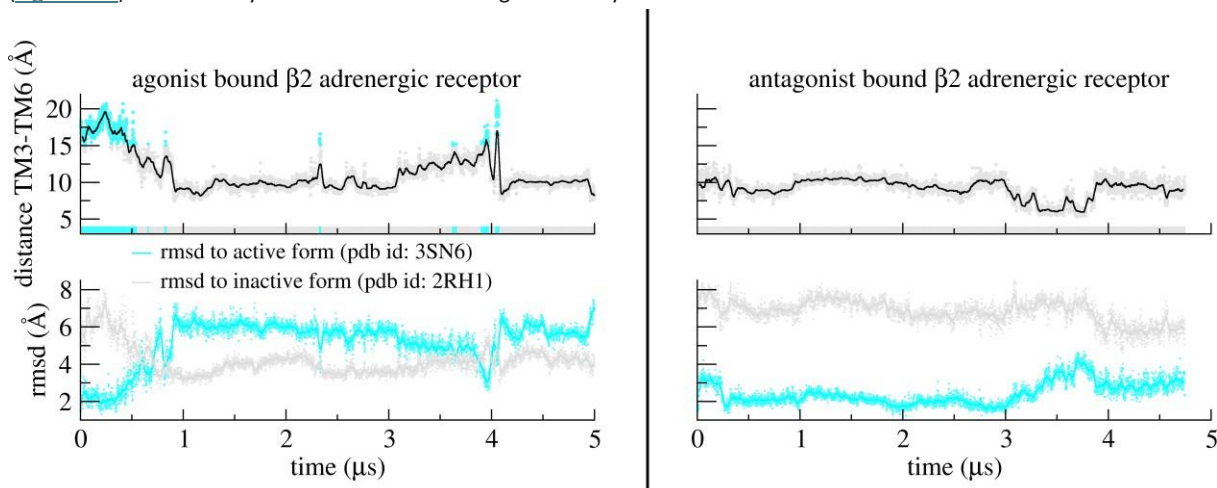


Figure S12. $\beta 2$ adrenergic receptor conformational dynamics. Top: Evolution of the intracellular cleft involving TM6 as measured by the distance between the CA atoms from the so-called ionic lock ($R^{3.50} D^{6.30}$). When the receptor is considered in its active state, the points are highlighted in cyan. They are otherwise shown in grey. Bottom: Evolution of the $\beta 2$ -ARreceptor rmsd (computed on the CA atoms from residues 211 to 222) when the reference frame is the active (cyan) or inactive state (grey). On the left, the agonist-bound active receptor deactivates after ~ 700 ns and get briefly reactivated for a short period after $\sim 4 \mu s$. The inactive state bound to an inverse agonist never activates.

A similar behavior is observed for the eight ORs systems studied. When an agonist is bound to the OR, one can observe switches from a conformation closer to an inactive state to a conformation closer to an active state ([Figure S13](#)). On the opposite, when the system is bound to decanal no switch is observed, the system remains in an inactive conformation during the 3 independent simulations. The same behavior is observed for the simulations on the WM mutant and on the H244A mutant which are both not responsive to androstadienone stimulation.

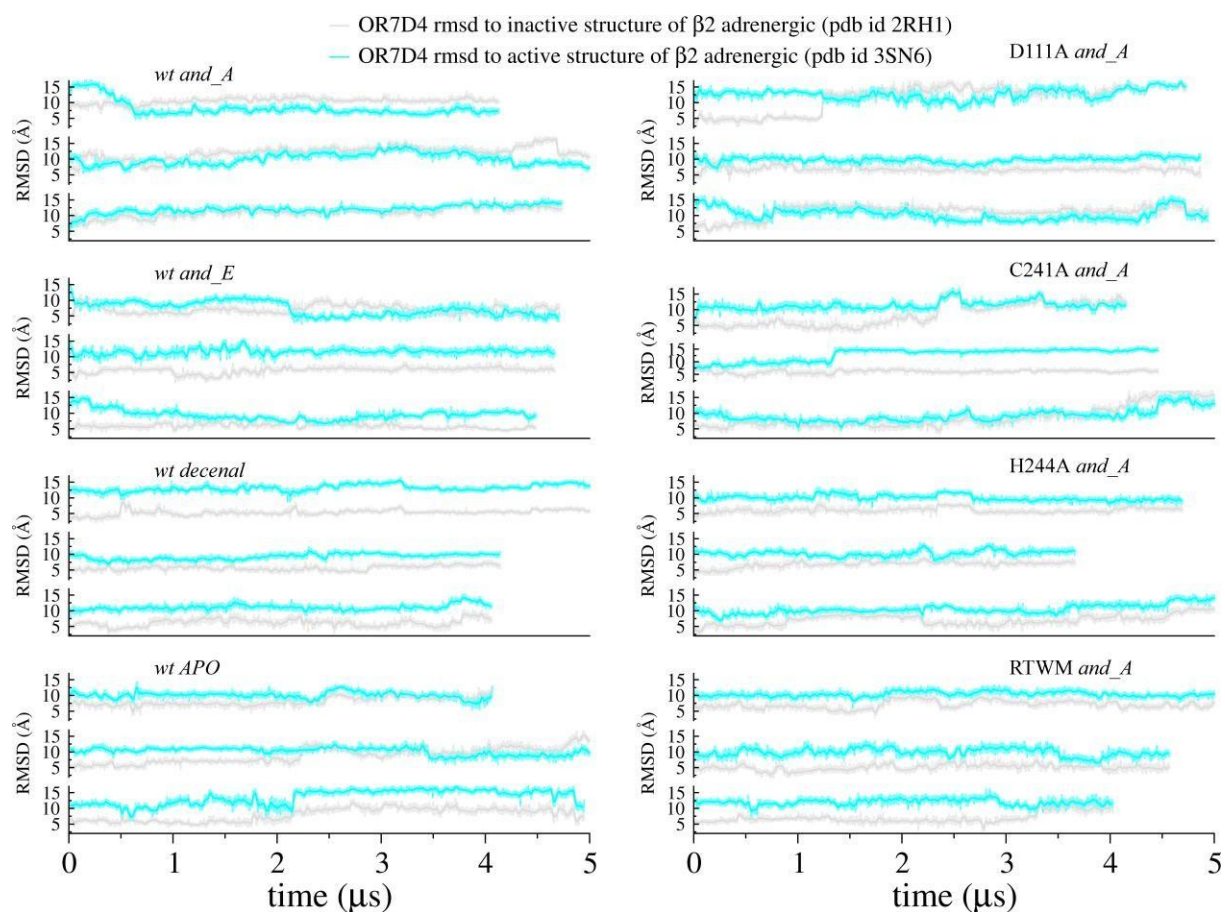


Figure S13. Evolution of the hOR7D4 rmsd (computed on the CA atoms from the KxKAFSTCASH motif) with respect to active (PDB id: 3SN6, cyan) and inactive (PDB id: 2RH1, grey) state of β 2-adrenergic receptor. When the cyan dots are under the grey ones, this means that the structure of hOR7D4 is closer to the active state with respect to the inactive state (1st trajectory of *wt and_A* for example).

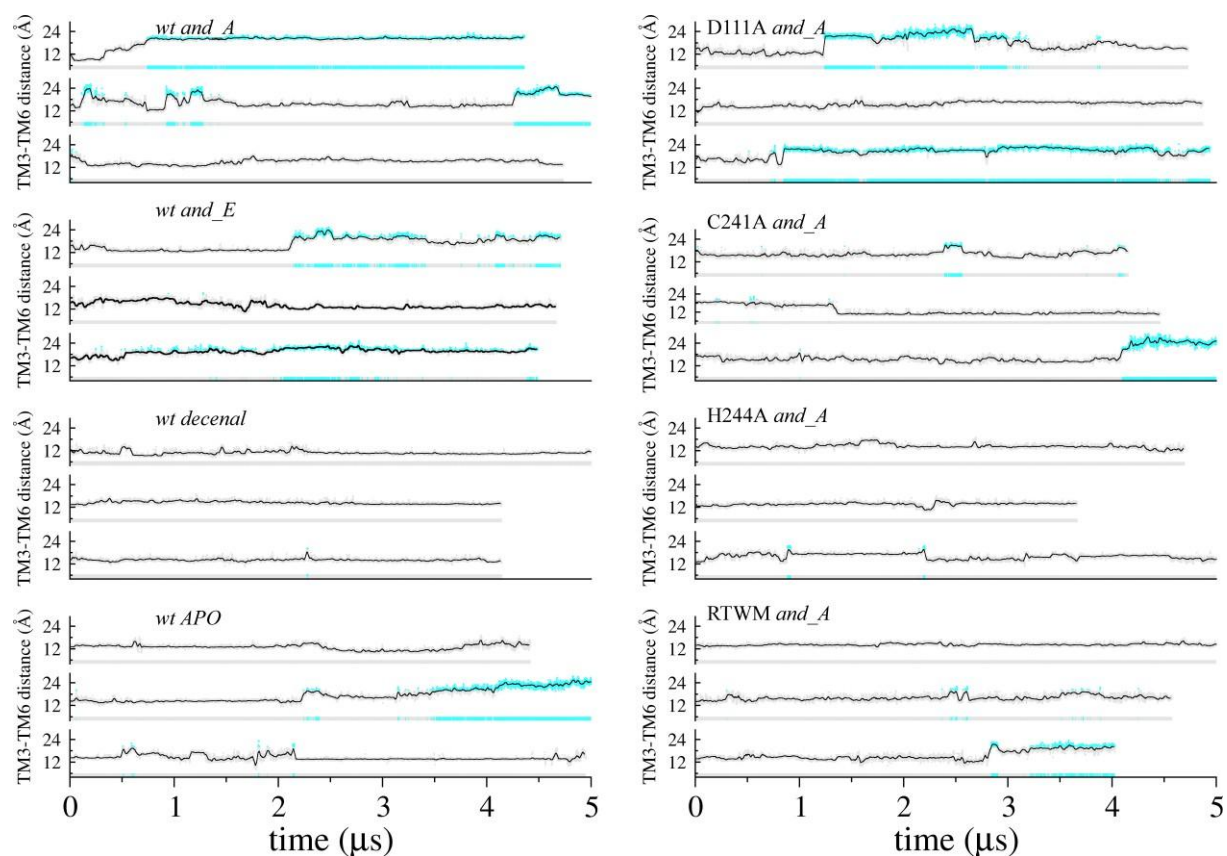


Figure S14. Time course of the cleft between TM3 and TM6 as measured by the distance between the C α atoms from the so-called ionic lock (D^{3.49}-K^{6.30}). The active state is highlighted in cyan and the inactive in grey.

Dynamic network analysis

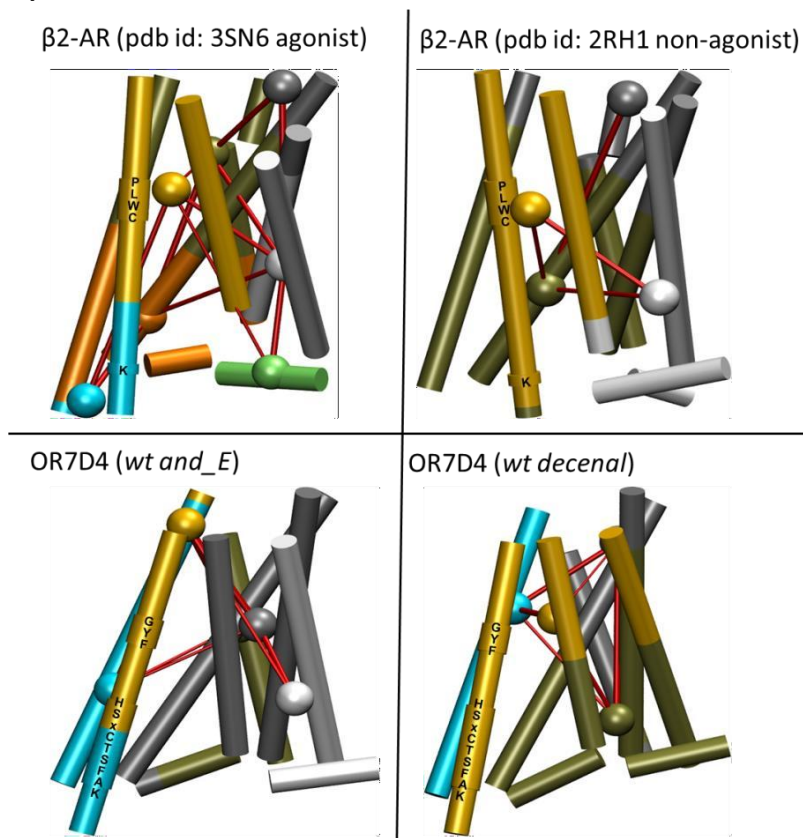


Figure S15. Network analysis comparison between the β 2-AR (top) and the OR7D4 system (bottom). Typical networks of an activated or inactivated system are shown on left and right, respectively. When an activation is observed, the KAFSTC_xSH motif (only K is conserved in the case of β 2-AR) of TM6 is involved in a separate node (cyan) which communicates with the node encompassing the so- called transmission switch (CWLP or FYG motif for β 2-AR or OR7D4, respectively).

Binding affinity evaluation using MM-GBSA calculations

The free energy of binding between the odorant and the OR was evaluated using the MM-GBSA method, according the following equation:

$$\Delta G_{\text{binding}} = \langle \Delta G_{\text{complex}} \rangle - \langle \Delta G_{\text{OR}} \rangle - \langle \Delta G_{\text{odorant}} \rangle \quad (1)$$

The $\langle \rangle$ corresponds to an average of a given value extracted from the MD trajectory.

Each term of the equation (1) can be written as:

$$\Delta G_X = \Delta G_{\text{MM}}^{\circ} + \Delta G_{\text{solv}} - T\Delta S \quad (2)$$

Where X corresponds to the complex, the odorant or the OR. The three terms are the gas-phase contribution to the binding energy, the solvation free energy upon binding and an entropic contribution respectively. The gas-phase contribution is the sum of three terms:

$$\Delta G_{\text{MM}}^{\circ} = \Delta E_{\text{intra}} + \Delta E_{\text{elec}} + \Delta E_{\text{vdW}} \quad (3)$$

The first term (ΔE_{intra}) is the difference in internal energy (bond, angle, and dihedral energy) of the OR and the odorant between the bound and unbound forms. The two other terms, (ΔE_{elec}) and (ΔE_{vdW}), are the non-bonded electrostatic and van der Waals energies between the bound and unbound form.

The solvation free energy includes two terms:

$$\Delta G_{\text{solv}} = \Delta G_{\text{pol}} + \Delta G_{\text{nonpol}} \quad (4)$$

The first terms ΔG_{pol} was estimated using the Generalised Born model of Onufriev, Bashford and Case. The value of the external dielectric constant was set to 80, while the internal one was set to 1. The non-polar term was considered as the solvent accessible surface area (SASA).

The terms $\Delta G_{\text{MM}}^{\circ} + \Delta G_{\text{solv}}$ represent the $\Delta G_{\text{MM-GBSA}}$ component of the free energy of binding.

Finally, the entropy component was evaluated by Normal Mode Analysis (NMA) using sander and nmode modules implemented in AMBER.

The MM-GBSA analysis was performed on 500 snapshots sampled every 10 ns and covering the whole MD simulation. Results for each studied system are gathered in [Table S1](#). A per-residue decomposition is shown in [Figure S16](#).

Table S1. Free energy of binding for the 2 full-agonists and the non-agonist. The agonists binding energies are comparable in all systems, suggesting that the mutations do not affect the affinity for the receptor.

		<i>and_A</i>	<i>and_E</i>	<i>decenal</i>	D111A	C241A	H244A	RTWM
	MD1	-41	-38	-27	-38	-35	-41	-38
$\Delta G_{\text{MM-GBSA}}$	MD2	-38	-40	-24	-37	-39	-35	-38
(kcal.mol ⁻¹)	MD3	-39	-40	-28	-36	-33	-33	-35
	<average>	-39.3	-39.5	-26.5	-37.0	-35.6	-36.2	-37.2

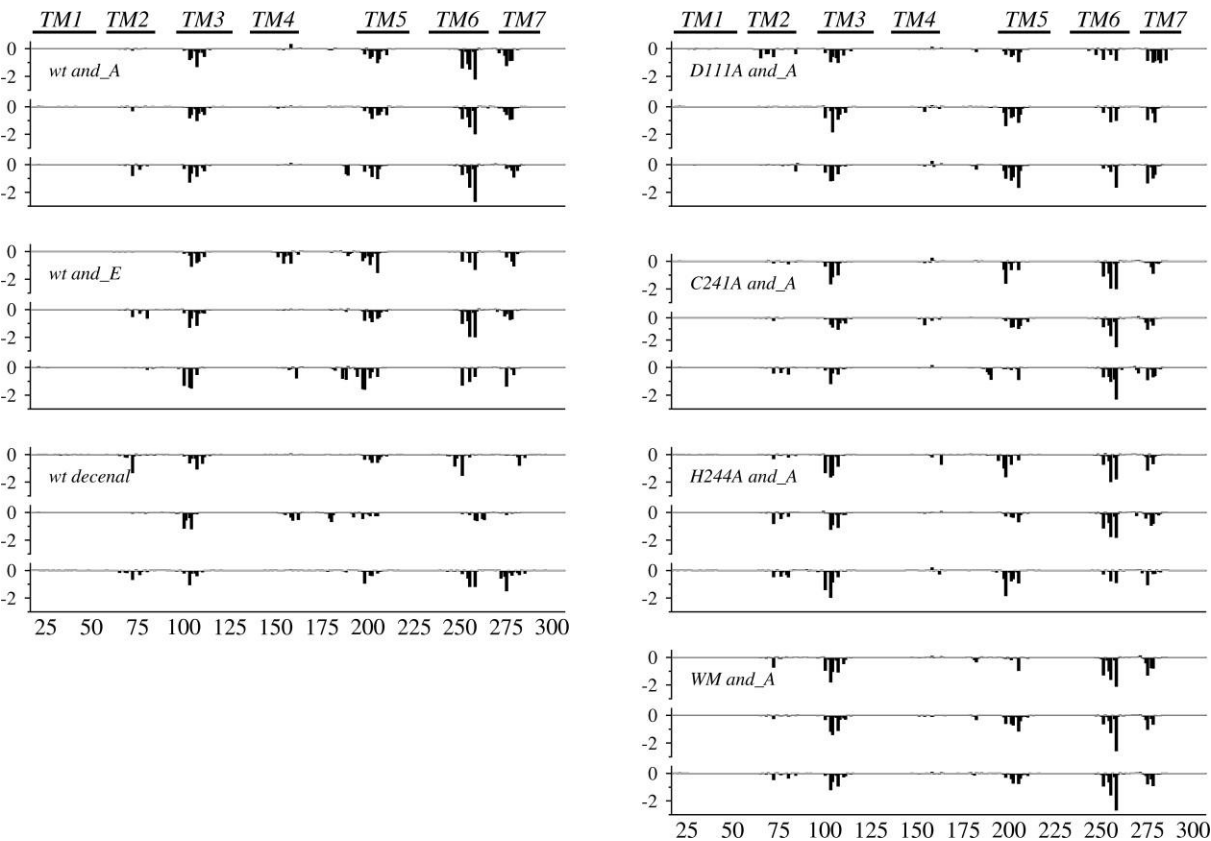


Figure S16. hOR7D4-ligands per-residue energy decomposition according to the MM-GBSA protocol. All patterns are highly similar except for decenal.

Conservation analysis

Table S 1. Conservation of residues involved in the network of TM3, TM6 and TM7.

TM3	A/G108 ^{3.36}	D/E111 ^{3.39}	L115 ^{3.43}	M118 ^{3.46}	R122 ^{3.50}	
	88%	94%	94%	97%	88%	
TM6	K234 ^{6.30}	A237 ^{6.33}	C241 ^{6.37}	H244 ^{6.40}	V248 ^{6.44}	Y252 ^{6.48}
	77%	89%	94%	97%	97%	92%
TM7	P283 ^{7.46}	N286 ^{7.49}	Y290 ^{7.53}	R293 ^{7.56}		
	95%	96%	95%	96%		

Conservation analysis

Table S 2. Conservation of residues involved in the network of TM3, TM6 and TM7.

TM3	A/G108 ^{3.36}	D/E111 ^{3.39}	L115 ^{3.43}	M118 ^{3.46}	R122 ^{3.50}	
	88%	94%	94%	97%	88%	
TM6	K234 ^{6.30}	A237 ^{6.33}	C241 ^{6.37}	H244 ^{6.40}	V248 ^{6.44}	Y252 ^{6.48}
	77%	89%	94%	97%	97%	92%
TM7	P283 ^{7.46}	N286 ^{7.49}	Y290 ^{7.53}	R293 ^{7.56}		
	95%	96%	95%	96%		

MATCHING RECEPTOR TO ODORANT WITH PROTEIN LANGUAGE AND GRAPH NEURAL NETWORKS

Matej Hladis*, Maxence Lalis, Sébastien Fiorucci, Jérémie Topin*

Institut de Chimie de Nice, Université Côte d’Azur, UMR 7272 CNRS, France

{matej.hladis, jeremie.topin}@univ-cotedazur.fr

ABSTRACT

Odor perception in mammals is triggered by interactions between volatile organic compounds and a subset of hundreds of proteins called olfactory receptors (ORs). Molecules activate these receptors in a complex combinatorial coding allowing mammals to discriminate a vast number of chemical stimuli. Recently, ORs have gained attention as new therapeutic targets following the discovery of their involvement in other physiological processes and diseases. To date, predicting molecule-induced activation for ORs is highly challenging since 43% of ORs have no identified active compound. In this work, we combine [CLS] token from protBERT with a molecular graph and propose a tailored GNN architecture incorporating inductive biases from the protein-molecule binding. We abstract the biological process of protein-molecule activation as the injection of a molecule into a protein-specific environment. On a newly gathered dataset of 46 700 OR-molecule pairs, this model outperforms state-of-the-art models on drug-target interaction prediction as well as standard GNN baselines. Moreover, by incorporating non-bonded interactions the model is able to work with mixtures of compounds. Finally, our predictions reveal a similar activation pattern for molecules within a given odor family, which is in agreement with the theory of combinatorial coding in olfaction.

1 INTRODUCTION

Mammalian sense of smell constantly provides information about the composition of the volatile chemical environment and is able to discriminate thousands of different molecules. At the atomic scale, volatile organic compounds are recognized by specific interactions with protein receptors expressed at the surface of olfactory neurons (Buck & Axel, 1991). Mammalian epithelium expresses hundreds of different olfactory receptors (ORs), belonging to the G protein-coupled receptors (GPCRs), which constitute the largest known multigene family (Niimura & Nei, 2003). The recognition of odorants by ORs is based on the complementarity of structures and hydrophobic or van der Waals interactions which leads to low molecular affinity (Katada et al., 2005). With the exception of a few conserved amino acids, the sequences of ORs show little identity. In particular, the ligand-binding pocket has hypervariable residues (Pilpel & Lancet, 1999) that are relatively well conserved between orthologs. This property gives ORs the ability to bind a wide variety of molecules that differ in structure, size, or chemical properties. The recognition of odorants is done according to the combinatorial code of activation (Malnic et al., 1999). Each odorant is recognized by several ORs, whereas an individual OR can bind several odorants with distinct affinities and specificities (Zhao et al., 1998). This combinatorial code is sensitive to subtle modifications, so the response of a single receptor can have a major influence on the smell perception. Even a small sequence modification could affect odorant responsiveness (Keller et al., 2007; Mainland et al., 2014). On the other hand, structural and functional modifications of an odorant can abolish the interaction with a specific receptor (Katada et al., 2005), and even lead to a different smell perception (Sell, 2006). So far the combinatorial code of the majority of odorants remains unknown. Identifying the recognition spectrum of each OR is therefore essential to decipher the mechanisms of the olfactory system and subsequently build models capable of cracking the combinatorial code of activation.

*Corresponding authors.

There is only a limited number of models designed to match ligands and ORs. Namely Kowalewski & Ray (2020) follow a molecule-oriented approach and predict agonists for a subset of 34 ORs (24 wild types and 10 variants) by representing molecules via fingerprints (Morgan, 1965; Klekota & Roth, 2008) and building an individual SVM model for each OR. On the other hand, Cong et al. (2022) focus on receptors and address a more complex problem of predicting active ORs for 4 given molecules. In their random forest model, each amino acid of the protein sequence is described by 3 physico-chemical properties and the molecules by a subset of Dragon descriptors (Mauri et al., 2006). In a more general approach, Gupta et al. (2021) consider any OR-molecule pair and use BiLSTM (Graves & Schmidhuber, 2005) to predict the binding of a molecule, represented by SMILES string, to an OR sequence.

In contrast, in this work we use a graph and a [CLS] token embedding from protBERT (Elnaggar et al., 2021) to represent molecules and receptors, respectively. We abstract receptor-molecule binding as the injection of a molecule into a protein specific environment. This is achieved by using molecular topology as a layout for the message passing process and copying the protein representation to each node of the molecular graph. As a result, the redundancy of protein information enables a local processing of the "protein environment" and achieves better performance than a common strategy of processing the receptor and the molecule in parallel. This abstraction leads to a graph with the number of nodes depending only on the size of the molecule.

Molecules with flexible moieties can undergo conformational changes upon binding to maximize interactions with the receptor binding cavity. This structural adaptation modifies the strength of internal non-bonded forces. However, standard GNN architectures are restricted to the molecular topology. Thus, we have built a tailored GNN architecture combining local interaction of bonded atoms, as done by standard GNNs, with multi-head attention, giving the model the ability to incorporate interactions between any pair of atoms. We show that this architecture outperforms other baselines as well as previous work on olfactory receptor-molecule activation prediction.

Finally, we found a relationship between human odor perception and model predictions, strengthening the biological relevance of the model. The results for humans are in full agreement with the experimental work done by Nara et al. (2011). By analyzing the predictions for human ORs, we observe that the combinatorial codes exhibit large diversity. The OR repertoire contains mostly narrow receptors with several broadly-tuned ones. The results also highlight the existence of odor-specific receptors, but most odors are coded in a complex activation pattern.

2 PRELIMINARIES

2.1 PROTEIN LANGUAGE MODELS

Recently, protein language models emerged as unsupervised structure learners (Rao et al., 2021b; Vig et al., 2021; Rives et al., 2021), allowing to extract abstract vector representations of proteins. As in natural language processing (NLP), large models with millions of parameters (e.g BERT (Devlin et al., 2019)) are trained on vast databases of amino acid sequences (Steinegger et al., 2019; Steinegger & Söding, 2018; UniProt Consortium, 2019; Suzek et al., 2007). Rao et al. (2019) trained and evaluated various natural language processing models on a set of structurally relevant tasks. Elnaggar et al. (2021) went further and trained a list of powerful NLP models ranging from 200M to 11B parameters. Recently proposed MSA Transformer (Rao et al., 2021a) exploits evolutionary relationships by using Multiple Sequence Alignment (MSA) as input rather than a single protein sequence at a time. AlphaFold2 (Jumper et al., 2021; Evans et al., 2021) extends the idea of using MSA and combines it with an experimentally obtained protein template in an end-to-end model trained on supervised structure prediction.

2.2 GRAPH NEURAL NETWORKS

In recent years, graph neural networks (Kipf & Welling, 2016; Gilmer et al., 2017; Simonovsky & Komodakis, 2017; Veličković et al., 2018; Wang et al., 2018; Zhou et al., 2018; Battaglia et al., 2018) have grown rapidly in popularity and received considerable attention in various domains such as drug design (Torng & Altman, 2019), physics (Shlomi et al., 2021), and chemistry (Gilmer et al., 2017; Yang et al., 2019; Gasteiger et al., 2020b;a). Chemistry is a particularly promising field for GNN applications since a molecule can be naturally represented as a graph $G = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is the set

of nodes (atoms) and $\mathcal{E}$ the set of edges (bonds). Each node $v \in \mathcal{V}$ and each edge $(u, v) \in \mathcal{E}$ in G have an initial vector of features, x_v and $e_{u,v}$, respectively. These feature vectors contain information about atom/edge properties such as atomic number, atomic weight, or bond type.

3 MODEL

To tackle the problem of receptor-molecule activation prediction we take a proteo-chemometric approach (Qiu et al., 2017) where the model estimates the probability $P(m, r)$ that a given molecule m activates a given receptor r . The input to the model is the amino acid sequence of the receptor on one side and the molecular graph on the other side. See Fig. 1 (a) for the model outline.

Protein sequence representation is obtained using protBERT (Elnaggar et al., 2021) pretrained on a dataset with 217M protein sequences (Suzek et al., 2007). Olfactory receptors are biopolymers of around 300 amino acids and modeling each amino acid explicitly is computationally demanding. To keep the costs relatively low, we take the advantage of the BERT’s classification token ([CLS]) (Devlin et al., 2019) which aggregates information about an entire protein sequence. This way we model amino acids implicitly as they contribute to the [CLS] token embedding, and at the same time, the sequence representation is reduced from roughly $300 \times d$ to only d , where d is an embedding dimension. We concatenate the [CLS] token embedding from the last 5 layers of protBERT and use this vector as the receptor input. A molecule is represented as a graph G with node and edge features in Tab. 5. Note that the representation relies solely on the molecular topology and does not take into account spatial characteristics.

The model architecture itself was designed by observing several inductive biases which emerge in molecule-receptor interactions:

Locality. The binding between a receptor and a molecule relies on subtle interactions between the binding groups of both protagonists. In particular, molecules contain functional groups that are crucial for forming intermolecular bonds with the receptor. Since we use an aggregated sequence representation, and thus do not explicitly model interactions between amino acids and atoms in the molecule, we emulate this "locality" by copying the receptor embedding to each node of the molecular graph (Fig. 1 (a)). This way the sequence representation is part of the node embedding and can be locally changed based on the connectivity and neighborhood properties of each node.

Non-bonded interactions. In addition to bonded interactions, atoms in the same molecule interact with each other via electrostatic and hydrophobic interactions. To account for these, we use the attention mechanism (Vaswani et al., 2017), which allows a flow of information between any pair of atoms (Fig. 1 (b)). This is particularly relevant for flexible molecules where two distant atoms in the topology could be close in the conformation. Such approach also inherently takes into account interactions between multiple molecules and thus allows for modeling mixtures.

Receptor-molecule complex. When a molecule binds to a receptor, the two form a stable complex that can be considered as a single entity. To accommodate this observation, we combine the molecule and receptor inputs early in the model pipeline rather than mixing them at a late stage. This facilitates the flow of information between them throughout the model.

Combining the above observations, we propose the model architecture in Fig. 1. Based on the *locality* and *receptor-molecule complex* observations, we inject receptor embedding into the molecular graph as additional node features, and we use the molecular topology as a layout for the message passing process (Fig. 1 (a)). This approach reduces the biological problem of protein-molecule interaction to a small-scale graph-level binary classification where a molecule is injected into a receptor environment. In this formulation, the size of the graph depends solely on the number of atoms in the molecule.

Given the above abstraction, any standard GNN architecture might be applied to predict a molecule-receptor activation. However, standard models do not allow for information to pass outside of the graph topology. To account for *non-bonded interactions*, we combine the local processing done by a GNN with the multi-head attention. To enhance the expressive power of the network, we use two separate identical GNNs for queries and keys/values. We denote these GNNs Q-MPNN for "queries" and KV-MPNN for "keys/values". Each row in the queries and keys/values is a node embedding resulting from the corresponding GNN (Fig. 1 (b)). Then we use the multi-head attention

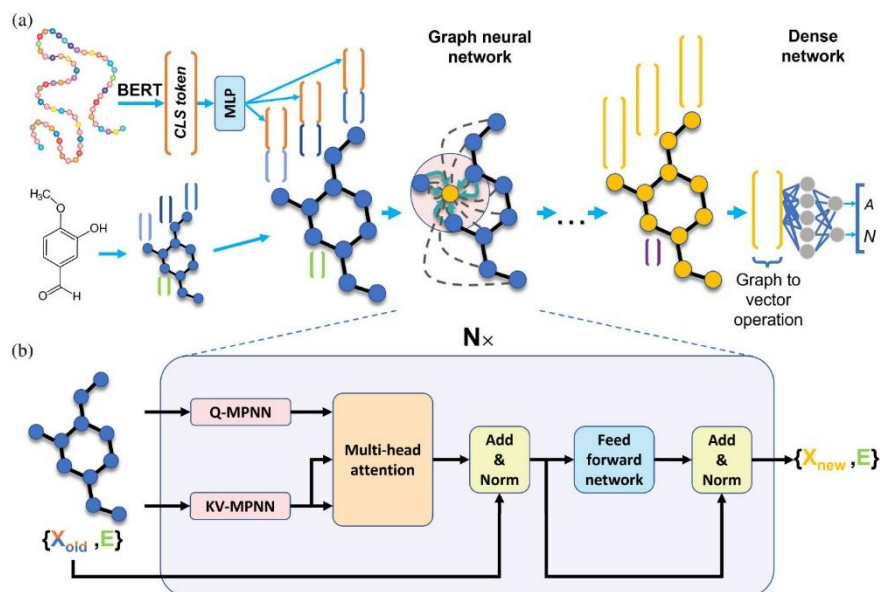


Figure 1: Model design. (a) Overview of the model. The input is a pair of protein sequence and molecular graph. The sequence is embedded using [CLS] token from protBERT (Elnaggar et al., 2021) and the resulting representation is concatenated to each node of the molecular graph. (b) Graph processing block. The node embeddings X_{old} of the input graph are first locally updated using two identical message passing neural networks (Q-MPNN and KV-MPNN), and then pairwise interactions are modeled by the multi-head attention. This is followed by residual connections, layer normalization and a feed forward network (Vaswani et al., 2017) to obtain the updated node embeddings X_{new} . Graph to vector operation is an ECC layer (Simonovsky & Komodakis, 2017) followed by an attention pooling (Section A, equation (2)). Edge features E are used in Q-MPNN and KV-MPNN and are reset at the end of the block.

followed by layer normalization (Ba et al., 2016) and feed forward network (Vaswani et al., 2017). A similar approach was taken in (Rong et al., 2020), but our methodologies differ. Rong et al. (2020) replace linear transformations in the multi-head attention with graph neural networks and thus have $3 * |heads| * |layers|$ different message passing steps. Instead, we use two GNNs to construct queries and keys/values, and the standard multi-head attention with linear transformations, so our model has $2 * N * |layers|$ steps, where N is the number of graph processing blocks.

The final model simply consists of N blocks followed by an ECC layer (Simonovsky & Komodakis, 2017) and an attention pooling as the graph to vector operation, and a dense layer for final output. We chose to keep the GNNs architecture simple since there is a subsequent processing. We use a dense layer followed by ReLU to construct messages from concatenated sender, receiver and edge features, sum aggregation and a GRU cell (Cho et al., 2014) to update nodes¹.

Training. To avoid overfitting we adopt similar strategy as in (Rong et al., 2020). During training, a random number of message passing steps T is sampled from a truncated normal distribution $N(\mu, \sigma, a, b)$ in each iteration of gradient descent. We use the Adam optimizer (Kingma & Ba, 2015) with the learning rate schedule from (Vaswani et al., 2017) and in all experiments we train the model for 10 000 epochs. In the first 8000 epochs we train on graphs with up to 32 nodes and 64 edges, and then we switch to the full dataset during the remaining 2000 epochs. For evaluation and prediction, we set the number of message passing steps to μ . In our experiments, we set $\mu = 6$, $\sigma = 1$ and the truncation interval between $a = 3$ and $b = 9$.

¹Model source code is available here: <https://github.com/MatejH1/Receptor2Odorant>

4 DATA

Only a limited amount of organized and curated data on OR-molecule pairs is available (Cong et al., 2022; Sharma et al., 2021; Liu et al., 2011). To fill this gap, we have gathered and curated a new dataset of 46 700 unique OR-molecule pairs from the literature, tripling the size of the currently largest database of this kind (Cong et al., 2022). The data puts together experiments for 1237 unique sequences for 11 mammalian species with additional information on experimental procedure and varying data quality. See Supplementary material G for further details about the dataset.

Mammalian olfaction is sensitive to chirality and two enantiomers can have different smell perception. However, 21% of the gathered data are actually pairs of an OR and a mixture of enantiomers. To address this, we consider two modeling strategies for the mixtures of enantiomers: a "mixture" approach where we exploit the ability of our proposed model to work with mixtures and a "single" approach where a mixture of enantiomers is treated implicitly, by changing the chiral property of the nodes. See further details in sections B and E.1.

In addition, there are several biases in the data that need to be addressed during training. Label noise coming from high-throughput screening, class imbalance, and the combinatorial nature of the task. In response, we construct sample weights for each sub-problem and use a weighted binary cross entropy loss where the final weights are the product of the partial weights. Further details for weights construction are discussed in Supplementary material C.

4.1 TEST SET

Because of the varying data quality, special attention needs to be paid to the test set construction. In order to have high confidence in the results we select only pairs from the highest quality dose-response data for the test set. Unless otherwise stated, in each cross-validation run in the experiments, we randomly select 30% of the dose-response samples for the test set. We run 5 cross-validation runs for all experiments.

5 EXPERIMENTS

We performed a series of experiments on the newly gathered dataset of OR-molecule pairs. In 5.1 we compare our proposed architecture with other GNNs and perform ablation studies on the inductive biases. We explore the generalization of the model to unseen receptors and molecules in Section 5.2, and in 5.3 we compare our approach with the previous studies of OR-molecule interactions. We present additional ablations and experiments on two other datasets in Supplementary material E. Due to the imbalance nature of molecule-protein activations, we use Mathews correlation coefficient (MCC) as the main evaluation metric. We also report others for comparison with previous work.

5.1 MODEL

GNN architecture We compare our model with 4 standard GNN models and we experiment with 2 additional modifications of our approach. We also compare our results with 2 state-of-the-art models on drug-target interaction prediction (DTI), namely HyperAttentionDTI (Zhao et al., 2021) and MolTrans (Huang et al., 2020), which based on Zhao et al. (2021) perform best on three different DTI datasets. We use ECC (Simonovsky & Komodakis, 2017), GGNN (Li et al., 2015), GAT (Veličković et al., 2018) and GIN (Xu et al., 2019) as standard GNN baselines. GAT is a particularly interesting baseline because it is a generalization of attention mechanism to an arbitrary graph. In the first ablation, we consider dropping the *non-bonded interactions* observation and we replace Q-MPNN, KV-MPNN and the multi-head attention in Fig. 1 (b) with a GAT layer (Transformer GAT in Tab. 1). This way the attention mechanism is kept, but we do not allow direct interactions between all pairs of atoms, and the model is restricted solely to the topology of the molecule. The second ablation is in the opposite direction. We drop the molecular topology and use only the multi-head attention by setting Q-MPNN and KV-MPNN to identity functions. Note that this reduces to the Transformer (Vaswani et al., 2017) encoder model.

As can be observed in the results in Tab. 1, our model outperforms the standard baselines and also proves to be the best compared to the DTI models. The ECC model failed to converge at all and its

Table 1: Summary of results for different architectures. *Transformer GAT* corresponds to replacing Q-MPNN, K-MPNN and the multi-head attention in Fig. 1 (b) with the GAT layer, and *Transformer* corresponds to setting Q-MPNN and K-MPNN to identity functions. All ablation models have 5 layers/blocks. Since baseline models cannot process mixtures, we use "single" approach to represent mixtures of enantiomers for all models except *Ours - mixture* (see Section B for details).

Model	AveP	Precision	Recall	F-score	MCC
ECC	0.307 (0.02)	0.325 (0.02)	0.527 (0.03)	0.402 (0.03)	0.187 (0.03)
GGNN	0.676 (0.02)	0.447 (0.02)	0.833 (0.02)	0.582 (0.02)	0.455 (0.02)
GAT	0.686 (0.03)	0.466 (0.03)	0.834 (0.03)	0.597 (0.02)	0.476 (0.02)
Transformer GAT	0.704 (0.02)	0.556 (0.01)	0.771 (0.04)	0.645 (0.01)	0.536 (0.01)
GIN	0.743 (0.02)	0.641 (0.03)	0.708 (0.04)	0.672 (0.02)	0.574 (0.02)
Transformer	0.748 (0.04)	0.654 (0.04)	0.695 (0.05)	0.671 (0.01)	0.576 (0.01)
MolTrans	0.638 (0.07)	0.402 (0.05)	0.822 (0.03)	0.556 (0.05)	0.476 (0.04)
HyperAttentionDTI	0.737 (0.02)	0.609 (0.03)	0.773 (0.02)	0.681 (0.02)	0.584 (0.02)
Ours - single	0.765 (0.02)	0.665 (0.01)	0.711 (0.02)	0.687 (0.02)	0.595 (0.02)
Ours - mixture	0.780 (0.01)	0.689 (0.02)	0.698 (0.04)	0.693 (0.02)	0.605 (0.02)

Table 2: Summary of results for strategies to combine sequence and molecule inputs. Mixtures of enantiomers are treated as a single graph.

Mixing	AveP	Precision	Recall	F-score	MCC
Concatenation	0.687 (0.03)	0.550 (0.05)	0.729 (0.02)	0.626 (0.03)	0.509 (0.03)
Attention	0.720 (0.02)	0.604 (0.02)	0.671 (0.03)	0.635 (0.02)	0.526 (0.02)
Ours - single	0.765 (0.02)	0.665 (0.01)	0.711 (0.02)	0.687 (0.02)	0.595 (0.02)

performance is poor compared to the others. The best from the baseline models is GIN with MCC 0.574, followed by GAT and GGNN. Compared to simple GAT, the performance increases in the first ablation when we combine the GAT layer with feed-forward network, normalization, and residual connections. Interestingly, ignoring the molecular topology and using Transformer yields better performance than any of the baseline GNNs or than Transformer GAT. This is surprising because Transformer has no access to the topology of the molecule and this suggests that the multi-head attention component is crucial for performance. However, combining both the multi-head attention and GNN together works the best (our model), implying that both the molecular topology and pairwise interactions are needed. Comparing to the DTI models, MolTrans performs worse than most of the baseline models and ablations. Our approach surpasses it by a large margin of 0.119 (0.129 with the "mixture" approach). HyperAttentionDTI performs better than MolTrans and the baseline GNNs, but is still outperformed by our approach.

Mixing It is not evident at which point in the model the protein and molecule inputs should be mixed. We argue that the molecule and the receptor create a complex, and since they constantly interact with each other, the information should be mixed in an early stage of the processing. We test this hypothesis against a more common strategy (Du et al., 2022) of late mixing of protein and molecule information. We compare our model with a standard strategy of processing molecular information separately via a GNN and then mixing the molecular embedding with the protein representation. This breaks the *locality* and *receptor-molecule complex* observations. We consider two mixing strategies: concatenation and attention. Concatenation is the most standard one and in this case the molecule is processed without the protein information and the resulting embedding from the graph to vector operation is concatenated to the protein [CLS] token. This concatenated vector is then the input to a dense network. The second strategy is to use the resulting molecule embedding as keys/values in the multi-head attention and the [CLS] representation of the protein as a query. This is followed by layer normalization and feed-forward network like in the Transformer block. We use the proposed blocks in Fig. 1 (b) to generate molecular embedding.

The results for mixing ablation are summarized in Tab. 2. Despite the redundancy of the protein representation in the nodes, our approach outperforms other mixing strategies. This suggests that by

incorporating the *locality* and *receptor-molecule complex* inductive biases, the model can learn to exploit the dependencies between various molecular binding groups and the receptor.

5.2 GENERALIZATION

An important and perhaps the most interesting experiment we conduct is to test the generalization of the model. 43% of mammalian ORs have no identified active compound, and only 206 out of 385 human ORs have a known responsive molecule. Therefore, the prediction of active compounds for so-called orphan receptors (i.e. deorphanization task) remains a great challenge. Moreover, among 6389 described odorant molecules (Castro et al., 2022), 504 have been tested so far, and finally only 336 of them are known to induce an OR response. Thus, it is of major interest to identify new active molecules for ORs (i.e. expanding chemical space).

Deorphanization. Model generalization to unseen ORs is tested in two scenarios. In the first one (called *random* in Tab. 3), several ORs are randomly chosen to be exclusively included in the test set and all their occurrences are removed from the training set. This scenario appears more realistic because deorphanized ORs are found in each OR family. In the second, more complex scenario, entire clusters of receptors are selected for the test set. The ORs are clustered according to their sequence similarity, and then all dose-response pairs for all receptors in 9 randomly picked clusters are placed in the test set. In this complex case, the model has to extrapolate to proteins with dissimilar sequences. In both scenarios, we also experiment with keeping negative primary and secondary screening pairs for the selected receptors (*OR-keep* in Tab. 3).

Expanding chemical space. Identifying new active compounds for olfactory receptors remains challenging. To test the ability of the model to generalize to new molecules, we follow the same strategy as for "deorphanization" (*Molecules* in Tab. 3). We either randomly select molecules and put all their occurrences to the test set, or we choose several clusters based on Tanimoto coefficient (i.e. structural similarity). See Section D for the full description of the experimental setup.

Generalization results. The results of the generalization experiments are summarized in Tab. 3. In the first *random* scenario, mainly recall is affected compared to an i.i.d. split and precision undergoes minor changes by 2-5%. Keeping non-active molecules has small impact and leads to a more conservative model with higher precision and lower recall. It appears that the model generalizes poorly to entirely new family of receptors (Tab. 3, *cluster* task). This is due to the fact that subtle variations in the amino acids sequence of ORs can utterly alter its recognition profile (de March et al., 2018; Trimmer et al., 2019). Interestingly, even though overall MCC is low for the *cluster* scenario, the precision is still maintained at a relatively high level with a difference of 11-15% compared to the i.i.d. split. From the molecule perspective, the model can generalize well to new compounds that have similar distribution to the training set. As expected, the performance decreases on an entirely unseen group of molecules, yet the model still keeps the precision of 54.4%, only 14% lower than in the i.i.d. case. The recall on the other hand deteriorates by half. This suggests that the model is conservative and reluctant to predict many responsive pairs, which can be an advantage for *in vitro* experiments. Although some responsive pairs are missed as a result of lower recall, the predictions keep similar reliability across scenarios, and thus the number of laboratory hits would remain stable.

5.3 COMPARISON WITH OR-MOLECULE MODELS

In Tab. 4, we compare our work with Kowalewski & Ray (2020), Cong et al. (2022), and Gupta et al. (2021). Overall, when generality is taken into account, our model outperforms previous approaches. In particular, Kowalewski & Ray (2020) perform slightly better in AUROC than others, but this approach fits a separate SVM model for each OR and can only be generalized to ORs that already have at least 3 known active compounds. This is a limitation for most of ORs. On the other hand, Cong et al. (2022) took a more general approach by fitting a single model to data for 10 molecules (4 targets and 6 similar compounds). It is applicable to a variety of ORs, but only for compounds structurally close to the 4 targets. Even though Cong et al. (2022) consider an easier setting with a limited number of molecules, our approach outperforms it by a large margin. Finally, our model and the BiLSTM model of Gupta et al. (2021) consider any given molecule-receptor pair. Our approach outperforms Gupta et al. (2021) in all the metrics and more than twice the precision.

Table 3: Summary of results for a generalization test set in case that a cluster of molecules/sequences is not present in the training set (*Cluster*) and randomly chosen molecules/sequences are not present in the training set (*Random*). *OR - keep* corresponds to a case where primary and secondary screening non-active molecules are kept in the training set. Mixtures of enantiomers are treated as a "mixture".

	Split	Num data	AveP	Precision	Recall	F-score	MCC
	i.i.d.	1565.0	0.780 (0.01)	0.689 (0.02)	0.698 (0.04)	0.693 (0.02)	0.605 (0.02)
Cluster	Molecule	1320.4	0.580 (0.08)	0.544 (0.07)	0.342 (0.06)	0.418 (0.06)	0.334 (0.07)
	OR	1062.0	0.558 (0.14)	0.535 (0.12)	0.132 (0.04)	0.203 (0.04)	0.088 (0.06)
	OR - keep	1062.0	0.625 (0.03)	0.576 (0.06)	0.095 (0.03)	0.161 (0.05)	0.091 (0.09)
Random	Molecule	1056.2	0.729 (0.08)	0.657 (0.11)	0.629 (0.04)	0.638 (0.06)	0.533 (0.07)
	OR	1217.8	0.684 (0.10)	0.636 (0.07)	0.491 (0.11)	0.552 (0.09)	0.417 (0.10)
	OR - keep	1217.8	0.710 (0.09)	0.670 (0.06)	0.470 (0.13)	0.548 (0.11)	0.430 (0.10)

Table 4: Comparison with previous studies on OR-molecule activation prediction. *ORs* and *Mols* are average numbers of unique receptors and molecules, respectively, in the test set.

Model	ORs	Mols	TNR	AUROC	Precision	Recall	F-score
SVM ^a	32	n/a	0.80 (0.06)	0.88 (0.07)	n/a	0.77 (0.10)	n/a
RF ^b	80	4	0.91	0.74	0.52	0.53	0.53
BiLSTM ^c	n/a	n/a	n/a	0.77	0.34	0.71	0.46
Ours - mixture	205	237	0.91 (0.01)	0.79 (0.03)	0.69 (0.02)	0.70 (0.04)	0.69 (0.02)

^aKowalewski & Ray (2020)

^bCong et al. (2022)

^cGupta et al. (2021), results taken from supplementary Fig. 3.

6 AGREEMENT WITH ODOR PERCEPTION

The combinatorial nature of olfaction was first hypothesized in 1973 by Polak (1973) and further confirmed by the discovery of ORs by Buck & Axel (1991). Malnic et al. (1999) demonstrated that different odorants are recognized by different subsets of ORs. Here, we refer to a subset of 385 human ORs activated by a given odorant as the combinatorial code, and in this section we assess a biological plausibility of our approach by analyzing consistency between the combinatorial code predicted by our model and odor perception of molecules (i.e. the odor families).

We evaluate the concordance between our model and the olfactory perception as follows: given a large dataset of known odorants (Castro et al., 2022), we construct an abstract smell characteristic of a molecule by taking an embedding of a GNN-based odor prediction model (Sanchez-Lengeling et al., 2019). This model was recently shown to outperform a median panelist in odor description task, and its embedding can capture perceptual similarity of molecules (Lee et al., 2022). The odorants are clustered based on their odor embedding to abstract odor families. For each family, we define an agreement score as the α -quantile of the distribution of median l_1 distances between the combinatorial codes within the family². In other words, let $A_i^k, A_j^k \in \{0, 1\}^{385}$ be the combinatorial codes for molecules i, j in family k . Then the agreement score Q_k for family k is given by

$$Q_k(\alpha) = \inf_q \{ \forall i \in k : P(\hat{m}\{\|A_i^k - A_j^k\|_1\} \leq q) > \alpha \} \quad (1)$$

where $\hat{m}$ denotes the median through $j \in k$. When considering α percent of the most consistent codes in family k , $Q_k(\alpha)$ represents the largest deviation of OR responses in that family. We report several values of α due to the known ambiguity of odor descriptors which depend on several cultural and personal factors of the panelists such as varying verbal expressions, current state and previous experience, or are even affected by OR mutations (Kaepler & Mueller, 2013).

Fig. 2 reports the distribution of $Q_k(\alpha)$ for all abstract odor families and several levels α . Most of the families have a consistent combinatorial code as illustrated by the density of Q_k (Fig. 2), highlighting

²There are 205 abstract odor families and 385 human ORs.

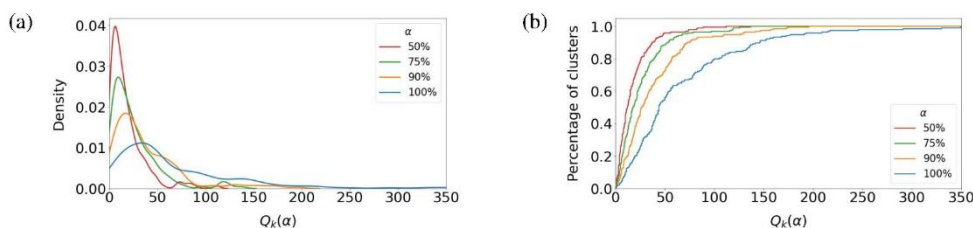


Figure 2: (a) Density and (b) cumulative distribution of agreement scores $Q_k(\alpha)$ for several levels α .

an agreement between the results of the two independent models: odor prediction and OR activation. For $\alpha = 75\%$, more than 60% of all families differ by less than 25 ORs and 84% by less than 50. Fig. 5 provides examples of families with different α -quantile values. Although the odor prediction model outperforms a median panelist, it is still prone to error. This may influence the quality of the clustering and consequently lead to outliers in the families, as illustrated in Fig. 5 (b).

A previous *in vitro* study on 125 odorants suggested that molecules recognized by a subset of ORs share the same odor quality (Nara et al., 2011). The consistency of activation patterns observed in the odor families allows us to generalize this experimental observation to all known odorants. Nara et al. (2011) found that most ORs are narrowly tuned with the exception of few broadly tuned ORs. Our predictions are consistent with this observation, as the results conclude that the vast majority of ORs have a very narrow recognition spectrum (Fig. 4). In addition, our model is able to identify specialized receptors for some specific odor families (e.g. Fig. 6). In agreement with Nara et al. (2011), odor is generally encoded in a combination of multiple active receptors and the size of the code varies among odorants.

7 CONCLUSION

In this study, we transfer the biological problem of protein-molecule interaction to a graph-level binary classification. We combine [CLS] token from the protein language model with graph neural networks and multi-head attention to predict active compounds for ORs. We design a tailored GNN architecture by incorporating inductive biases that are observed in receptor-ligand interactions and we demonstrate that this architecture outperforms other baselines and previous works. Our results show that [CLS] token contains valuable information to characterize proteins and at the same time lowers computational and memory demands. In addition, by taking into account non-bonded interactions, the model is suited to deal with mixtures of compounds.

Although [CLS] token reduces computational costs, it comes with the limitation that one cannot infer the importance of individual amino acids. In the future, we plan to explore the possibility to use MSA-based protein language models to explicitly model amino acids. This may lead to better interpretability with the drawback of higher computational demands. While our predictions are in good agreement with experimental results, our model does not take into account the influence of the concentration on either the perceived odor or the pattern of activated ORs. Recent experiments have shown that inhibition of ORs by odorants is also part of the combinatorial code of odors and, more specifically, that the interaction between odorants in a mixture modulates the activation of ORs relative to that of isolated molecules (Kurian et al., 2021). However, data on OR inhibition remain scarce and do not allow training a model predicting molecular inhibition.

To date, 43% of the ORs are still orphans. The prediction of the combinatorial code for all ORs and odorant molecules opens the way to the generalization of experimental results obtained on subsets of ORs or a limited number of odorants. By linking activity patterns to olfactory perception, we confirm that odorants are recognized by a unique subset of ORs and that these subsets are specific to an olfactory characteristic. Finally, we expect this work to have an impact beyond the olfaction research. ORs are also expressed in non-nasal tissues and are involved in the regulation of different metabolic functions or found in cancerous cells. Thus, they are becoming a promising drug targets, and matching ligand to specific ORs is of major importance.

Published as a conference paper at ICLR 2023

ACKNOWLEDGMENTS

We thank Jana Bocková and Jody Pacalon for helpful feedback on drafts of this paper and for insightful discussions on the activation mechanism of ORs. This work was funded by the French National Research Agency (ANR) under reference number ANR-19-107-0044 (PhD fellowship to Maja Hladik), by the Fondation Rouzniska under the auspices of Fondation de France (PhD fellowship to Mance Lalić), by GIRA2T, and it has also received financial support from the CNRS through the MITI interdisciplinary program. This work was granted access to the HPC resources of IDRIS under the allocation 2022-[AD010713580] made by GENCI and it was also supported by the French government, through the ICAJEDI Investments in the Future project managed by the ANR under reference number ANR-15-IDEX-01. The authors are grateful to the OPAL infrastructure from Université Côte d’Azur and the Université Côte d’Azur’s Center for High-Performance Computing for providing resources and support.

REFERENCES

Ricardo C. Arancada, Zita Peterlin, Xinmin Zhang, Alex Chesler, and Stuart Firestein. A pharmacological profile of the aldehyde receptor repertoire in rat olfactory epithelium. *The Journal of Physiology*, 555(3):743–756, 2004. doi: <https://doi.org/10.1113/jphysiol.2003.058040>. URL <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.2003.058040>.

Jimmy Ba, Jamie R. Kiros, and Geoffrey E. Hinton. Layer normalization. *ArXiv*, abs/1607.06450, 2016. URL <https://arxiv.org/abs/1607.06450>.

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Flores Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, H. Francis Song, Andrew J. Ballard, Justin Gilmer, George E. Dahl, Ashish Vaswani, Kelsey R. Allen, Charles Nash, Victoria Lammington, Chris Dyer, Nicolas Heess, Duan Wierstra, Pushmeet Kohli, Matthew M. Bortvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks. *CoRR*, abs/1806.01261, 2018. URL <http://arxiv.org/abs/1806.01261>.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL <http://github.com/google/jax>.

Linda Buck and Richard Axel. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, 65(1):175–187, April 1991.

Jason B. Castro, Travis J. Gould, Robert Pellegrino, Zhiwei Liang, Liyah A. Coleman, Farnesh Patel, Derek S. Wallace, Tanushri Bhatnagar, Joel D. Mainland, and Richard C. Gerkin. Pyrfume: A win down to the world’s olfactory data. *BioRxiv*, 2022. doi: <https://doi.org/10.1101/2022.09.28.507170>. URL <https://www.biorxiv.org/content/early/2022/09/28/2022.09.28.507170>.

Kyungyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dumitru Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation, 2014.

Xiaojing Cong, Wenwen Ren, Jody Pacalon, Rui Xu, Lun Xu, Xuesen Li, Claire A. de March, Hiroaki Matsumani, Hongmeng Yu, Yiqun Yu, and Jérôme Golchowski. Large scale G protein coupled olfactory receptor–ligand pairing. *ACS Central Science*, 8(3):379–387, 2022. doi: <https://doi.org/10.1021/acscentsci.1c01495>. URL <https://doi.org/10.1021/acscentsci.1c01495>.

Mindy I. Davis, Jeremy P. Hunt, Sanna Hergard, Pietro Cicci, Lisa M. Wodicka, Gabriel Pallares, Michael Hocker, Daniel K. Treiber, and Patrick P. Zarrinkar. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.*, 29(11):1046–1051, October 2011.

Claire A. de March, Jérémie Topin, Elise Bruguera, Gleb Novikov, Kentaro Ikegami, Hiroaki Matsumani, and Jérôme Golchowski. Odorant receptor 7D4 activation dynamics. *Angew. Chem. Int. Ed Engl.*, 57(17):4554–4558, April 2018.

Published as a conference paper at ICLR 2023

Kathrin Kaepler and Friedrich Mueller. Odor classification: A review of factors influencing perception-based odor arrangements. *Chemical Senses*, 38(3):189–209, 01 2013. ISSN 0379-864X. doi: <https://doi.org/10.1093/chemeset/341>. URL <https://doi.org/10.1093/chemeset/341>.

Kentaro Kajiya, Koichiro Inaki, Motonari Tanaka, Tatsuya Hagi, Hiroshi Katakuchi, and Kazushige Tsuchida. Molecular bases of odor discrimination: Reconstitution of olfactory receptors that recognize overlapping sets of odors. *Journal of Neuroscience*, 21(16):6018–6025, 2001. ISSN 0270-6474. doi: <https://doi.org/10.1523/JNEUROSCI.121-16.2001>. URL <https://www.jneurosci.org/content/21/16/6018>.

Sayako Katada, Takatsugu Hirokawa, Yuki Oka, Makiko Sawa, and Kazushige Tsuchida. Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: Mapping the odorant-binding site. *Journal of Neuroscience*, 25(7):1806–1815, 2005. ISSN 0270-6474. doi: <https://doi.org/10.1523/JNEUROSCI.4723-04.2005>. URL <https://www.jneurosci.org/content/25/7/1806>.

Andreas Keller, Hanyi Zhuang, Qiuyi Chi, Leslie B. Vosshall, and Hiroaki Matsumani. Genetic variation in a human odorant receptor alters odor perception. *Nature*, 449(7161):468–472, Sep 2007. ISSN 1476-4687. doi: <https://doi.org/10.1038/nature06162>. URL <https://doi.org/10.1038/nature06162>.

Sunghwan Kim, Jie Chen, Tiejun Cheng, Asti Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A. Shoemaker, Paul A. Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E. Bolton. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research*, 49(D1):D1388–D1395, 11 2020. ISSN 0305-1048. doi: <https://doi.org/10.1093/nar/gkaa971>. URL <https://doi.org/10.1093/nar/gkaa971>.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015. Conference Track Proceedings*, 2015. URL <https://arxiv.org/abs/1412.6980>.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *CoRR*, abs/1609.02907, 2016. URL <https://arxiv.org/abs/1609.02907>.

Justin Klekota and Frederick P. Roth. Chemical substructures that enrich for biological activity. *Bioinformatics*, 24(21):2518–2525, November 2008.

Joel Kowalewski and Anandasankar Ray. Predicting human olfactory perception from activities of odorant receptors. *iScience*, 23(8):101361, 2020. ISSN 2589-0042. doi: <https://doi.org/10.1016/j.isci.2020.101361>. URL <https://www.sciencedirect.com/science/article/pii/S2589004220300484>.

Smajda M. Kurian, Ratiella G. Naresi, Diogo Manoel, Ann-Sophie Barwich, Bettina Malnic, and Luis R. Saravia. Odor coding in the mammalian olfactory epithelium. *Cell Tissue Res*, 383(1): 445–456, January 2021.

Greg Landrum. Rdkit: Open-source cheminformatics software, 2020. URL https://github.com/rdkit/rdkit/releases/tag/release_2020_09_3.

Brian K. Lee, Emily J. Mayhew, Benjamin Sanchez-Lengeling, Jennifer N. Wei, Wesley W. Qian, Kelsie Little, Matthew Andres, Brimley B. Nguyen, Theresa Moloy, Jane K. Parker, Richard C. Gerkin, Joel D. Mainland, and Alexander B. Wiltschko. A principal odor map unifies diverse tasks in human olfactory perception. *BioRxiv*, 2022. doi: <https://doi.org/10.1101/2022.09.01.504602>. URL <https://www.biorxiv.org/content/early/2022/09/01/2022.09.01.504602>.

Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks, 2015. URL <https://arxiv.org/abs/1511.03433>.

Xinyi Liu, Xubo Su, Fei Wang, Zhimin Huang, Qi Wang, Zhen Li, Ruihan Zhang, Lifang Wu, Yi Pan, Yingyi Chen, Hanyi Zhuang, Guoqiang Chen, Ting Shi, and Jian Zhang. OR2ORactor: a web server for deciphering olfactory coding. *Bioinformatics*, 27(16):2302–2303, 06 2011. ISSN 1367-4803. doi: <https://doi.org/10.1093/bioinformatics/btr585>. URL <https://doi.org/10.1093/bioinformatics/btr585>.

Published as a conference paper at ICLR 2023

Daniela C. Gonzalez-Kristeller, João B. P. do Nascimento, Pedro A. F. Galante, and Bettina Malnic. Identification of agonists for a group of human odorant receptors. *Premiers in Pharmacology*, 6, 2015. ISSN 1663-9812. doi: <https://doi.org/10.3389/fphar.2015.00035>. URL <https://www.frontiersin.org/article/10.3389/fphar.2015.00035>.

Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.*, 18(5-6):602–610, June 2005.

Ria Gupta, Ayusha Mittal, Vishesh Agrawal, Sushant Gupta, Krishan Gupta, Rishi Raj Jain, Pratik Garg, Narjaya Kumar Mohanty, Riya Sogani, Harsh Singh Chhabra, Vrshakha Gautam, Tripi Mishra, Debarka Sengupta, and Gautam Ahuja. Olfactory: A conglomerate of artificial intelligence-driven prediction engines for olfactory decoding. *Journal of Biological Chemistry*, 297(2): 100956, 2021. ISSN 0021-9258. doi: <https://doi.org/10.1016/j.jbc.2021.100956>. URL <https://www.sciencedirect.com/science/article/pii/S0021925821007360>.

Franziska Haag, Sandra Hoffmann, and Dietmar Krautwurst. Key food furanones furanone and so-tolone specifically activate distinct odorant receptors. *Journal of Agricultural and Food Chemistry*, 69(37):10999–11005, 2021. doi: <https://doi.org/10.1021/acs.jafc.1c03314>. URL <https://doi.org/10.1021/acs.jafc.1c03314>.

Franziska Haag, Antonella Di Pizio, and Dietmar Krautwurst. The key food odorant receptive range of broadly tuned receptor OR2W1. *Food Chemistry*, 375:131680, 2022. ISSN 0308-8146. doi: <https://doi.org/10.1016/j.foodchem.2021.131680>. URL <https://www.sciencedirect.com/science/article/pii/S0308814621026688>.

Jonathan Heek, Anselm Levskaya, Avital Oliver, Marvin Ritter, Bertrand Rondepierre, Andreas Steiner, and Marc van Zee. Flax: A neural network library and ecosystem for JAX, 2020. URL <http://github.com/google/flax>.

Kevin Huang, Cao Xiao, Lucas M. Glass, and Jiming Sun. MolTrans: Molecular interaction transformer for drug-target interaction prediction. *Bioinformatics*, 37(6):830–836, 10 2020. ISSN 1367-4803. doi: <https://doi.org/10.1093/bioinformatics/btaa880>. URL <https://doi.org/10.1093/bioinformatics/btaa880>.

Amara Jaheen, Claire A. de March, Hiroaki Matsumani, and Sheba Ranganathan. Machine learning assisted approach for finding novel high activity agonists of human ectopic olfactory receptors. *International Journal of Molecular Sciences*, 22(21), 2021. ISSN 1422-0067. doi: <https://doi.org/10.3390/ijms22111546>. URL <https://www.mdpi.com/1422-0067/22/21/11546>.

Valérie Jacquier, Horst Pick, and Horst Vogel. Characterization of an extended receptive ligand repertoire of the human olfactory receptor OR17-40 comprising structurally related compounds. *Drug Neurochemistry*, 97(2):537–544, 2006. doi: <https://doi.org/10.1111/j.1471-4159.2006.03771.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-4159.2006.03771.x>.

Éric M. Jones, Rishi Jhoo, Daniel Cancelli, Nathan B. Lubock, Jeffrey Wang, Megan Satyadi, Rocky Cheung, Claire de March, Joshua S. Bloom, Hiroaki Matsumani, and Brian Kouri. A scalable, multiplexed assay for decoding GPCR–ligand interactions with RNA sequencing. *Cell Systems*, 8(3): 254–260.e6, 2019. ISSN 2405-4712. doi: <https://doi.org/10.1016/j.cels.2019.02.009>. URL <https://www.sciencedirect.com/science/article/pii/S2405471219300703>.

John Junger, Richard Evans, Alexander Fritz, Tim Green, Michael Higgins, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zidek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romero-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michael Zielinski, Martin Steinegger, Michael Paschold, Danas Berghammer, Sebastian Rodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Kory Kavukcuoglu, Pushmeet Kohli, and Denis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, Aug 2021. ISSN 1476-4687. doi: <https://doi.org/10.1038/s41586-021-03819-2>. URL <https://doi.org/10.1038/s41586-021-03819-2>.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. doi: <https://doi.org/10.18653/v1/N19-1423>. URL <https://aclanthology.org/N19-1423>.

Bing-Xue Du, Yuan Qin, Yan-Feng Jiang, Yi Xu, Siu-Ming Yiu, Hai Yu, and Jian-Yu Shi. Compound-protein interaction prediction by deep learning: Databases, descriptors and models. *Drug Discovery Today*, 27(5):1350–1366, 2022. ISSN 1359-6446. doi: <https://doi.org/10.1016/j.drudis.2022.02.023>. URL <https://www.sciencedirect.com/science/article/pii/S1359644622000818>.

Ahmed Elmaghrabi, Michael Heinzeinger, Christian Dallago, Ghali Rehawi, Yi Wang, Lijon Jones, Tom Gibbs, Tamas Fehér, Christoph Angerer, Martin Steinegger, Debanshi Bhowmik, and Burkhard Rost. ProtTrans: Towards cracking the language of life code through self-supervised deep learning and high performance computing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021. doi: <https://doi.org/10.1109/TPAMI.2021.3095381>.

Richard Evans, Michael O’Neill, Alexander Fritz, Natasha Antropova, Andrew Senior, Tim Green, Augustin Zidek, Russ Bates, Sam Blackwell, Jason Yin, Olaf Ronneberger, Sebastian Rodenstein, Michael Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Junger, and Denis Hassabis. Protein complex prediction with alphafold-multimer. *BioRxiv*, 2021. doi: <https://doi.org/10.1101/2021.10.04.463034>. URL <https://www.biorxiv.org/content/early/2021/10/04/2021.10.04.463034>.

Tim Frey, Charles A. Kwahia, Erika A. Wallin, Elsa Holgersson, Erik Hedström, Björn Bohman, Marie Bengtsson, Paul G. Becker, Dietmar Krautwurst, and Peter Witzgall. The human odorant receptor OR10A6 is tuned to the pheromone of the commensal fruit fly *Drosophila melanogaster*. *BioRxiv*, 2020. doi: <https://doi.org/10.1101/2020.12.07.414714>. URL <https://www.biorxiv.org/content/early/2020/12/07/2020.12.07.414714>.

Yoshihumi Fujita, Tomoko Takahashi, Akiko Suzuki, Kayo Kawashima, Futoshi Nara, and Ryuta Koishi. Desorption of drosophila G protein-coupled receptor for an odorant receptor. *Journal of Receptors and Signal Transduction*, 27(4):323–334, 2007. doi: <https://doi.org/10.1080/10799890701534180>. URL <https://doi.org/10.1080/10799890701534180>.

Johannes Gasteiger, Shankari Giri, Johannes T. Margraf, and Stephan Günemann. Fast and uncertainty-aware directional message passing for non-equilibrium molecules, 2020a. URL <https://arxiv.org/abs/2011.14115>.

Johannes Gasteiger, Janek Groß, and Stephan Günemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020b. URL <https://openreview.net/forum?id=21c9kx2F7E>.

Christiane Geithe, Gaby Andersen, Agne Malki, and Dietmar Krautwurst. A butter aroma recombinant activates human class I odorant receptors. *Journal of Agricultural and Food Chemistry*, 63(43): 9410–9420, 2015. doi: <https://doi.org/10.1021/acs.jafc.5b01884>. URL <https://doi.org/10.1021/acs.jafc.5b01884>.

Christiane Geithe, Franziska Noc, Johanna Kreissl, and Dietmar Krautwurst. 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. *Chemical Senses*, 42(3):181–193, 12 2016. ISSN 0379-864X. doi: <https://doi.org/10.1093/chemeset/bjw117>. URL <https://doi.org/10.1093/chemeset/bjw117>.

Christiane Geithe, Jonas Probst, Franziska Kerschwig, Gerd Krause, and Dietmar Krautwurst. Structural determinants of a conserved enantioselective carvone binding pocket in the human odorant receptor OR1A1. *Cellular and Molecular Life Sciences*, 74(22):4209–4229, 2017.

Justin Gilmer, Samuel S. Schoenholz, Patrick L. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML’17*, pp. 1263–1272, 2017.

Published as a conference paper at ICLR 2023

Joel D. Mainland, Andreas Keller, Yun R. Li, Ting Zhou, Casey Trimmer, Lindsey L. Snyder, Andrew H. Meyerle, Kaylin A. Adipietro, Wen Ling L. Liu, Hanyi Zhuang, Seungho Zhao, Sonmin S. Lee, Abigail Liu, and Hiroaki Matsunami. The mis-sense of smell: functional variability in the human odorant receptor repertoire. *Nature Neuroscience*, 17(1):114–120, Jan 2014. ISSN 1546-1726. doi: 10.1038/nn.3598. URL <https://doi.org/10.1038/nn.3598>.

Joel D. Mainland, Yun R. Li, Ting Zhou, Wen Ling L. Liu, and Hiroaki Matsunami. Human olfactory receptor responses to odorants. *Sci. Data*, 2(1):150002, February 2015.

Bettina Malaic, Junzo Hirono, Takaaki Sato, and Linda B. Buck. Combinatorial receptor codes for odors. *Cell*, 96(5):713–723, March 1999.

Patrick Marcinek, Franziska Haag, Christine Geithe, and Dietmar Krautwurst. An evolutionary conserved olfactory receptor for foodborne and semiochemical alkylpyrazines. *The FASEB Journal*, 35(6):e21638, 2021. doi: <https://doi.org/10.1096/fj.202100224R>. URL <https://www.onlinelibrary.wiley.com/doi/abs/10.1096/fj.202100224R>.

Andrea Mauri, Viviana Consonni, Manuela Pavan, and Roberto Todeschini. DRAGON software: An easy approach to molecular descriptor calculations. volume 56, pp. 237–248, 2006.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2018. URL <https://arxiv.org/abs/1802.03426>.

IL L. Morgan. The generation of a unique machine description for chemical structures—A technique developed at chemical abstracts service. *Journal of Chemical Documentation*, 5(2):107–113, 1965. doi: 10.1021/c160017a018. URL <https://doi.org/10.1021/c160017a018>.

Kiyomitsu Nara, Luis R Saravia, Xiaolan Ye, and Linda B. Buck. A large-scale analysis of odor coding in the olfactory epithelium. *J. Neurosci.*, 31(25):9179–9191, June 2011.

Eva M. Neuhaus, Anasudha Mashukova, Weiwei Zhang, Joe Harbour, and Hanne Hart. A specific heat shock protein enhances the expression of mammalian olfactory receptor proteins. *Chemical Senses*, 31(5):445–452, 03 2006. ISSN 0379-864X. doi: 10.1093/chemse/bhj049. URL <https://doi.org/10.1093/chemse/bhj049>.

Yoshihito Niumura and Masatoshi Nei. Evolution of olfactory receptor genes in the human genome. *Proceedings of the National Academy of Sciences*, 100(21):12235–12240, 2003. doi: 10.1073/pnas.1635157100. URL <https://www.pnas.org/doi/abs/10.1073/pnas.1635157100>.

Franziska Nue, Johannes Pulster, Christiane Geith, Matthias Kolbhoff, Peter Schieberle, and Dietmar Krautwurst. OR2M3: A highly specific and narrowly tuned human odorant receptor for the sensitive detection of onion key food odorant 3-methylpentan-1-ol. *Chemical Senses*, 42(3):195–210, 12 2016. ISSN 0379-864X. doi: 10.1093/chemse/bjw118. URL <https://doi.org/10.1093/chemse/bjw118>.

Franziska Nue, Tim Frey, Julia Fiedler, Christiane Geithe, Bettina Nowak, and Dietmar Krautwurst. Il-6. *Journal of Biological Methods*, 4(4):e81, Nov. 2017. doi: 10.14440/jbm.2017.206. URL <https://jbmchods.org/jbm/article/view/206>.

Tejvija Olander, Doron Lancet, and Daniel W Nebert. Update on the olfactory receptor (OR) gene superfamily. *Hum. Genomics*, 3(1):87–97, September 2008.

Yitzhak Pilpel and Doron Lancet. The variable and conserved interfaces of modeled olfactory receptor proteins. *Protein Science*, 8(5):969–977, 1999. doi: <https://doi.org/10.1110/ps.8.5.969>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1110/ps.8.5.969>.

Ernest H. Polak. Multiple profile-multiple receptor site model for vertebrate olfaction. *Journal of Theoretical Biology*, 40(3):469–484, 1973. ISSN 0022-5193. doi: [https://doi.org/10.1016/0022-5193\(73\)90005-2](https://doi.org/10.1016/0022-5193(73)90005-2). URL <https://www.sciencedirect.com/science/article/pii/0022519373900052>.

14

Published as a conference paper at ICLR 2023

Kristin Schmiedberg, Elena Shirokova, Hans-Peter Weber, Boris Schilling, Wolfgang Meyerhoff, and Dietmar Krautwurst. Structural determinants of odorant recognition by the human olfactory receptors OR1A1 and OR1A2. *Journal of Structural Biology*, 159(3):400–412, September 2007. ISSN 1047-8477. doi: 10.1016/j.jsb.2007.04.013. URL <https://www.sciencedirect.com/science/article/pii/S1047847707000872>.

Charles S. Sell. On the unpredictability of odor. *Angew. Chem. Int. Ed Engl.*, 45(38):6254–6261, September 2006.

Anju Sharma, Bishal Kumar Saha, Rajnish Kumar, and Pritish Kumar Varadwaj. OlfactionBase: a repository to explore odors, receptors, olfactory receptors and odorant-receptor interactions. *Nucleic Acids Research*, 50(D1):D678–D686, 09 2021. ISSN 0958-1048. doi: 10.1093/nar/gkab763. URL <https://doi.org/10.1093/nar/gkab763>.

Miki Shirao, Keiichi Yoshikawa, Yoshiki Takai, Ai Nakashima, Haruki Takeuchi, Hitoshi Sakano, and Kazushige Tsubara. Olfactory receptor and neural pathway responsible for highly selective sensing of musk odors. *Neuron*, 81(1):165–178, 2014. ISSN 0896-6273. doi: <https://doi.org/10.1016/j.neuron.2013.10.021>. URL <https://www.sciencedirect.com/science/article/pii/S0896627313009322>.

Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2(2):021001, Jan 2021. doi: 10.1088/2632-2153/ab6b9a. URL <https://doi.org/10.1088/2632-2153/ab6b9a>.

Martin Simonovsky and Nikos Komodakis. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 29–38, Jul 2017. doi: 10.1109/CVPR.2017.11.

Hwanjun Song, Minseok Kim, Dongmin Park, Yoojin Shin, and Jae-Gil Lee. Learning from noisy labels with deep neural networks: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–19, 2022. doi: 10.1109/TNNLS.2022.3152521.

Marc Spehr, Günter Gisselmann, Alexandra Poplawski, Jeffrey A. Riffell, Christian H. Wetzel, Richard K. Zimmer, and Hanne Hart. Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science*, 299(5615):2054–2058, 2003. doi: 10.1126/science.1080376. URL <https://www.sciencemag.org/doi/abs/10.1126/science.1080376>.

Martin Steinegger and Johannes Söding. Clustering huge protein sequence sets in linear time. *Nature Communications*, 9(1):2542, Jun 2018. ISSN 2041-7233. doi: 10.1038/s41467-018-04964-5. URL <https://doi.org/10.1038/s41467-018-04964-5>.

Martin Steinegger, Milot Mirdita, and Johannes Söding. Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. *Nature Methods*, 16(7):603–606, Jul 2019. ISSN 1548-7095. doi: 10.1038/s41592-019-0437-4. URL <https://doi.org/10.1038/s41592-019-0437-4>.

Baris E. Suzek, Hongzhan Huang, Peter McGarvey, Raja Mazumder, and Cathy H. Wu. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, 23(10):1282–1288, May 2007.

Jing Tang, Agnieszka Szwajda, Soshi Shikayama, Tao Xu, Peter H Harsanyi, Kristi Wennerberg, and Tero Aittokallio. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J. Chem. Inf. Model.*, 54(3):735–743, March 2014.

Wen Tong and Russ B. Altman. Graph convolutional neural networks for predicting drug-target interactions. *Journal of Chemical Information and Modeling*, 59(10):4131–4149, 2019. doi: 10.1021/acs.jcim.9b00628. URL <https://doi.org/10.1021/acs.jcim.9b00628>.

Casey Trimmer, Andreas Keller, Nicole R. Murphy, Lindsey L. Snyder, Jason R. Miller, Maira H. Negri, Nicholas Katsaris, Leslie B. Vosshall, Hiroaki Matsunami, and Joel D. Mainland. Genetic variation across the human olfactory receptor repertoire alters odor perception. *Proc. Natl. Acad. Sci. U. S. A.*, 116(19):9475–9480, May 2019.

16

Published as a conference paper at ICLR 2023

Tianyi Qiu, Jingxuan Qiu, Jun Feng, Dingfeng Wu, Yiyun Yang, Kailin Tang, Zhiwei Cao, and Ruixin Zhu. The recent progress in proteinchemometric modelling: focusing on target descriptors, cross-term descriptors and application scope. *Brief. Bioinform.*, 18(1):125–136, January 2017.

Roshan Rao, Nicholas Bhattacharya, Neil Thomas, Yun Duan, Xi Chen, John Canny, Pieter Abbeel, and Yun S. Song. Evaluating protein transfer learning with TAPE. *Adv. Neural Inf. Process. Syst.*, 32:9689–9701, December 2019.

Roshan Rao, Jason Liu, Robert Verkuil, Joshua Meier, John Canny, Pieter Abbeel, Tom Sercu, and Alexander Rives. MSA transformer. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8844–8856, Jul 2021a. URL <https://proceedings.mlr.press/v139/rao21a.html>.

Roshan Rao, Joshua Meier, Tom Sercu, Sergey Ovchinnikov, and Alexander Rives. Transformer protein language models are unsupervised sequence learners. In *International Conference on Learning Representations*, May 2021b. URL <https://openreview.net/forum?id=zyicEggyva>.

Sarah E. Repicky and Charles W. Lucife. Molecular receptive range variation among mouse odorant receptors for aliphatic carboxylic acids. *Journal of Neurochemistry*, 109(1):193–202, 2009. doi: <https://doi.org/10.1111/j.1471-4159.2009.05925.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-4159.2009.05925.x>.

Alexander Rives, Joshua Meier, Tom Sercu, Siddhant Goyal, Zeming Lin, Jason Liu, Dersi Cao, Myk Oll, C. Lawrence Zitnick, Jerry Ma, and Rob Fergus. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15), 2021. ISSN 0027-8424. doi: 10.1073/pnas.2016239118. URL <https://www.pnas.org/content/118/15/2016239118>.

Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. In *Advances in Neural Information Processing Systems*, volume 33, pp. 12559–12571, 2020. URL <https://proceedings.nips.cc/paper/2020/file/94ae13541efca3389a3bed3fa1c199d5d-Paper.pdf>.

Harumi Saito, Monoka Kubota, Richard W. Roberts, Qiuyi Chi, and Hiroaki Matsunami. RTP family members induce functional expression of mammalian odorant receptors. *Cell*, 119(5): 679–691, 2004. ISSN 0092-8674. doi: <https://doi.org/10.1016/j.cell.2004.11.021>. URL <https://www.sciencedirect.com/science/article/pii/S009286740400693X>.

Harumi Saito, Qiuyi Chi, Hanyi Zhuang, Hiroaki Matsunami, and Joel D. Mainland. Odor coding by a mammalian receptor repertoire. *Science Signaling*, 2(60):ra9–ra10, 2009. doi: 10.1126/signal.2009016. URL <https://www.sciencemag.org/doi/abs/10.1126/scisig.ra9>.

Benjamin Sanchez-Lengeling, Jennifer N. Wei, Brian K. Lee, Richard C. Gerkin, Alán Aspuru-Guzik, and Alexander B. Wittke. Machine learning for scent: Learning generalizable perceptual representations of small molecules, 2019. doi: 10.26434/chemrxiv-2019-10665. URL <https://arxiv.org/abs/1910.10665>.

Guenhael Sanz, Claire Schlegel, Jean Claude Pernollet, and Loïc Briand. Comparison of odorant specificity of two human olfactory receptors from different phylogenetic classes and evidence for antagonism. *Chemical Senses*, 30(1):69–80, 01 2005. ISSN 0379-864X. doi: 10.1093/chemse/bji002. URL <https://doi.org/10.1093/chemse/bji002>.

Narumi Sato-Akamura, Nao Horio, Aya Kato-Nambu, Keiichi Yoshikawa, Yoshihito Niumura, Svyoko Phara, Mika Shirao, and Kazushige Tsubara. Ligand specificity and evolution of mammalian musk odor receptors: Effect of single receptor deletion on odor detection. *Journal of Neuroscience*, 36(16):4482–4491, 2016. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1259-15.2016. URL <https://www.jneurosci.org/content/36/16/4482>.

15

Published as a conference paper at ICLR 2023

UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47 (D1):D506–D515, January 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL <https://proceedings.nips.cc/paper/2017/file/3d96f23494dee9bfb3e063c583e3aa-Paper.pdf>.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL <https://openreview.net/forum?id=xGC1kPz9Qz>.

Jesse Vig, Ali Madani, Lav R. Varshney, Caiming Xiong, Richard Socher, and Nazneen F. Rajani. BIOLOGY meets biology: Interpreting attention in protein language models. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=yvkt_zvfmcd7.

X. Wang, R. Girshick, A. Gupta, and K. He. Non-local neural networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803, 2018. doi: 10.1109/CVPR.2018.00813.

Christian H. Wetzel, Markus Oles, Christiane Welterdieck, Michael Kuczkowski, Günter Gisselmann, and Hanne Hart. Specificity and sensitivity of a human olfactory receptor functionally expressed in human embryonic kidney 293 cells and xenopus laevis oocytes. *Journal of Neuroscience*, 19 (17):7426–7433, 1999. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.19-17.07426.1999. URL <https://www.jneurosci.org/content/19/17/7426>.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=zyG6B1R5Gn>.

Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hus Gao, Angel Guzman-Perez, Timothy Hopper, Brian Kelley, Miriam Mathea, Andrew Palmer, Volker Settel, Tommi Jaakkola, Klavs Jensen, and Regina Barzilay. Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8):3370–3388, 2019. doi: 10.1021/acs.jcim.9b00237. URL <https://doi.org/10.1021/acs.jcim.9b00237>.

Emily A. Yasi, Sara L. Ewen, Hanfei Wang, Widiati Sugianto, Anita R. Mimicfield, Kaitlyn A. Hoover, Paul J. Brauhm, and Pamela Peralta-Yahya. Rapid deorphanization of human olfactory receptors in yeast. *Biochemistry*, 58(16):2160–2166, 2019. doi: 10.1021/acs.biochem.8b01208. URL <https://doi.org/10.1021/acs.biochem.8b01208>.

Huajing Zhao, Lidija Ivic, Joji M. Oaki, Mitsuhiko Hashimoto, Katsuhiko Mikoshiba, and Stuart Firestein. Functional expression of a mammalian odorant receptor. *Science*, 279(5348):237–242, January 1998.

Qichang Zhao, Haochen Zhao, Kai Zheng, and Jianxin Wang. HyperAttentionDTI: improving drug-protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics*, 38(3):655–662, October 2021.

Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications, 2018. URL <https://arxiv.org/abs/1812.10643>.

17

A MODEL DETAILS

In all our experiments, we consider node and bond features in Tab. 5. Molecular graphs were constructed using RDkit (Landrum, 2020) from SMILES obtained from PubChem (Kim et al., 2020). Atomic number, chiral tag and hybridization were embedded in $\mathbb{R}^{72}$, and Bond type and Stereo type in $\mathbb{R}^{36}$.

Table 5: Features of the molecular graph. Features were treated as categorical variables except for *formal charge*, *number of implicit Hs*, *explicit valence* and *mass* which are continuous variables.

Atom features	Bond features
Atomic number	Bond type
Chiral tag	Stereo type
Hybridization	Is aromatic
Formal charge	
Num. of implicit Hs	
Explicit valence	
Mass	
Is aromatic	

The attention pooling layer in graph to vector operation for graph g is defined as

$$\alpha = \text{softmax}(Xw) \quad x' = \sum_{i=1}^n \alpha_i X_i \quad (2)$$

where $X \in \mathbb{R}^{n \times d}$ is a matrix of node embeddings and $w \in \mathbb{R}^{d \times 1}$ are trainable parameters.

A.1 TRAINING DETAILS

Dropout of 0.1 was applied in each multi-head attention layer and 0.5 just before the output layer. Padding was set to 32 nodes and 64 edges in the first 8000 epochs and 128 and 256, respectively, in the remaining 2000 epochs. The initial learning rate was set to 0.001 and we use 6000 warm-up steps for the scheduler. The model was implemented in JAX (Bradbury et al., 2018) and FLAX (Heek et al., 2020) and trained on Nvidia V100 SXM2 32GB or Nvidia A100 SXM4 80GB GPUs.

B ENANTIOMERS

The olfactory system of primates is sensitive to the chirality of molecules: they can discriminate between enantiomers. By definition, a pair of enantiomers³ are mirror isomers that are not superimposable chemical structures. They have the same atomic composition and they share identical physical and chemical properties in an achiral environment. However, they will have different properties when interacting with a chiral environment (for example, polarized light or the binding site of a protein). Two enantiomers can have a completely different odor, for example *R*-carvone and *S*-carvone which smell like caraway and mint, respectively. At the molecular level, these two molecules are recognized by distinct OR subsets. The specific selectivity of ORs for a given enantiomer is at the origin of the differences in perception of chiral molecules. Even if chirality has a high influence on ORs recognition, the distinction between enantiomers is often overlooked in laboratory experiments. As a result, 21% of our gathered data are actually pairs of an OR and a mixture of two (or more) enantiomers. Thus, it appears crucial to inject molecular chirality into our model.

Molecular graphs of enantiomers are identical and the only difference is in their chiral centers. These may have clockwise (CW) or counter-clockwise (CCW) chirality and one way of distinguishing them is to put *chiral tag* property of one chiral center to CW and CCW for the other. Since the molecular graphs are identical, a mixture of enantiomers can be modeled in two ways: as a single graph where chiral centers have "unknown" chirality or a mixture of independent molecules. In the latter case, we can exploit the fact that combining GNN with the multi-head attention is capable

³With a slight abuse of notation, we refer to both enantiomers and diastereoisomers as *enantiomers*.

Table 6: Screening quality. Values in each row correspond to the conditional probability that the true response is positive/negative (i.e. the outcome of the dose-response measurement) given the result from the primary and secondary screening. These probabilities were estimated based on the data from Mainland et al. (2015). EC_{50} denotes the true label from the dose-response measurements.

		Primary		Secondary	
		pos.	neg.	pos.	neg.
EC_{50}	pos.	0.40	0.31	0.72	0.23
	neg.	0.60	0.69	0.28	0.77

of modeling mixtures. Thus, a mixture of enantiomers can be treated as a graph with $2^{|chiral\ centers|}$ identical disconnected components that differ only in the *chiral tag* property of chiral centers.

We test both "single" and "mixture" approaches in Section E.1. Our experiments indicate that the two strategies have similar performance on OR-molecule pairs (Tab. 7). Interestingly, when evaluated on a limited set of complex mixtures (i.e. pairs of OR and mixture containing more than two topologically different compounds), the performance of the model trained in the "mixture" approach is higher even though the only mixtures seen during training were the mixtures of enantiomers (Tab. 8). Note that the data on mixtures are limited and further experiments are needed.

C SAMPLE WEIGHTS

Data quality. Conducting an experiment with high confidence on responsiveness is costly and time-consuming. The standard procedure for testing molecule-receptor pairs is to first perform a noisy primary screening where many different pairs are tested with just one injection for a single concentration. Then, promising pairs from the primary screening are further tested at several concentrations in a secondary screening. Finally, the pairs that are considered responsive from the secondary screening are tested in the most precise tertiary screening in multiple concentrations and multiple injections per concentration (i.e. dose-response curve). If neglected, varying sample quality could lead to fitting corrupted labels and impairing the model performance and generalization. Therefore, during training we employ a simple strategy to counteract this label noise. For each type of screening, we estimate probability $P_{c,\hat{c}}^{type}$ that the observed label $\hat{c}$ corresponds to the true label c . By treating a dose-response label as true we estimate $\hat{P}_{c,\hat{c}}^{type}$ based on (Mainland et al., 2015), where several pairs were tested in all three types of experiments (Tab. 6). Then we use the diagonal elements as data quality weights for a given type and a given class y

$$w_{quality}(y, type) = \hat{P}_{y,y}^{type} \quad (3)$$

We set the weight for dose-response samples to 1. In the future, a more elaborated label noise treatment could be used (see (Song et al., 2022) for a non-exhaustive list of label noise strategies).

Class imbalance. Naturally, most of the molecule-receptor pairs are non-responsive and only 6.2% of the samples in our data are responsive. To account for this class imbalance, we use standard class weights based on the imbalance ratio

$$w_{class}(y) = \begin{cases} \frac{|non-responsive|}{|responsive|} & \text{if } y = 1 \\ 1 & \text{if } y = 0 \end{cases} \quad (4)$$

Pair imbalance. On top of the class imbalance, some olfactory receptors have been tested with many molecules, whereas others are yet to be explored and were tested only on a few specific compounds (Fig. 7). We refer to this type of imbalance as *pair imbalance* and to avoid the bias resulting from the uneven distribution of experiments, we adopt a heuristic weighting scheme to down-weight extensively tested receptors and up-weight less-known ones. For each molecule-receptor pair, we set pair imbalance weight as the logarithm of the harmonic mean

$$w_{pair}(m, r) = \ln \left(1 + \frac{K}{2} \left(\frac{1}{|M(r)|} + \frac{1}{|OR(m)|} \right) \right) \quad (5)$$

where $|M(r)|$ is the number of molecules tested for a given receptor r , $|OR(m)|$ is the number of receptors tested for a given molecule m and K is a constant. In the experiments, we set $K = 100$.

The final sample weight is then given by

$$w(m, r, y, type) = w_{quality}(y, type)w_{class}(y)w_{pair}(m, r) \quad (6)$$

D EXPERIMENTAL DETAILS

D.1 GENERALIZATION

Deorphanization. To test the generalization to unseen receptors we consider two scenarios. The first one (referred to as *random* in Tab. 3), where the model predicts active compounds for unseen receptors chosen randomly, and the second, more complex scenario, where entire groups of similar receptors are unseen during the training (*cluster* in Tab. 3).

We test the *random* scenario by constructing a test set for each cross-validation run in the following way. From pairs tested in dose-response experiments, we filter receptors with at least 3 known active compounds. Then from these filtered receptors we randomly choose 35% and place all dose-response pairs containing these into the test set. We then discard remaining primary and secondary screening pairs containing the selected receptors from the training set (OR in Tab. 3) or we discard only the responsive pairs containing these receptors (OR - keep in Tab. 3). The reasoning behind keeping negative examples is that, in reality, we have access to non-responsive molecules from primary screenings for all human olfactory receptors. Following this procedure, the model has never seen the selected receptors and it needs to extrapolate their molecule-receptor interaction from the remaining proteins.

Test set construction in the *cluster* scenario is similar to the *random* case, but we do not choose receptors for the test set uniformly. Instead, we cluster receptors based on their sequence similarity and then place dose-response pairs for all receptors from 9 randomly picked clusters into the test set. We then follow the same procedure of discarding either all or only the responsive primary and secondary screening pairs containing these receptors from the training set. This scenario is more demanding than the first setting because the model would need to extrapolate to entirely dissimilar proteins than the ones in the training set.

Expanding chemical space. Extrapolation to the unseen molecules is tested by an identical procedure as the deorphanization. The only differences are that we select molecules instead of receptors (25% molecules with at least 2 responsive pairs in *random*) and in the *cluster* scenario we cluster molecules based on their Tanimoto coefficient (i.e. structural similarity). We take compounds from 6 randomly selected clusters. In both scenarios, we then discard all other pairs containing the selected compounds from the training set.

E ADDITIONAL EXPERIMENTS

E.1 DATA ABLATIONS

Enantiomers. As we argued previously, olfaction is sensitive to the chirality and modeling enantiomers poses a great challenge. Here, we test the two approaches to incorporate mixtures of enantiomers and their performance is summarized in Tab. 7. We report two settings where we either test the model on single molecules only (i.e. discarding mixtures of enantiomers from the test set) or on the full test set. We also train a separate model on the training set with mixtures of enantiomers discarded. As can be seen in the results, modeling mixtures of enantiomers explicitly as a mix of independent molecules slightly outperforms the implicit "single" approach. In both cases keeping mixtures of enantiomers is beneficial for modeling single molecules.

Complex mixtures. Several lab experiments have been performed to test a receptor with a mixture of topologically different compounds. In Tab. 8, we report the performance of our model on a limited dataset of complex mixtures. In particular, we investigate impact of the two strategies for treating mixtures of enantiomers on the performance on complex mixtures. Although the two approaches

Table 7: Summary of results for strategies to treat mixtures of enantiomers. Mixtures of enantiomers are either modeled as a mixture of distinct molecules (*mixture*) or as a single graph with chiral centers having "unknown" chirality (*single*). *Discard* corresponds to the model trained only on data without the mixtures. *Discard mix of isomers* denotes a test set without the mixtures and *full test set* is a test set containing all pairs, including mixtures of enantiomers.

	Isomers	Num data	AveP	Precision	Recall	F-score	MCC
Discard mix of isomers	Discard	1425.0	0.773 (0.01)	0.666 (0.03)	0.689 (0.03)	0.677 (0.02)	0.589 (0.03)
	Single	1425.0	0.759 (0.02)	0.656 (0.01)	0.715 (0.03)	0.685 (0.02)	0.597 (0.03)
	Mixture	1425.0	0.778 (0.01)	0.681 (0.02)	0.704 (0.04)	0.691 (0.02)	0.608 (0.02)
Full test set	Single	1565.0	0.765 (0.02)	0.665 (0.01)	0.711 (0.02)	0.687 (0.02)	0.595 (0.02)
	Mixture	1565.0	0.780 (0.01)	0.689 (0.02)	0.698 (0.04)	0.693 (0.02)	0.605 (0.02)

Table 8: Summary of results for a small dataset of complex mixture-receptor pairs for dose-response pairs only (EC_{50}) and all available data (*All*).

	Isomers	Num data	AveP	Precision	Recall	F-score	MCC
EC_{50}	Single	37.0	0.938 (0.05)	0.895 (0.04)	0.836 (0.02)	0.864 (0.01)	0.494 (0.10)
	Mixture	37.0	0.994 (0.01)	0.915 (0.07)	0.836 (0.07)	0.869 (0.02)	0.538 (0.10)
All	Single	112.0	0.799 (0.08)	0.716 (0.08)	0.774 (0.03)	0.741 (0.03)	0.638 (0.05)
	Mixture	112.0	0.940 (0.04)	0.789 (0.10)	0.755 (0.06)	0.768 (0.04)	0.684 (0.07)

have identical training sets, the results suggest that the model trained with the "mixture" strategy outperforms the "single" approach.

Data quality weights. Label noise originating from different quality of experiments can bias the model and negatively impact its generalization. To test the label noise treatment, we experiment with discarding primary screening data (the most noisy ones in Tab. 6), with weights presented in Section C, and with ignoring the label noise (i.e. a naive approach). Note that we use only dose-response data for test sets and these are assumed to have no noise.

According to the results in Tab. 9, neglecting the label noise lowers the performance. It can be observed that when using weights and when discarding the primary screening data the results are on par. This is due to the fact that most of the corrupted labels are coming from the primary screening (Tab. 6) and so discarding these cures a large portion of the label noise. Nevertheless, we hypothesize that keeping the primary screening data while taking the noise into account could be beneficial for model generalization. Primary screening experiments have relatively low cost and can cover a large number of different pairs. Thus, the model can have access to a large sample of the data distribution.

Table 9: Summary of results for data quality weights. *Naive* corresponds to equal weights. Mixtures of enantiomers are treated as a "mixture".

Data quality	AveP	Precision	Recall	F-score	MCC
Discard primary	0.754 (0.02)	0.667 (0.03)	0.724 (0.04)	0.694 (0.01)	0.603 (0.02)
Naive	0.740 (0.04)	0.631 (0.05)	0.715 (0.02)	0.669 (0.03)	0.570 (0.04)
Weight	0.780 (0.01)	0.689 (0.02)	0.698 (0.04)	0.693 (0.02)	0.605 (0.02)

E.2 PERFORMANCE ON OTHER DATASETS

In addition to OR-molecule data, we conduct experiments on two drug-target interaction prediction datasets, namely KIBA (Tang et al., 2014) and DAVIS (Davis et al., 2011), and we compare performance of our approach with two state-of-the-art drug-target interaction models HyperAttentionDTI (Zhao et al., 2021) and MolTrans (Huang et al., 2020), which based on Zhao et al. (2021) perform best on the two datasets. Both KIBA and DAVIS record kinase inhibition binding affinities measured in laboratory assays. We follow data preprocessing and binarization scheme from (Zhao et al., 2021).

Table 10: Statistics of DAVIS (Davis et al., 2011), KIBA (Tang et al., 2014) and M2OR (ours) datasets. *Seqs.* and *Mols.* respectively refer to unique number of sequences and molecules in the datasets. *Orphans* is a number of sequences without any known active compound and *% mols. tested per sequence* and *Ligands per sequence* are, respectively, average portion of molecules tested on a given sequence and average number of known active compounds for a sequence.

Dataset	Num. pairs	Seqs.	Mols.	Orphans	% mols. tested per sequence	Ligands per sequence
DAVIS	25 772	379	68	0	100.0%	19.3
KIBA	116 350	225	2068	5	25.0%	98.5
M2OR	46 717	1237	596	527	6.3%	2.3

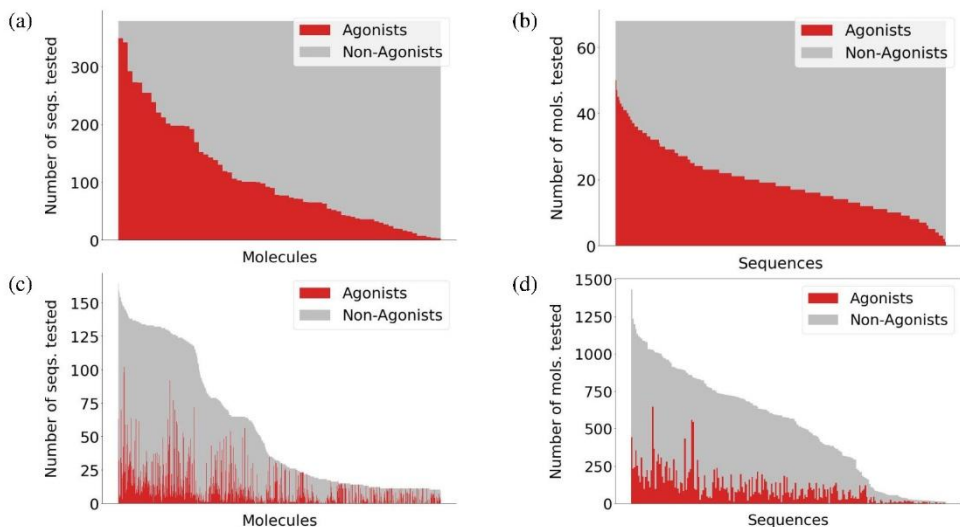


Figure 3: Distributions of the number of laboratory experiments: (a) per molecule and (b) per sequence for DAVIS, (c) per molecule and (d) per sequence for KIBA. Note that DAVIS dataset contains results on all possible pairs of 68 molecules and 379 proteins.

In Tab. 10, Fig. 3 and Fig. 7 a comparison between distributions of the two datasets and of our data (M2OR in Tab. 10) is reported. As can be observed by shapes of the gray areas in Fig. 3 and Fig. 7, OR data have a more scarce distribution of tests per sequence and per molecule and a more imbalanced class distribution than what can be observed in KIBA or DAVIS.

Table 11: Performance of HyperAttentionDTI (Zhao et al., 2021), MolTrans (Huang et al., 2020) and our model on KIBA (Tang et al., 2014), DAVIS (Davis et al., 2011) and M2OR (ours) data. Results for MolTrans and HyperAttentionDTI on KIBA and DAVIS are taken from (Zhao et al., 2021).

	Model	AveP	Precision	Recall	MCC
DAVIS	MolTrans	0.784 (0.002)	0.782 (0.003)	0.617 (0.004)	n/a
	HyperAttentionDTI	0.839 (0.002)	0.754 (0.002)	0.780 (0.001)	n/a
	Ours - mixture	0.744 (0.015)	0.699 (0.013)	0.666 (0.016)	0.564 (0.011)
KIBA	MolTrans	0.708 (0.003)	0.710 (0.003)	0.645 (0.003)	n/a
	HyperAttentionDTI	0.814 (0.002)	0.689 (0.003)	0.798 (0.002)	n/a
	Ours - mixture	0.704 (0.010)	0.610 (0.014)	0.683 (0.005)	0.555 (0.012)
M2OR	MolTrans	0.638 (0.066)	0.402 (0.053)	0.822 (0.027)	0.476 (0.042)
	HyperAttentionDTI	0.737 (0.015)	0.609 (0.028)	0.773 (0.020)	0.584 (0.022)
	Ours - mixture	0.780 (0.012)	0.689 (0.016)	0.698 (0.042)	0.605 (0.017)

The performance on all three datasets is summarized in Tab. 11. Overall, on DAVIS HyperAttentionDTI and MolTrans have higher AveP value than our model, and on KIBA HyperAttentionDTI performs best and our approach is on par with MolTrans. Our model outperforms other approaches on the OR-molecule dataset.

F ANALYSIS OF THE COMBINATORIAL CODE

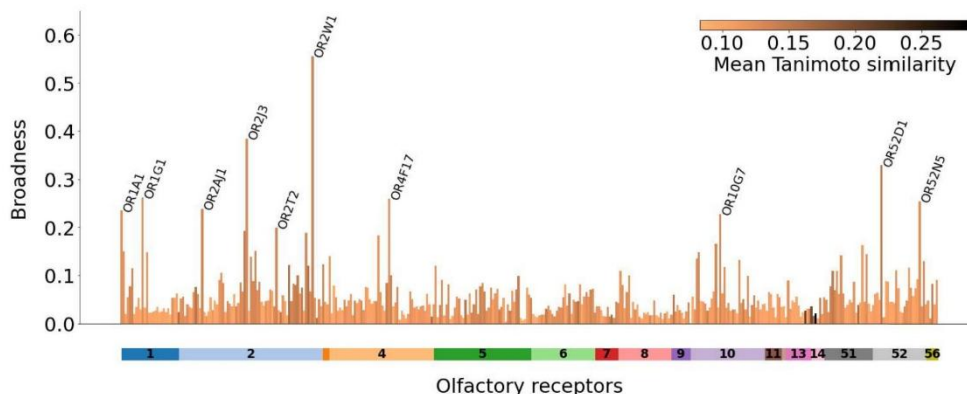


Figure 4: Distribution of the estimated broadness of human ORs. Broadness for each OR is defined as a portion of its predicted active compounds out of all known odorant molecules as given by the pyrfume database (Castro et al., 2022). The olfactory receptors are grouped by receptor families (Olender et al., 2008) and colored by mean Tanimoto similarity between their active compounds.

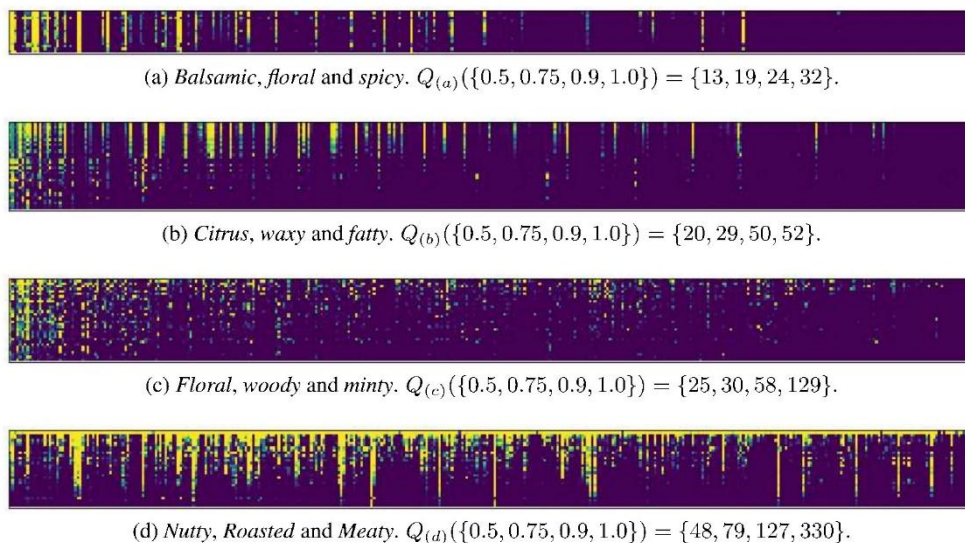


Figure 5: Examples of the predicted combinatorial codes for 4 abstract odor families and 385 human ORs⁴. Each row is an odorant molecule and each column is an olfactory receptor. Each molecule-OR pair is colored by the probability of activation ranging from dark blue for no activation to yellow for an active pair. The ORs are sorted in the figures in descending order from the left to right by their predicted broadness. Captions under the figures correspond to the most frequent descriptors in a given abstract odor family, ordered by frequency.

⁴Figures for all families available here: <https://github.com/MatejH1/Receptor2Odorant>.

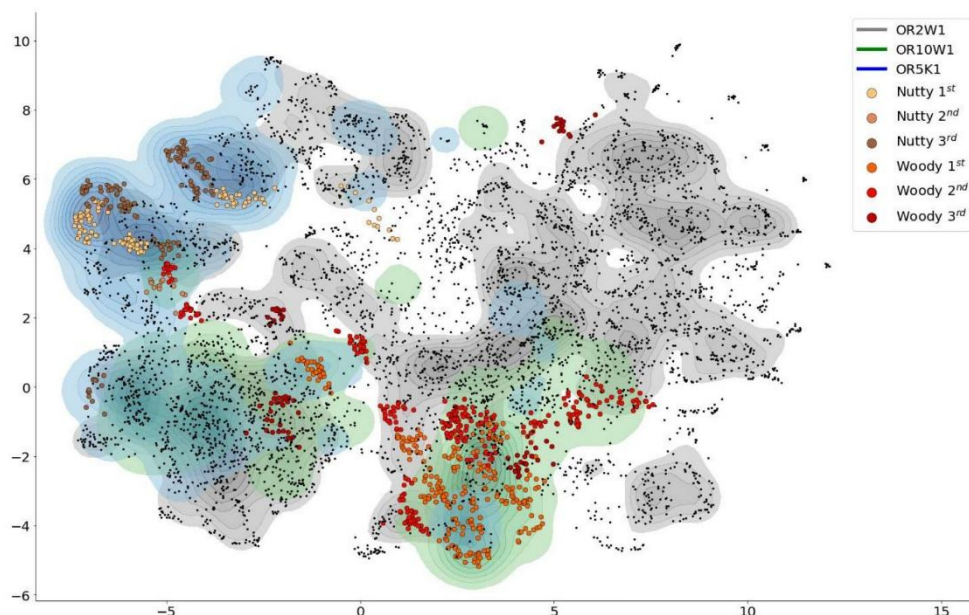


Figure 6: UMAP (McInnes et al., 2018) projection of the odor embedding with estimated densities of active compounds. Each point is an embedded odorant from pyrume database (Castro et al., 2022) and highlighted points are molecules that belong to the abstract odor families with the given smell descriptor being 1. the most occurring one, 2. the second most occurring, and 3. the third most occurring in the family. The contour regions are estimated densities of active compounds for a given receptor. We identify specific receptors OR10W1 (green) and OR5K1 (blue). OR10W1 is activated by molecules with the *woody* descriptor whereas OR5K1 by *nutty* compounds. OR2W1 (grey) is an example of a broad receptor activated by odorants with all kinds of different descriptors.

G DATASET

Our newly gathered dataset of OR-molecule pairs⁵, called M2OR for Molecule to Olfactory Receptor, currently contains 46 700 unique pairs that were collected from 31 scientific papers (Malnic et al., 1999; Wetzel et al., 1999; Kajiya et al., 2001; Spehr et al., 2003; Saito et al., 2004; Araneda et al., 2004; Sanz et al., 2005; Neuhaus et al., 2006; Jacquier et al., 2006; Schmiedeberg et al., 2007; Keller et al., 2007; Fujita et al., 2007; Saito et al., 2009; Repicky & Luetje, 2009; Shirasu et al., 2014; Mainland et al., 2015; Gonzalez-Kristeller et al., 2015; Geithe et al., 2015; 2016; 2017; Sato-Akuhara et al., 2016; Noe et al., 2016; 2017; Jones et al., 2019; Yasi et al., 2019; Frey et al., 2020; Jabeen et al., 2021; Marcinek et al., 2021; Haag et al., 2021; 2022; Cong et al., 2022). It combines 596 different molecules with 1237 unique sequences and in total consists of 68 837 experiments. Wild type, variant and mutant receptors are assembled across 11 different mammal species. For each experiment, several types of information are available:

- **Receptor:** Species, Mutation, Gene Name, Uniprot ID, Sequence. When available, primary sequence or Uniprot ID is obtained from the publication or else Uniprot ID is inferred from the gene name. We keep original gene name as was stated in the article. For mutants, Uniprot ID/Sequence corresponds to the reference sequence and the mutation is annotated in the dedicated column. Then mutated sequence should be deduced by combining the reference sequence and the mutation.
- **Molecule:** Name, CID, CAS, InChI Key, canonicalSMILES, Mixture. Compounds are identified by InChI Key based on the Pubchem database (Kim et al., 2020). Distinction between mono-molecular compounds, mixture, and sum of isomers is made, and it is derived

⁵Data available here: <https://github.com/chemosim-lab/M2OR>

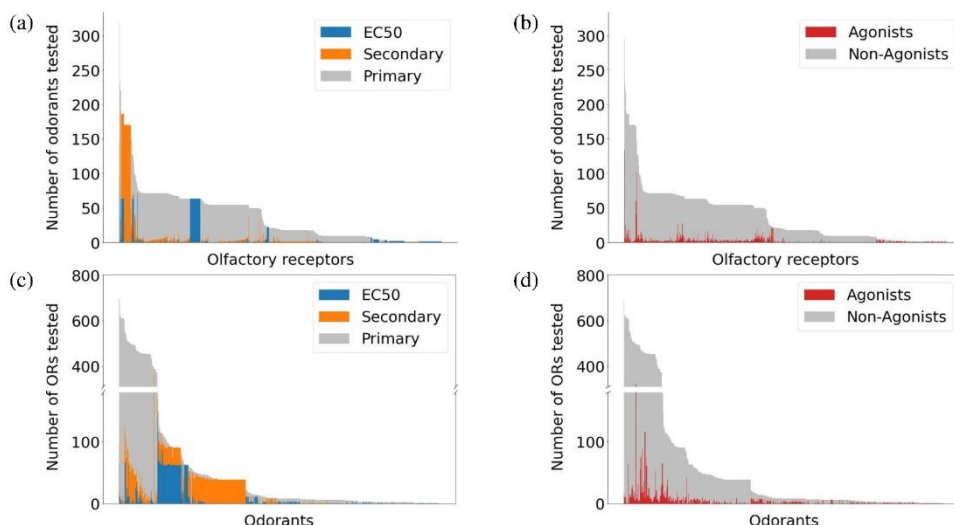


Figure 7: Distributions of the number of lab experiments in the M2OR dataset: (a-b) per receptor, (c-d) per molecule. Sever imbalance in the quality, number of experiments, and class distribution can be observed.

using *EnumerateStereoisomers* function from RDKit (Landrum, 2020). When no InChI Key is available, canonicalSMILES is inferred from the chemical structure.

- **Response:** Value, Unit, Value_Screen, Unit_Screen, Responsive, Nbr_measurements. Value corresponds to the experiment’s raw response if available and Value_Screen is a tested concentration. If a sample is non-responsive in a dose-response measurement, Value_Screen is set to the maximum tested concentration. Nbr_measurements corresponds to the number of injections for a given experiment. Responsiveness is encoded as 1 for agonists and 0 for non-agonists.
- **Bioassay:** Type, Cell_line, Delivery, Assay, G-protein, Co_transfection, Assay System, Tag. Information about the type of assay done, the cell line used, the delivery method of the odorant and about the system such as the presence of G protein, tags or co-transfection of other proteins.
- **Reference:** Reference, DOI, Reference Position. Article name, authors, DOI and location where the information was found.

Table 12: Dataset statistics.

Experiments	68 110
Unique pairs	46 717
Unique sequences	1237
Unique molecules	596
Species	11

Data preparation. Some OR-molecule pairs may be tested in several different experiments with possibly different outcomes. In order to train and evaluate our model, we extracted unique pairs from the dataset as follows: If a given pair has a dose-response experiment, we take its result and ignore others. Otherwise, if a pair has multiple measurements but not a full dose-response curve, we consider it as a secondary screening and we take this pair and its response only if there is no conflict between the experiments. If there is a conflict, we discard the pair entirely (in total 166 pairs were discarded).

M2OR: a database of olfactory receptor–odorant pairs for understanding the molecular mechanisms of olfaction

Maxence Lalis^{1,†}, Matej Hladiš^{1,†}, Samar Abi Khalil¹, Loïc Briand², Sébastien Fiorucci¹ and Jérémie Topin^{1,*}

¹Institut de Chimie de Nice, Université Côte d'Azur, UMR 7272 CNRS, 06108 Nice, France

²Centre des Sciences du Goût et de l'Alimentation, CNRS, INRAE, Institut Agro, Université de Bourgogne, F-21000 Dijon, France

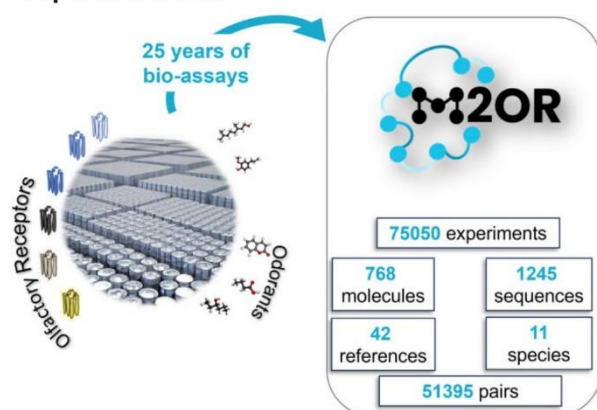
*To whom correspondence should be addressed. Tel: +33 4 89 15 01 32; Fax: +33 4 92 07 61 51; Email: jeremie.topin@univ-cotedazur.fr

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Abstract

Mammalian sense of smell is triggered by interaction between odorant molecules and a class of proteins, called olfactory receptors (ORs). These receptors, expressed at the surface of olfactory sensory neurons, encode myriad of distinct odors via a sophisticated activation pattern. However, determining the molecular recognition spectrum of ORs remains a major challenge. The Molecule to Olfactory Receptor database (M2OR, <https://m2or.chemsensem.fr/>) provides curated data that allows an easy exploration of the current state of the research on OR-molecule interaction. We have gathered a database of 75,050 bioassay experiments for 51 395 distinct OR-molecule pairs. Drawn from published literature and public databases, M2OR contains information about OR responses to molecules and their mixtures, receptor sequences and experimental details. Users can obtain information on the activity of a chosen molecule or a group of molecules, or search for agonists for a specific OR or a group of ORs. Advanced search allows for fine-grained queries using various metadata such as species or experimental assay system, and the database can be queried by multiple inputs via a batch search. Finally, for a given search query, users can access and download a curated aggregation of the experimental data into a binarized combinatorial code of olfaction.

Graphical abstract



Introduction

Mammalian sense of smell is based on a sophisticated mechanism for detection of volatile molecules by olfactory receptors (ORs) expressed in the membrane of sensory neurons (1). The first OR gene was discovered in 1991 (2) and since then, a thorough investigation has been carried out for various species including humans (3–7), mice (8), rats (9) and elephants (10). The total number of OR genes has been estimated to be approximately 400 in humans (11) and can reach nearly 5000 in elephants (10), making ORs the largest family of G-protein coupled receptors (GPCR).

Odorants are recognized by ORs according to a combinatorial code: a molecule can activate a subset of receptors and each receptor can respond to several different molecules (12). Understanding the mechanism by which a relatively small number of receptors, typically a few hundred, enable the discrimination of at least tens of thousands of odors (13) remains a major challenge. Emerging research indicates that even minor alterations in the functionality of a single receptor can lead to notable perceptual consequences (14–16). Despite our understanding that the OR gene family displays a significant genetic and functional diversity, the complex nature of the ol-

Received: August 13, 2023. Revised: September 13, 2023. Editorial Decision: September 27, 2023. Accepted: October 4, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. Summary of the existing OR-molecule databases listed by creation year

Database	Pairs	Mols.	ORs	Species	Web interface	Non-responsive pairs	Bioassay description	Year
OdorDB	402	95	812	27	✓	✗	✗	2000
ODORactor	4223	3000	1608	2	✓	✗	✗	2011
OlfactionDB	400	85	83	2	✓	✗	✗	2012
Cong <i>et al.</i>	15 693	244	720	2	✗	✓	✗	2022
OlfactionBase	874	330	150	2	✓	✗	✗	2022
M2OR (this paper)	51 395	768	1246	11	✓	✓	✓	2023

Pairs corresponds to a number of distinct OR-molecule pairs in the database. Mols. and ORs are, respectively, numbers of unique molecules and protein sequences. Check mark and cross mark indicate a presence of a web interface, an inclusion of non-responsive pairs and details about the experimental procedure (bioassay description).

factory code and our limited knowledge about specific OR-odorant pairings pose challenges in translating this variability into perceptual distinctions.

In order to decipher the combinatorial code of odours, a considerable effort has been devoted to understand the interactions between ORs and molecules. Since the pioneering work led by G.M. Shepherd, several databases and datasets containing genomic, proteomic and functional information on olfactory receptors have been created (17–23). However, majority of these platforms serve as repositories for genomic and proteomic data related to olfaction and only a few of these online resources offer a list of known interactions between olfactory receptors and molecules. In addition, only one single dataset contains non-responsive OR-molecule pairs and none of the existing databases gather crucial information on the bioassays, nor on the stereochemistry or concentration of the tested molecules. To address these limitations we propose Molecule to Olfactory Receptor (M2OR) database, which is the largest and most comprehensive database of OR-molecule experiments available (Table 1).

M2OR compiles experimental data from 42 scientific articles (12,22,24–63) and contains information about OR-molecule interactions including non-responsive experiments, details on experimental procedures, concentrations tested, as well as curated stereochemistry properties of the molecules. The web interface includes an intuitive search engine with advanced search options for fine-grained queries based on molecules, receptors or experimental details. Users can access raw experimental results or a curated aggregation of the experiments to OR-molecule pairs. Batch search can be used to obtain results in bulk using several standard molecule and receptor identifiers. This database offers an easy access to the results of 25 years of OR-molecule bioassays with an unprecedented level of detail which will facilitate both *in vitro* and *in silico* deorphanization.

Comparison with previous work

During the last decades, efforts to understand how an OR recognises an odorant have generated an impressive amount of data. Table 1 compares the different databases which have been created to provide access to these results. To date, only OdorDB (19) and OlfactionBase (23) allow for web-searching for OR-molecule pairs. However, these two databases offer a limited access to experimental results (874 pairs are available in OlfactionBase and OdorDB lists 95 odorants which are known to activate ORs). Developed in early 2010s, OlfactionDB (20) and ODORactor (21) webserver are not main-

tained anymore and a recent work by Cong *et al.* (22) only provided a raw dataset in a CSV format without any web-server.

Except Cong *et al.* (22), none of the available data sources contain non-responsive OR-molecule pairs, which are an integral part of the combinatorial code of odorants. In addition, the existing databases do not gather information on bioassays used to assess the OR-molecule response, stereochemistry properties of the molecules, nor their concentrations. In contrast, the philosophy of M2OR is to provide all available information which is pertinent to study olfaction, such as non-responsive experiments, details on experimental procedures, concentrations tested, as well as curated stereochemistry properties of the molecules.

New features

Olfactory perception is dependent on odorant concentration and changes in concentration can lead to different perception of hedonicity or olfactory quality (64). From a molecular point of view, concentration of a ligand has a considerable influence on the response of ORs. An increase in the ligand concentration results in a higher probability of OR activation, ultimately leading to an increase in cellular signaling. This way, a molecule will not induce any cellular response at low concentration, whereas it will become an agonist for a large subset of ORs when its concentration increases. Contrary to previous databases, M2OR includes either screening concentration or EC₅₀ for all gathered experimental data. This allows users to analyse OR-molecule interaction beyond simple responsiveness. Additionally, a great deal of effort has been invested in providing stereochemistry of molecules. Indeed, certain ORs, such as OR1A1, have a different response to enantiomers (39), making the complete curation of stereochemistry an essential aspect of M2OR.

Functional studies of ORs have relied on a variety of bioassays, including the use of native olfactory sensory neurons (OSNs) and several different heterologous expression systems that have facilitated the deorphanisation of ORs. However, it is crucial to take into consideration the instances of assay-dependent bias when interpreting OR responses (65). A recent demonstration illustrates this phenomenon: new ligands for ORs were successfully identified in human prostate carcinoma cell lines (LNCaP), whereas they were not recognised when ORs were expressed in HEK293 cells (66). As a result, M2OR incorporates information about experimental procedures, enabling users to identify and filter experiments based on various assay metadata, such as cell lines or assay types.

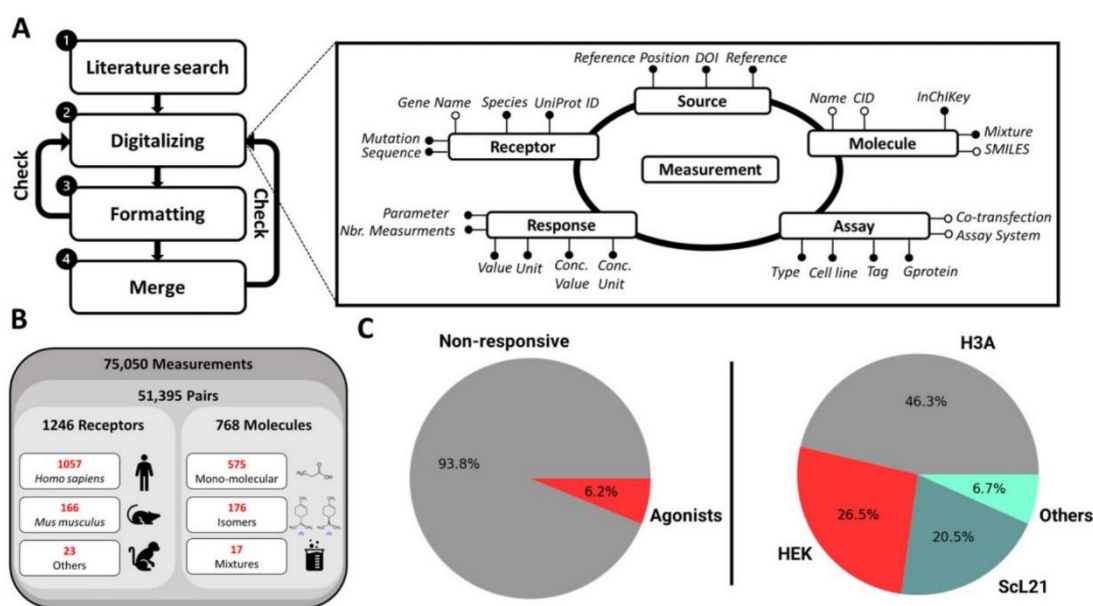


Figure 1. Summary of the construction and architecture of M2OR. **(A)** Database construction workflow: diagram shows the assembly process with framed figures for each category, subdivided further. Solid circles indicate mandatory columns and hollow circles are optional information. **(B)** Database overview of the distribution of data in terms of sequences and molecules. **(C)** Distribution of the number of agonists and non-agonists and the type of bioassay. H3A (Hana3a cell line) and ScL21 (stable cell line) are both derived from HEK (human embryonic kidney cell line).

These unique features significantly increase the overall value and information provided by M2OR. For example, the assay metadata can be used to estimate response confidence level in a subsequent applications such as training a machine learning model (67). Information such as concentration, cell line or assay type facilitate bioassay experimental design and it provides a necessary data to study the impact of different experimental procedures on the OR-molecule pair response.

Database construction

Data extraction and curation

A thorough and detailed workflow was employed to ensure the accuracy, consistency and reliability of the M2OR database. Firstly, a comprehensive list of scientific papers was collected using literature tools such as PubMed, Google Scholar and ResearchRabbit, and then sorted based on the estimated number of experiments. This approach allowed for the identification of valuable sources of data, which could be further curated to provide a robust and comprehensive database. To populate M2OR, the main text and supplementary material of each article are parsed through, and complementary information is, when needed, requested from the corresponding authors. This information is then carefully reviewed, checked and formatted to ensure its accuracy and consistency. A series of checks are also employed to highlight any errors such as typos or inconsistencies in the data (see Data consistency checks in Supporting material). This approach ensures that the data could be easily compared and analysed across different studies, thus providing a valuable resource for researchers interested in receptor-molecule interactions and their effects on the olfactory system. Figure 1A shows the pipeline followed for the construction of M2OR. Checking and formatting code as well as the raw information for each

data source is freely available on M2OR GitHub repository (<https://github.com/chemosim-lab/M2OR>).

Molecular identifier

The International Chemistry Identifier Key (InChIKey) (68) serves as a standardized molecular identifier in the M2OR database. It is retrieved from PubChem (69) using the provided molecular identifiers found in the articles. Identifiers used by the authors were preferentially kept as stated in the original publication, and other identifiers, such as CID, were inferred from PubChem using InChIKey. Name and CAS are preferably retained as they are in the source, although there might be some inconsistencies due to CAS being a proprietary identifier and name not standardized. If only the molecular structure is available, a SMILES representation is inferred and further used to search for the corresponding InChIKey. In one study, the authors used a newly-synthesised molecule which is only identified by its SMILES string.

Compound classification

The diversity of the composition of the tested compounds is defined by three classes: 'mixture', 'sum of isomers' and 'monomolecular'. The term 'mixture' indicates that the experiments were carried out using a blend of multiple molecules (e.g. essential oil or artificial composition) and mixtures are identified by a space-separated list of InChIKeys of each compound. For all the molecules, information on isomerism is carefully researched. Number of chiral centres and geometric isomerism are automatically determined from SMILES using the RDKit package. If a molecule has at least one chiral center and no specific information is provided about the enantiomery or the diastereoisomery, the molecule is considered to be a combination of isomers, and thus labelled 'sum of isomers'. In cases where the authors explicitly mentioned isomerism of a

tested molecule it is identified as ‘monomolecular’. All achiral molecules are also labeled as ‘monomolecular’.

Receptor identifier

Protein sequence serves as a standardized OR identifier in M2OR. If it is not available in the original publication, it is retrieved from UniProt (70) using the name or other identifier provided by the authors. For mutated receptors, the wild-type identifier and sequence is retrieved, and the mutation is separately indicated using *XpositionY* format where amino acid *X* from the wild-type sequence is mutated to amino acid *Y* at the given *position*. Additions are indicated by setting *X* to ‘+’ and deletion with ‘–’ as *Y*. The taxon name of the species from which the OR originates is also included.

Response

The results of the experiments are divided into two categories according to the authors conclusion. An experiment is labelled ‘responsive’ if the measured signal is considered as statistically different from the baseline by the authors. Otherwise it is labelled ‘non-responsive’. If available, different experimental parameters are collected such as the compound concentration, the assay raw response and EC₅₀. Finally, if authors did not provide conclusion on the response (e.g. for screening experiments), a specific workflow is used to determine the response (see Mainland *et al.* (27) data treatment in Supporting material).

Bioassay

Exploring the recognition spectrum of ORs have required the development of several cell biology protocols to overcome the challenge associated with expressing ORs on the surface of heterologous systems cells (71). In the majority of OR assays, compounds are typically evaluated individually against each receptor using transient mammalian cell line-based luciferase assays. However, there are multiple methods to measure the molecule-induced activity of an OR and we have carefully extracted assay information from the publications. Briefly, we gather several parameters such as measured quantity, type of cell line for OR expression, type of G-protein, co-transfected protein or type of tags (i.e. N-terminal modifications). The experimental conditions and tools used are also included in the database.

Resources

The original source of data, including the reference and DOI, is collected. Additionally, a specific location of the information, such as table, figure or supplementary material is included in the database.

Web implementation

M2OR is a dynamic web application developed using the Laravel framework (<https://laravel.com/docs/10.x>), which utilizes HTML, CSS, JavaScript and PHP for development. M2OR offers a user-friendly and interactive experience, enabling seamless data exploration and downloads. Additionally, a Content Management System (CMS) has been developed also on the Laravel framework to facilitate efficient management, editing, addition and deletion of data from the administration part. M2OR is powered by MySQL relational database management system (<https://www.mysql.com/>).

Data content

M2OR contains 1246 unique protein sequences across 11 different mammalian species. *Homo sapiens* and *Mus musculus* ORs represent most of the experiments, with respectively 66% and 33% of the data (Figure 1B). The database includes 768 compounds, where 176 are sum of isomers, 407 are non-stereoisomeric molecules, and 168 are specific isomers. Moreover, 17 mixtures are also included in the database. Combination of these receptors and molecules results in a total of 51 395 unique pairs. With some pairs tested multiple times, the database contains 75 050 different experiments and represents the largest database available in the literature on OR bioassays (Table 1). With 3100 active pairs and, 48 295 non-responsive pairs, M2OR comprises 6% of agonist (Figure 1C). 41% of the bioassay results are luciferase assays using the Hana3A cell line, which expresses chaperon proteins like RTP1 or RTP2, olfactory G-protein and rho tag. This collection gives an overview of what has been achieved in terms of ligand binding assays since 1999.

Experiments table

When searching in M2OR, the user is navigated to Experiments table which contains all relevant bioassay data. It includes several experiments per OR-molecule pair depending on the number of different types of experimental procedures performed on the pair. Molecules and receptors in Experiments table are respectively described by *Name*, *CID*, *CAS*, *InChIKey*, *SMILES* and *Mixture* for molecules, and *Gene name*, *UniProt ID*, *Sequence*, *Mutation* and *Species* for receptors (Figure 2D).

Details about the response are provided via columns: *Parameter*, *Responsive*, *Value*, *Unit*, *Concentration*, *Concentration unit* and *Nbr. measurements*. Responses are binary-coded in the *Responsive* column with 0 representing a non-agonist and 1 an agonist. The *Concentration* and *Concentration unit* columns contain either a concentration and its units, in the case of a screening experiment, or a maximum concentration and its units used in a dose-response assay (i.e. experiments performed at several concentrations). When a mixture of compounds was used for the experiment, the entry in the *Concentration* column is a comma-separated list of concentrations for each individual component. For screening experiments, the *Value* and *Unit* columns contain raw value reported in the source article, and for dose-response assays, the EC₅₀ value and its units are reported in these columns. Which type of value is reported in the *Value* column is described by *Parameter*, which can be EC₅₀ for dose-response measurements, Raw if the raw response is available or norm_rec, norm_pair, or norm_other for responses normalized by the receptor’s baseline, the response of a given pair, or an unknown normalization denominator, respectively. Additionally, number of repetitions is reported in the *Nbr. measurement* column.

Bioassays are fully described in the following columns: *Type*, *Cell line*, *Gprotein*, *Co-transfection*, *Tag*, *Delivery*, *Assay* and *Assay system*. Each bioassay is described by the measured quantity (*Type*), which could be a concentration of cAMP (CAMP), Ca²⁺, or Secreted embryonic alkaline phosphatase (SEAP), the fluorescence emitted by Green fluorescent protein (GFP) or by luciferase, or the membrane activity measured by intensity (Intensity) or conductance (Conductance). *Cell line* used for OR expression is also specified. This could be HEK293T cells (HEK), engineered OR-specific derivatives

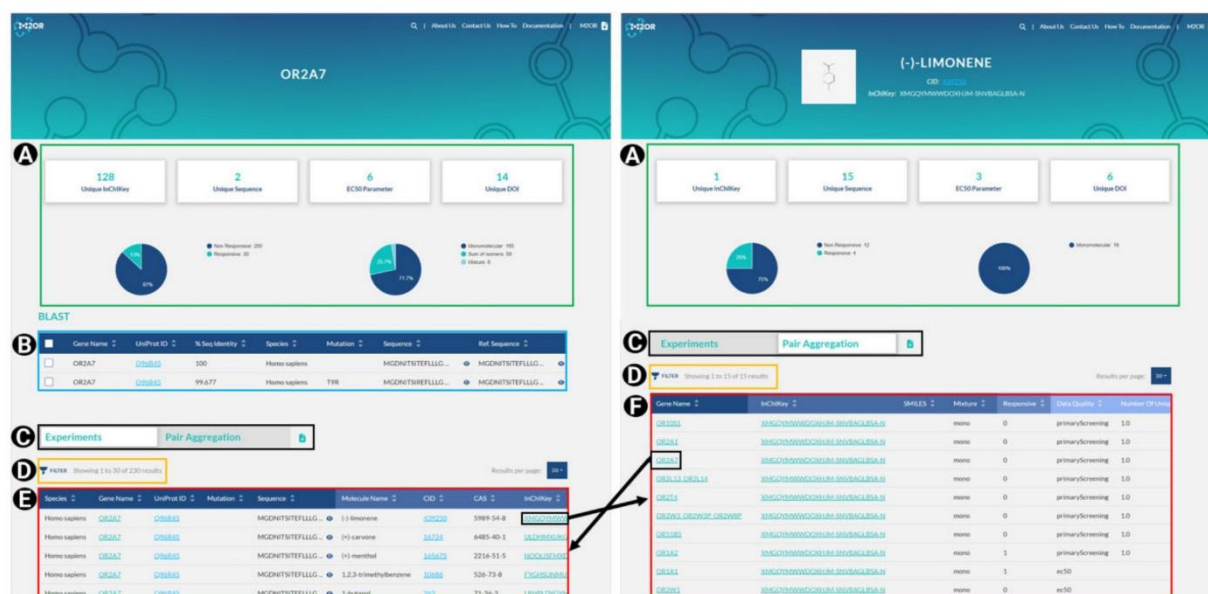


Figure 2. Search results. Receptor page (left) and Molecule page (right). They are subdivided as follows: **(A)** statistics for the displayed table, from left to right the number of unique molecules (Unique InChIKey), number of unique sequence (Unique Sequence), number of dose-response experiments (EC₅₀ Parameter), number of associated references (Unique DOI). Pie chart on the left represents the ratio of responsive/non-responsive records for the displayed table. Pie chart on the right shows the ratio between sum of isomers, monomolecular compounds and mixtures. **(B)** BLAST table in the case of a receptor search. User can select one or more sequences related to the initial search. **(C)** Button to switch between Experiments table and Pair aggregation table. **(D)** Button to select filters. **(E)** Experiments table corresponding to the search query. Gene name and InChIKey link to Experiments table for a given receptor and a given molecule, respectively. UniProt ID and CID redirect the user to the corresponding record on UniProt and PubChem. **(F)** Pair aggregation table corresponding to the search query.

like Hana3A cells (H3A) and ScL21 (SCL21), yeast-based systems, oocytes, olfactory sensory neurons (OSN), Neuroblastoma x Glioma hybrid (NxG108CC15), or human cancer cells (HeLa/Olf). When available, the type of G-protein (*Gprotein*) is detailed (*G_{olf}*, *G_{alpha16}*, *G_{alpha15/16}*, *G_{alphaq}*). Any protein co-transfected with the ORs, as mentioned in the source, is listed in the *Co-transfection* column. N-terminal modifications, known as tags (Il-6-Halotag®, Flag, Rho, GFP, MYC, Rho Lucy), which can influence the response, are frequently used in these bioassays and are reported in the *Tag* column. Experimental conditions, such as *Assay (in vitro, ex vivo)* or *Delivery* method (liquid, gas), are also noted. The tools used, such as FLIPR or Glosensor, are mentioned in the *Assay system* column.

Finally, resources are indicated in the columns: *Reference*, *DOI* and *Reference Position*. The original source of the experiments, including the reference and DOI, is provided. Additionally, a specific location of the information, such as table, figure or supplementary material is noted in the *Reference position* column.

BLAST table

Human genome includes approximately 1000 OR genes, of which around 60% are considered pseudogenes (7,72). Each OR has a distinct recognition spectrum and alterations in one or more amino acids can significantly change its response (73,74). Multiple variants and mutants of the same gene have been tested in the literature and they are gathered in the M2OR database. For instance, 42 different sequences share more than 99% sequence identity with the sequence annotated as OR1A1 in the UniProt database (70) (i.e. with the

reference sequence) and some of these sequences show different responses compared to the reference. To facilitate the comparison of such cases, similar sequences are grouped under the same reference sequence. BLAST algorithm (75) is used to compare each sequence in M2OR against the UniProt database (2023 release 03). Name and UniProt ID of the best match in terms of identity are then associated with these sequences. They are subsequently grouped by their best match name, resulting in the BLAST table (Figure 2B). The BLAST table includes the name of the best match (*Gene name*), its UniProt ID (*UniProt ID*), its UniProt sequence (*Ref. sequence*), the query sequence (*Sequence*), an identity ratio between the query sequence and the sequence of the best match (*Seq. identity*), the species of the receptor (*Species*) and the mutations required to transit from the Ref. sequence to the query sequence (*Mutation*).

Pair aggregation table

Multiple experiments for the same OR-molecule pair can be found in Experiments table (e.g. multiple concentrations tested or different experimental procedures). However, users are often interested in an aggregated view on the OR-molecule pairs for applications such as analysis of the combinatorial code, or new active pair prediction (22,67,76,77). Pair aggregation table is created to provide this consensus response for each OR-molecule pair. Creation of this table relies on the following decisions:

- (1) Dose-response measurements are prioritized over screening data. When there are contradictory responses between multiple dose-response measurements for a

given pair, then the pair is discarded. So far, this resulted in elimination of 27 pairs.

- (2) In case of the screening data, we exclude pairs that are responsive at low concentrations but not at higher concentrations. We also discard pairs that exhibit inconsistencies at the same concentration. This leads to elimination of additional 246 pairs.
- (3) A consistent screening pair is considered responsive if it is responsive in at least one concentration.

By applying these rigorous and systematic criteria, we obtain a set of 51 395 consistent OR–molecule pairs.

Pair aggregation table (Figure 2F) comprises *Responsive*, *Data quality* and *Number unique value screen* columns. In addition, there are the molecule and receptor information columns as in Experiments table. The *Responsive* column contains the consensus response of the pair and the *Data quality* column indicates reliability of the response decision. The most reliable are dose-response experiments (EC₅₀). If only a screening was done, then secondary screening (secondaryScreening), with mediocre reliability, indicates that the pair was tested in at least two distinct concentrations and the least reliable primary screening pairs (primaryScreening) were tested in a single concentration. In case of a screening-based decision, *Number unique value screen* shows the number of distinct concentrations tested for a given pair.

Navigating through M2OR

M2OR can be navigated via a web interface using four options:

- (1) Basic search: It enables the user to search for specific molecules or receptors using identifiers such as InChIKey, CID, SMILES, molecule name, receptor sequence, receptor name or UniProt ID. The results are displayed on a detailed ‘Listing’ page.
- (2) Advanced search: This feature allows the user to conduct detailed searches within the M2OR database using dropdown menu options or auxiliary search queries for each column. Boolean operators AND, OR and NOT are available for more granular control over search queries.
- (3) Batch search: With this functionality, the user can execute multiple search queries simultaneously by inputting identifiers such as sequences, UniProt IDs, Gene names, InChIKeys, CIDs or SMILES. The user can combine search criteria using boolean operators AND, OR and NOT.
- (4) Browsing table: All available experiments in the M2OR database can be browsed through Browsing table without any specific search.

Basic search

The user can conduct a search for either molecules or ORs using the search bar located on the home page (Figure 3A). For molecules, InChIKey, CID, SMILES or common names can be entered. For ORs, a complete sequence, gene name or UniProt ID may be an input. Additionally, ORs can be identified using a segment of the sequence, such as specific motifs. After initiating search, the user will be directed to a ‘Listing’ page, which compiles all the records that correspond to the search criteria.

Listing

In the Listing page, each tile represents either a molecule or an OR and includes their main identifiers (Figure 3B). For molecules, these are a common name, a clickable CID that redirects to the corresponding PubChem page, and an InChIKey identifier. For receptors, the identifiers are the gene name, a clickable UniProt ID that redirects to the corresponding UniProt page, and the associated species.

The number of search results is displayed at the bottom left of the listing page, with additional pages available for navigation if the number of matches exceeds 30. The user can select a single tile or multiple tiles using a clickable box in the top right corner of each tile. Once the selection is finalized, the user can navigate to the corresponding Experiments table (see Experiments table section for further details). If a single tile was selected, the user is directed to *Molecule page* or *Receptor page*, depending on the choice. Otherwise a general *Experiment page* is shown.

Experiment/Molecule/Receptor page

Statistics pertaining to the displayed Experiments table are shown on top of the *Experiment page* (Figure 2A). There are four sections that respectively show the number of displayed unique molecules, unique receptors, EC₅₀ experiments and sources linked to the table. Two pie charts offer a swift overview of the proportions between agonists and non-agonists and monomolecular compounds, sum of isomers and mixtures.

The table can be downloaded by clicking on the download icon. Above the table, the total number of records is shown, and the user can choose number of rows to be displayed per page. Each column can be sorted in ascending or descending order for ease of navigation. Filter (Figure 2D) is available for users to select a specific value for certain columns. For example, to get all known agonists with their corresponding EC₅₀ values, the user can search for a given receptor of interest, and then via the Filter function select only responsive records obtained in a dose-response experiments. The statistics on top of the page are updated according to the selected filters.

Cells in the columns such as the *UniProt ID*, *CID* and *DOI* are clickable and link directly to the relevant UniProt, PubChem and reference web pages, respectively. *InChIKey* and *Gene Name* redirect to Experiments table corresponding to the desired molecule or receptor. Additionally, an ‘eye’ icon is available for users to view the full content of each cell. For *Molecule page* and *Receptor page*, the user can find a banner which provides information about the search query, similar to the details available on the tile in the listing page.

The user will also find a supplementary BLAST table in the *Receptor page*. This table enables selection of variants or mutants associated with the receptor by using the *Check box* column. The user can select multiple entries from this column, and conveniently select or deselect all entries with a single click at the top of the column. A complete description is provided in the BLAST table section.

Batch search

Batch search is a feature that allows users to execute multiple search queries simultaneously (Figure 3C). For instance, the user can search for responses of multiple receptors by inputting a set of sequences, each separated by a line. Batch search can be used for receptors using various identifiers, namely the full sequence, UniProt ID or receptor name. The

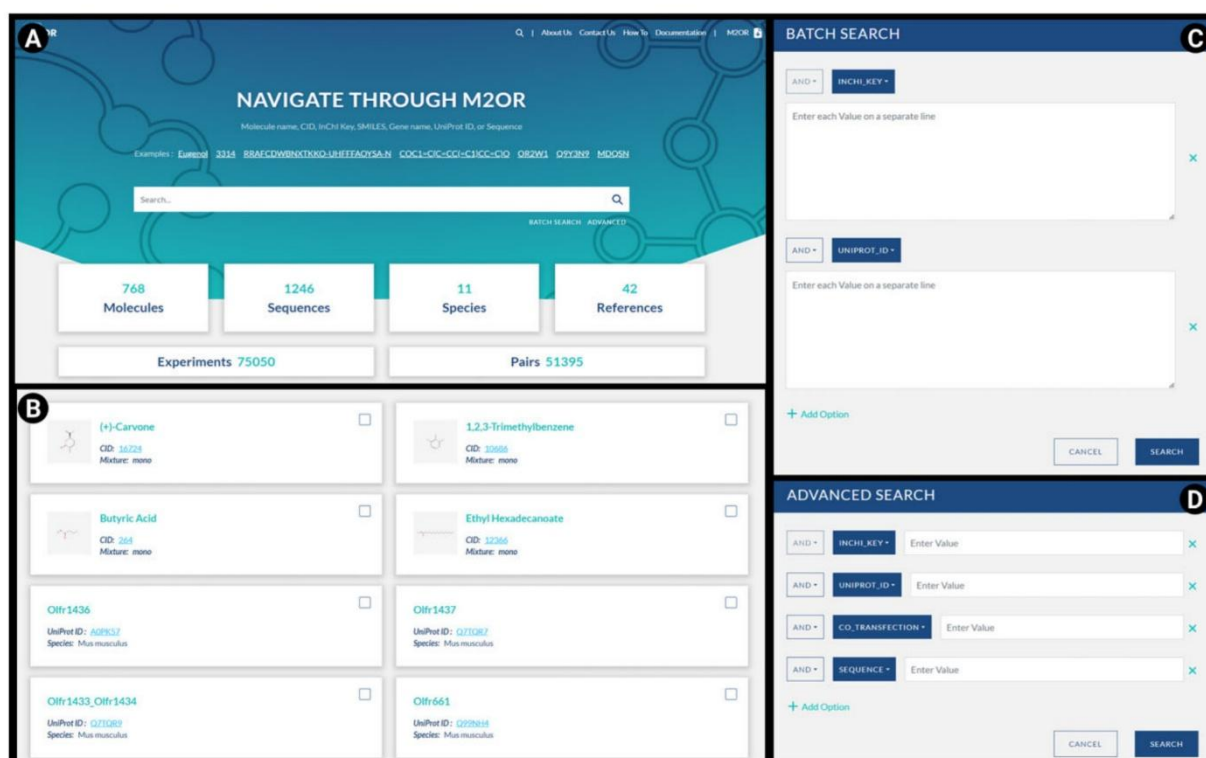


Figure 3. Search options within M2OR. **(A)** M2OR front page displaying the statistics and examples for Basic search. **(B)** Listing page resulting from the basic search. User can select several molecules or receptors and get access to the experimental results. **(C)** Batch search. The database can be explored by providing list of molecules (identified by name, CID, InChIkey or SMILES) or receptors (identified by sequence, gene name or UniProt ID). **(D)** Advanced search. Detailed queries can be executed within the M2OR database based on molecule, receptor, type of bioassays or reference.

same applies to molecules, where the corresponding InChIKey, CID or SMILES can be used as inputs. Furthermore, Batch search feature supports use of multiple criteria, and the user can manage their combination using boolean operators such as AND, OR or NOT.

Advanced search

Advanced search allows the user to execute detailed queries within the M2OR database (Figure 3D). Each row offers two options for input: the user can either manually enter a specific value or choose from the available options in a dropdown menu. Moreover, Advanced search function provides granular control over the search query through boolean operators AND, OR or NOT.

Browsing table

Browsing table can be accessed through the 'Experiments' tile at the main page. All the records in the M2OR database can be browsed from here. Columns can be sorted and the number of displayed records adjusted.

Future development

Moving forward, M2OR aims to continuously incorporate new bioassay results. The database will be expanded to include additional content such as antagonists and inverse agonists, a list of orphan ORs, and details of genes and their chromosomal positions. The perceptual dimension of odours

will also be taken into account by adding information on the odour of molecules or groups of molecules.

To improve the user experience and ease of exploration, we are working on introducing new features for navigating receptor sequences. This will involve an implementation of Balles-teros Weinstein numbering (78) allowing users to compare the influence of point mutations across different ORs. In addition, a snake-plot representation will be included to visualise important features along receptor sequences.

The main planned update is an integration of a web server based on our previously developed predictive model (67) which will provide users with a complete hypothetical combinatorial code for a given molecule. This will offer a valuable insight about interactions between molecules and receptors.

With these advances, M2OR will strive to remain a leading and comprehensive resource for scientists as well as enthusiasts exploring the fascinating world of olfactory receptors and their interaction with various molecules.

Conclusion

We design Molecule to Olfactory Receptor (M2OR) database, which brings together 75 050 bioassay experiments for 51 395 distinct OR-molecule pairs. It provides a user-friendly and intuitive interface to allow for fast and easy navigation through the current state of the research on OR-molecule interaction. Users can obtain information on the activity of a chosen molecule or a group of molecules, or search for ligands for a specific OR or a group of ORs. Advanced search

allows for fine-grained queries using various metadata such as species or experimental assay system, and the database can be queried by multiple inputs via a batch search. Finally, for a given search query, users can access and download a curated aggregation of the experimental data into a binarized combinatorial code of olfaction. Having an unprecedented level of details about molecules, receptors and their interaction, M2OR provides an essential tool to facilitate the exploration of the response spectrum of ORs. This database will provide a hub for researchers in olfaction and GPCR research.

Data availability

M2OR is freely available at <https://m2or.chemsensim.fr/>. Raw data and source codes are accessible at <https://github.com/chemosim-lab/M2OR> and <https://doi.org/10.5281/zenodo.8385484>.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We would like to thank Dietmar Krautwurst, Joel Mainland and Keishi Yoshikawa who kindly provided details about their results and experiments.

Funding

French National Research Agency (ANR) [ANR-19-CE07-0044] (PhD fellowship to M.H.); Fondation Roudnitska under the aegis of Fondation de France (PhD fellowship to M.L.); GIRACTION; Centre National de la Recherche Scientifique (CNRS) through the MITI interdisciplinary program. Funding for open access charge: Agence Nationale de la Recherche [ANR-19-CE07-0044].

Conflict of interest statement

None declared.

References

- Mombaerts, P. (1999) Molecular biology of odorant receptors in vertebrates. *Annu. Rev. Neurosci.*, **22**, 487–509.
- Buck, L. and Axel, R. (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, **65**, 175–187.
- Zozulya, S., Echeverri, F. and Nguyen, T. (2001) The human olfactory receptor repertoire. *Genome Biol.*, **2**, 1–12.
- Niimura, Y. and Nei, M. (2003) Evolution of olfactory receptor genes in the human genome. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 12235–12240.
- Malnic, B., Godfrey, P.A. and Buck, L.B. (2004) The human olfactory receptor gene family. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 2584–2589.
- Glusman, G., Sosinsky, A., Ben-Asher, E., Avidan, N., Sonkin, D., Bahar, A., Rosenthal, A., Clifton, S., Roe, B., Ferraz, C., *et al.* (2000) Sequence, structure, and evolution of a complete human olfactory receptor gene cluster. *Genomics*, **63**, 227–245.
- Olender, T., Waszak, S.M., Viavant, M., Khen, M., Ben-Asher, E., Reyes, A., Nativ, N., Wysocki, C.J., Ge, D. and Lancet, D. (2012) Personal receptor repertoires: olfaction as a model. *BMC Genomics*, **13**, 1–16.
- Zhang, X. and Firestein, S. (2002) The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.*, **5**, 124–133.
- Quignon, P., Giraud, M., Rimbault, M., Lavigne, P., Tacher, S., Morin, E., Retout, E., Valin, A.-S., Lindblad-Toh, K., Nicolas, J., *et al.* (2005) The dog and rat olfactory receptor repertoires. *Genome Biol.*, **6**, 1–9.
- Niimura, Y., Matsui, A. and 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.
- Olender, T., Nativ, N. and Lancet, D. (2013) HORDE: comprehensive resource for olfactory receptor genomics. *Olfactory Receptors: Methods and Protocols*. Humana Press, Totowa, NJ, pp. 23–38.
- Malnic, B., Hirono, J., Sato, T. and Buck, L.B. (1999) Combinatorial receptor codes for odors. *Cell*, **96**, 713–723.
- Gerkin, R.C. and Castro, J.B. (2015) The number of olfactory stimuli that humans can discriminate is still unknown. *Elife*, **4**, e08127.
- Menashe, I., Abaffy, T., Hasin, Y., Goshen, S., Yahalom, V., Luetje, C.W. and Lancet, D. (2007) Genetic elucidation of human hyperosmia to isovaleric acid. *PLoS Biol.*, **5**, e284.
- Mainland, J.D., Keller, A., Li, Y.R., Zhou, T., Trimmer, C., Snyder, L.L., Moberly, A.H., Adipietro, K.A., Liu, W. L.L., Zhuang, H., *et al.* (2014) The missense of smell: functional variability in the human odorant receptor repertoire. *Nat. Neurosci.*, **17**, 114–120.
- Trimmer, C., Keller, A., Murphy, N.R., Snyder, L.L., Willer, J.R., Nagai, M.H., Katsanis, N., Vossahl, L.B., Matsunami, H. and Mainland, J.D. (2019) Genetic variation across the human olfactory receptor repertoire alters odor perception. *Proc. Natl. Acad. Sci.*, **116**, 9475–9480.
- Healy, M.D., Smith, J.E., Singer, M.S., Nadkarni, P.M., Skoufos, E., Miller, P.L. and Shepherd, G.M. (1997) Olfactory receptor database (ORDB): a resource for sharing and analyzing published and unpublished data. *Chem. Senses*, **22**, 321–326.
- Skoufos, E., Healy, M.D., Singer, M.S., Nadkarni, P.M., Miller, P.L. and Shepherd, G.M. (1999) Olfactory Receptor Database: a database of the largest eukaryotic gene family. *Nucleic Acids Res.*, **27**, 343–345.
- Skoufos, E., Marengo, L., Nadkarni, P.M., Miller, P.L. and Shepherd, G.M. (2000) Olfactory receptor database: a sensory chemoreceptor resource. *Nucleic Acids Res.*, **28**, 341–343.
- Modena, D., Trentini, M., Corsini, M., Bombaci, A. and Giorgetti, A. (2011) OlfactionDB: a database of olfactory receptors and their ligands. *Adv. Lif. Sci.*, **1**, 1–5.
- Liu, X., Su, X., Wang, F., Huang, Z., Wang, Q., Li, Z., Zhang, R., Wu, L., Pan, Y., Chen, Y., *et al.* (2011) ODORactor: a web server for deciphering olfactory coding. *Bioinformatics*, **27**, 2302–2303.
- Cong, X., Ren, W., Pacalon, J., Xu, R., Xu, L., Li, X., de March, C.A., Matsunami, H., Yu, H., Yu, Y., *et al.* (2022) Large-scale G protein-coupled olfactory receptor–ligand pairing. *ACS Central Sci.*, **8**, 379–387.
- Sharma, A., Saha, B.K., Kumar, R. and Varadwaj, P.K. (2021) OlfactionBase: a repository to explore odors, odorants, olfactory receptors and odorant–receptor interactions. *Nucleic Acids Res.*, **50**, D678–D686.
- Yasi, E.A., Eisen, S.L., Wang, H., Sugianto, W., Minniefield, A.R., Hoover, K.A., Branham, P.J. and Peralta-Yahya, P. (2019) Rapid deorphanization of human olfactory receptors in yeast. *Biochemistry*, **58**, 2160–2166.
- Marcinek, P., Haag, F., Geithe, C. and Krautwurst, D. (2021) An evolutionary conserved olfactory receptor for foodborne and semiochemical alkylpyrazines. *FASEB J.*, **35**, e21638.
- Sato-Akuhara, N., Horio, N., Kato-Namba, A., Yoshikawa, K., Niimura, Y., Ihara, S., Shirasu, M. and Touhara, K. (2016) Ligand specificity and evolution of mammalian musk odor receptors: effect of single receptor deletion on odor detection. *J. Neurosci.*, **36**, 4482–4491.

27. Mainland, J.D., Li, Y.R., Zhou, T., Liu, W. L.L. and Matsunami, H. (2015) Human olfactory receptor responses to odorants. *Sci. data*, **2**, 1–9.
28. Saito, H., Kubota, M., Roberts, R.W., Chi, Q. and Matsunami, H. (2004) RTP family members induce functional expression of mammalian odorant receptors. *Cell*, **119**, 679–691.
29. Schmiedeberg, K., Shirokova, E., Weber, H.-P., Schilling, B., Meyerhof, W. and Krautwurst, D. (2007) Structural determinants of odorant recognition by the human olfactory receptors OR1A1 and OR1A2. *J. Struct. Biol.*, **159**, 400–412.
30. Noe, F., Polster, J., Geithe, C., Kotthoff, M., Schieberle, P. and Krautwurst, D. (2016) OR2M3: a highly specific and narrowly tuned human odorant receptor for the sensitive detection of onion key food odorant 3-Mercapto-2-methylpentan-1-ol. *Chem. Senses*, **42**, 195–210.
31. Haag, F., Pizio, A.D. and Krautwurst, D. (2022) The key food odorant receptive range of broadly tuned receptor OR2W1. *Food Chem.*, **375**, 131680.
32. Haag, F., Hoffmann, S. and Krautwurst, D. (2021) Key food furanones furanol and sotolone specifically activate distinct odorant receptors. *J. Agr. Food Chem.*, **69**, 10999–11005.
33. Noe, F., Frey, T., Fiedler, J., Geithe, C., Nowak, B. and Krautwurst, D. (2017) IL-6–HaloTag® enables live-cell plasma membrane staining, flow cytometry, functional expression, and de-orphaning of recombinant odorant receptors. *J. Biol. Meth.*, **4**, e81.
34. Shirasu, M., Yoshikawa, K., Takai, Y., Nakashima, A., Takeuchi, H., Sakano, H. and Touhara, K. (2014) Olfactory receptor and neural pathway responsible for highly selective sensing of musk odors. *Neuron*, **81**, 165–178.
35. Jabeen, A., de March, C.A., Matsunami, H. and Ranganathan, S. (2021) Machine learning assisted approach for finding novel high activity agonists of human ectopic olfactory receptors. *Int. J. Mol. Sci.*, **22**, 11546.
36. Repicky, S.E. and Luetje, C.W. (2009) Molecular receptive range variation among mouse odorant receptors for aliphatic carboxylic acids. *J. Neurochem.*, **109**, 193–202.
37. Neuhaus, E.M. (2006) A specific heat shock protein enhances the expression of mammalian olfactory receptor proteins. *Chem. Senses*, **31**, 445–452.
38. Sanz, G. (2005) Comparison of odorant specificity of two human olfactory receptors from different phylogenetic classes and evidence for antagonism. *Chem. Senses*, **30**, 69–80.
39. Geithe, C., Protze, J., Kreuchwig, F., Krause, G. and Krautwurst, D. (2017) Structural determinants of a conserved enantiomer-selective carvone binding pocket in the human odorant receptor OR1A1. *Cell. Mol. Life Sci.*, **74**, 4209–4229.
40. Aranceda, R.C., Peterlin, Z., Zhang, X., Chesler, A. and Firestein, S. (2004) A pharmacological profile of the aldehyde receptor repertoire in rat olfactory epithelium. *J. Physiol.*, **555**, 743–756.
41. Saito, H., Chi, Q., Zhuang, H., Matsunami, H. and Mainland, J.D. (2009) Odor coding by a mammalian receptor repertoire. *Sci. Signal.*, **2**, ra9.
42. Keller, A., Zhuang, H., Chi, Q., Vosshall, L.B. and Matsunami, H. (2007) Genetic variation in a human odorant receptor alters odour perception. *Nature*, **449**, 468–472.
43. Geithe, C., Andersen, G., Malki, A. and Krautwurst, D. (2015) A butter aroma recombinant activates human Class-I odorant receptors. *J. Agr. Food Chem.*, **63**, 9410–9420.
44. Fujita, Y., Takahashi, T., Suzuki, A., Kawashima, K., Nara, F. and Koishi, R. (2007) Deorphanization of Dresden G protein-coupled receptor for an odorant receptor. *J. Recept. Sig. Transd.*, **27**, 323–334.
45. Wetzel, C.H., Oles, M., Wellerdieck, C., Kuczkowiak, M., Gisselmann, G. and Hatt, H. (1999) Specificity and sensitivity of a human olfactory receptor functionally expressed in human embryonic kidney 293 cells and xenopus laevis oocytes. *J. Neurosci.*, **19**, 7426–7433.
46. Kajiya, K., Inaki, K., Tanaka, M., Haga, T., Kataoka, H. and Touhara, K. (2001) Molecular bases of odor discrimination: reconstitution of olfactory receptors that recognize overlapping sets of odorants. *J. Neurosci.*, **21**, 6018–6025.
47. Jones, E.M., Jajoo, R., Cancilla, D., Lubock, N.B., Wang, J., Satyadi, M., Cheung, R., de March, C., Bloom, J.S., Matsunami, H., et al. (2019) A scalable, multiplexed assay for decoding GPCR-ligand interactions with RNA sequencing. *Cell Syst.*, **8**, 254–260.
48. Frey, T., Kwadha, C.A., Haag, F., Pelletier, J., Wallin, E.A., Holgersson, E., Hedenström, E., Bohman, B., Bengtsson, M., Becher, P.G., et al. (2022) The human odorant receptor OR10A6 is tuned to the pheromone of the commensal fruit fly *Drosophila melanogaster*. *iScience*, **25**, 105269.
49. Spehr, M., Gisselmann, G., Poplawski, A., Riffell, J.A., Wetzel, C.H., Zimmer, R.K. and Hatt, H. (2003) Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science*, **299**, 2054–2058.
50. Geithe, C., Noe, F., Kreissl, J. and Krautwurst, D. (2016) The broadly tuned odorant receptor OR1A1 is highly selective for 3-Methyl-2,4-nonanedione, a key food odorant in aged wines, tea, and other foods. *Chem. Senses*, **42**, 181–193.
51. Jacquier, V., Pick, H. and Vogel, H. (2006) Characterization of an extended receptive ligand repertoire of the human olfactory receptor OR17-40 comprising structurally related compounds. *J. Neurochem.*, **97**, 537–544.
52. Gonzalez-Kristeller, D.C., do Nascimento, J.B.P., Galante, P.A.F. and Malnic, B. (2015) Identification of agonists for a group of human odorant receptors. *Front. Pharmacol.*, **6**, 35.
53. Baud, O., Yuan, S., Veya, L., Filipek, S., Vogel, H. and Pick, H. (2015) Exchanging ligand-binding specificity between a pair of mouse olfactory receptor paralogs reveals odorant recognition principles. *Sci. Rep.*, **5**, 14948.
54. Yoshikawa, K., Deguchi, J., Jieying, H., Lu, H.-Y. and Matsunami, H. (2022) An odorant receptor that senses four classes of musk compounds. *Current Biology*, **32**, 5172–5179.
55. McRae, J.F., Mainland, J.D., Jaeger, S.R., Adipietro, K.A., Matsunami, H. and Newcomb, R.D. (2012) Genetic variation in the odorant receptor OR2J3 is associated with the ability to detect the “Grassy” smelling Odor, cis-3-hexen-1-ol. *Chem. Senses*, **37**, 585–593.
56. Weber, L., Al-Refae, K., Ebbert, J., Jägers, P., Altmüller, J., Becker, C., Hahn, S., Gisselmann, G. and Hatt, H. (2017) Activation of odorant receptor in colorectal cancer cells leads to inhibition of cell proliferation and apoptosis. *PLoS One*, **12**, e0172491.
57. Weber, L., Schulz, W.A., Philippou, S., Eckardt, J., Ubrig, B., Hoffmann, M.J., Tannapfel, A., Kalbe, B., Gisselmann, G. and Hatt, H. (2018) Characterization of the Olfactory Receptor OR10H1 in Human Urinary Bladder Cancer. *Front. Physiol.*, **9**, 456.
58. Saito, N., Yamano, E., Ishii, A., Tanaka, M., Nakamura, J. and Watanabe, Y. (2018) Involvement of the olfactory system in the induction of anti-fatigue effects by odorants. *PLoS One*, **13**, e0195263.
59. Abaffy, T., Bain, J.R., Muehlbauer, M.J., Spasojevic, I., Lodha, S., Bruguera, E., O’Neal, S.K., Kim, S.Y. and Matsunami, H. (2018) A testosterone metabolite 19-Hydroxyandrostenedione induces neuroendocrine trans-differentiation of prostate cancer cells via an ectopic olfactory receptor. *Front. Oncol.*, **8**, 162.
60. Huysseune, S., Locrel, M., Veithen, A., Philippeau, M. and Quesnel, Y. (2020) XXIXth Annual Meeting of the European Chemoreception Research Organization, ECRO 2019. *Chem. Senses*, **45**, 119–174.
61. Alex, V. (2015) Abstracts from the 24th Annual Meeting of the European Chemoreception Research Organization (ECRO 2014), Dijon, France, September 10–13th, 2014. *Chem. Senses*, **40**, 211–297.
62. Pierre, C., Françoise, W., Magali, P., Alex, V., Olivier, N. and Yannick, Q. (2017) XXVth Annual Meeting of the European Chemoreception Research Organization, ECRO 2016. *Chem. Senses*, **42**, e1–e59.

63. Ieki,T., Yamanaka,Y. and Yoshikawa,K. (2022) Functional analysis of human olfactory receptors with a high basal activity using LNCaP cell line. *PLoS One*, **17**, e0267356.
64. Laing,D.G., Legha,P., Jinks,A.L. and Hutchinson,I. (2003) Relationship between molecular structure, concentration and odor qualities of oxygenated aliphatic molecules. *Chem. Senses*, **28**, 57–69.
65. Peterlin,Z., Firestein,S. and Rogers,M.E. (2014) The state of the art of odorant receptor deorphanization: a report from the orphanage. *J. Gen. Physiol.*, **143**, 527–542.
66. Ieki,T., Yamanaka,Y. and Yoshikawa,K. (2022) Functional analysis of human olfactory receptors with a high basal activity using LNCaP cell line. *Plos one*, **17**, e0267356.
67. Hladiš,M., Lalis,M., Fiorucci,S. and Topin,J. (2023) Matching receptor to odorant with protein language and graph neural networks. In: *The Eleventh International Conference on Learning Representations*.
68. Goodman,J.M., Pletnev,I., Thiessen,P., Bolton,E. and Heller,S.R. (2021) InChI version 1.06: now more than 99.99% reliable. *J. Cheminformatics*, **13**, 40.
69. Kim,S., Chen,J., Cheng,T., Gindulyte,A., He,J., He,S., Li,Q., Shoemaker,B.A., Thiessen,P.A., Yu,B., *et al.* (2023) PubChem 2023 update. *Nucleic Acids Res.*, **51**, D1373–D1380.
70. The UniProt Consortium (2023) UniProt: the Universal Protein knowledgebase in 2023. *Nucleic Acids Res.*, **51**, D523–D531.
71. Lu,M., Echeverri,F. and Moyer,B.D. (2003) Endoplasmic reticulum retention, degradation, and aggregation of olfactory G-protein coupled receptors. *Traffic*, **4**, 416–433.
72. Ignatieva,E., Levitsky,V., Yudin,N., Moshkin,M. and Kolchanov,N. (2014) Genetic basis of olfactory cognition: extremely high level of DNA sequence polymorphism in promoter regions of the human olfactory receptor genes revealed using the 1000 Genomes Project dataset. *Front. Psychol.*, **5**, 257.
73. de March,C.A., Topin,J., Bruguera,E., Novikov,G., Ikegami,K., Matsunami,H. and Golebiowski,J. (2018) Odorant receptor 7D4 activation dynamics. *Angew. Chem.*, **130**, 4644–4648.
74. Bushdid,C., De March,C., Fiorucci,S., Matsunami,H. and Golebiowski,J. (2018) Agonists of G-protein-coupled odorant receptors are predicted from chemical features. *J. Phys. Chem. Lett.*, **9**, 2235–2240.
75. Johnson,M., Zaretskaya,I., Raytselis,Y., Merezukh,Y., McGinnis,S. and Madden,T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.*, **36**, W5–W9.
76. Kowalewski,J. and Ray,A. (2020) Predicting human olfactory perception from activities of odorant receptors. *iScience*, **23**, 101361.
77. Gupta,R., Mittal,A., Agrawal,V., Gupta,S., Gupta,K., Jain,R.R., Garg,P., Mohanty,S.K., Sogani,R., Chhabra,H.S., *et al.* (2021) OdoriFy: a conglomerate of artificial intelligence-driven prediction engines for olfactory decoding. *J. Biol. Chem.*, **297**, 100956.
78. Ballesteros,J.A. and 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: *Methods in Neurosciences*. Elsevier, Vol. **25**, pp. 366–428.

Received: August 13, 2023. Revised: September 13, 2023. Editorial Decision: September 27, 2023. Accepted: October 4, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

3.2 – ODORANT BINDING PROTEIN

- M. Lalis, L. Moitrier, M. Jäger, C. Meinert, M. Brulé, C. Belloir, N. C. Jones, S. V. Hoffmann, S. Fiorucci, S. Wolf, L. Briand, J. Topin*. How Allosteric Mutations Control Ligand Binding in Lipocalin Protein: Odorant Binding Protein as a Test Case, [*Cellular and Molecular Life Science*](#), **2025**.

In this study, I led by a multidisciplinary consortium to investigate how lipocalins recognize and bind ligands. Using computational modeling and experimental validation, we identified key structural regions that control ligand entry and demonstrated the impact of mutations on binding. Our results also show how allosteric effects modulate ligand interactions. These insights, supported by a genome-wide analysis, reveal conserved mechanisms across the lipocalin family and open new avenues for biotechnological applications.

ORIGINAL ARTICLE



How allosteric mutations control ligand binding in Lipocalin protein: odorant binding protein as a test case

Maxence Lalis¹ · Lucie Moitrier² · Miriam Jäger³ · Cornelia Meinert¹ · Marine Brulé² · Christine Belloir² · Nykola C. Jones⁴ · Søren V. Hoffmann⁴ · Sébastien Fiorucci¹ · Steffen Wolf³ · Loïc Briand² · Jérémie Topin¹

Received: 22 January 2025 / Revised: 6 May 2025 / Accepted: 1 June 2025
 © The Author(s) 2025

Abstract

Lipocalins are a family of proteins found in mammals that are essential for the binding and transport of various molecules, but the mechanisms underlying their target recognition are still unclear. To answer this question, we studied odorant-binding proteins (OBPs), a specific type of lipocalin involved in chemical communication and olfaction. Using an integrative approach combining numerical modelling and experimental validation, we identified key structural regions that regulate the entry of molecules into the binding pocket. Modification of these regions disrupts molecular recognition, highlighting their importance for function. In addition, we found that changes in distant parts of the protein influence binding, shedding light on allosteric mechanisms. These results advance our understanding of lipocalin function and open up avenues for the design of proteins with targeted binding properties.

Keywords Molecular dynamics · Site-directed mutagenesis · Circular dichroism · Isothermal titration calorimetry · Lipocalin · Odorant binding protein

Introduction

Lipocalins form a group of small, soluble proteins that play a crucial role in the transport, storage or sequestration of small molecules in mammals [1], that recognize a wide range of ligand [2–4]. Their ability to bind a wide range of lipophilic and other organic compounds underlines their essential role in mammals [5, 6]. They share a conserved

tertiary structure, known as the lipocalin fold, which consists of an eight-stranded antiparallel β -barrel enclosing an internal binding cavity designed as the calyx [7–9]. However, the translocation process involved in ligand binding in lipocalin is still poorly understood [10], and this lack of knowledge limits the full use of lipocalins for potential applications. Advances in research into the control of the translocation mechanism could open up new avenues for the use of lipocalin in various biotechnology applications such as diagnostics, quality control and therapeutics [11–15].

To investigate the ligand translocation pathway, we used mammalian odorant binding proteins (OBPs) as a model for the lipocalin family [16]. Elucidation of OBP structures has shown that this protein family adopts a similar fold with a large hydrophobic cavity in its core (Fig. 1A) [17, 18]. These proteins are involved in various functions, including removing odorants from the olfactory epithelium [19], implicated in the immune response [20], or scavenging free radicals in the nasal mucosa [21].

For OBPs, the calyx has been shown to be involved in the capture and transport of odorants as the presence of this hydrophobic cavity is a key feature that allows reversible binding of odorants [22, 23]. Previous studies combining site-directed mutagenesis with in vitro assays have

Maxence Lalis and Lucie Moitrier contributed equally to this work.

✉ Loïc Briand
loic.briand@inrae.fr

✉ Jérémie Topin
jeremie.topin@univ-cotedazur.fr

¹ Institut de Chimie de Nice UMR7272, Université Côte d'Azur, CNRS, 28 Avenue Valrose, Nice 06108, France

² Centre des Sciences du Goût et de l'Alimentation, Institut Agro, CNRS, INRAE, Université de Bourgogne, Dijon F-21000, France

³ Biomolecular Dynamics, Institute of Physics, University of Freiburg, 79111 Freiburg, Germany

⁴ Department of Physics and Astronomy, Aarhus University, Aarhus 8000, Denmark

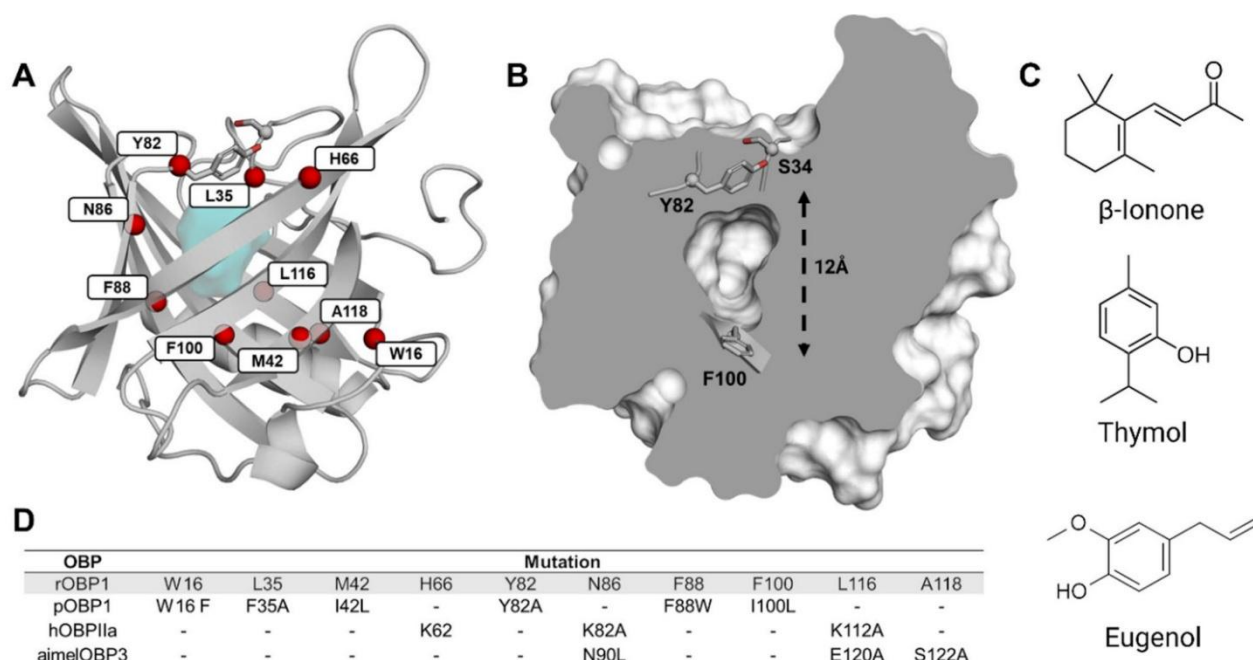


Fig. 1 Structure of rOBP1 and position of the known mutants. **(A)** Cartoon representation of rOBP1 (pdb code 3fiq [27]). Positions previously subjected to point mutation are denoted by red spheres. The hydrophobic calyx appears as a light blue surface in the core of the protein. The two amino acids (Y82 and S34) predicted to act as a gate controlling ligand-binding are represented as sticks with carbon and

oxygen atoms in grey and red, respectively. **(B)** Sliced surface representation of rOBP1 in the same orientation as A. The hydrophobic calyx appears as a void in the center of the protein. **(C)** Lewis representation of the molecules studied in this work. **(D)** Table summarizing positions which have been subjected to mutation in different mammalian OBPs

demonstrated that the binding specificity of OBPs could be largely affected by replacing amino acids in the calyx (Fig. 1A, D and table S1). In addition, post-translational modifications at residues not directly involved in ligand recognition could also affect the different binding affinities of OBPs for specific ligands [24, 25], suggesting an allosteric modulation of the binding. However, the binding pathway of odorants to OBPs calyx remains relatively under explored, particularly regarding the molecular mechanisms that govern the entry and exit of ligands from the binding pocket. Structural studies of OBP revealed that the cavity inside the β -barrel structure is not directly exposed to the surrounding solvent [17, 22] (Fig. 1 and fig. S1). This implies the existence of a dynamic process inducing a protein reorganisation to create a channel through which odorants will diffuse. However, this translocation pathway has never been observed in structural studies in which the calyx remains occluded from the solvent (fig. S1). Molecular dynamics simulations have suggested that a conserved tyrosine (Y82 in Fig. 1) controlled the access to the calyx [26]. Despite its high conservation (82%) in OBPs and other lipocalins, the precise molecular mechanisms governing the binding pathway, and its role in different lipocalins remain to be fully elucidated. Moreover, as most site-directed mutagenesis has been performed inside the calyx and only a few in the

vicinity of the gate (Fig. 1A, D), possible allosteric effects on ligand binding remain unknown.

In this study, we performed extensive equilibrium MD simulations [28–30] to decipher the mechanism of odorant binding to OBPs. OBPs from rat (rOBP1) and pig OBP1 (pOBP1) were considered for in silico experiments. This step provides molecular insights into the odorant binding pathway for 3 known rOBP1 binders [31] (Fig. 1C) and one known binder (2,6-dimethyl-7-octen-2-ol) for pOBP1 [32]. Additionally, we performed dissipation-corrected targeted MD simulations (dcTMD) [33, 34] together with pathway clustering [35, 36] to determine the free energy profile underlying the binding and unbinding process, as well as friction factors that reveal ligand hydration shell changes. The simulations revealed a conserved pathway linking the bulk to the calyx.

In order to confirm that odorants follow this single pathway to reach the binding site, a double mutant was constructed to block the gate in closed conformation. Residues serine34 and tyrosine82 were mutated to cysteine to create a disulfide bridge that would close the OBP gateway and prevent binding to odorants. The fold and stability of this mutant was assessed by synchrotron radiation circular dichroism spectroscopy [37] and confirmed the formation of a disulfide bridge. The loss of function of this mutant

was assessed by isothermal calorimetry (ITC) experiments [38]. Finally, guided by protein network analysis, we identified an allosteric site regulating odorant binding, which was confirmed through mutagenesis and ITC measurements. Altogether, our results provide a detailed description at the atomic level of the binding process in OBP and show how specific mutations could alter the binding process by blocking the translocation pathway or disturbing protein dynamics. By analyzing the conservation of these residues we propose that this binding mechanism is a conserved feature across the lipocalin family.

Materials and methods

Molecular Dynamics

All unbiased simulations were carried out using the AMBER20 package and the FF14SB force field [39]. Protein structures of rOBP1 and pOBP1 were taken from X-ray structures (accession code: 3fiq [27] and 1e00 [22]), the protonation state at pH 7.5 of their amino acids was computed using the H++ webserver [40]. Ligands were optimized at the B3LYP/6-31G* level using Gaussian 19 [41]. Partial charges for each atom were then obtained using the antechamber module. Finally, ligand parameters were obtained from the Generalized Amber Force Field (GAFF2).

In order to increase the sampling of binding events, 5 ligands were introduced randomly around the protein (fig. S2). The simulation box was completed using TIP3P water molecules and neutralized using Na⁺. The total system is made up of 35,779 atoms, in a 358,566 Å³ periodic box. Shake algorithm was used to apply bonds involving hydrogen atoms allowing to set a time step of 2 fs during MDs simulations.

Systems were minimized by 10,000 minimization steps and 20 ps of MD simulations at 100 K using a restraint of 20 kcal·mol⁻¹·Å⁻² on the protein were performed, followed by four rounds of 10,000 minimization steps reducing the restraints by 5 kcal·mol⁻¹·Å⁻² at each round, with 20 ps MD simulation in between each step. Then, the systems were slowly heated to 300 K over a period of 20 ps. Equilibration runs were continued over 20 ns prior to the production simulations. Production simulations of 200 ns were extended if no binding event was observed. The entrance of the ligand inside the binding cavity was monitored by its distance to the F100 residue for rOBP1 (fig. S3). During the 10 replicas, rOBP1 underwent small fluctuations ($\text{RMSD}_{\text{backbone}} < 2 \text{ Å}$) showing great stability (fig. S4 and S5) of the OBP fold. Similar behavior was observed for pOBP1-wt (fig. S6). During the MD simulations, the C-terminal part is stabilized

near the calyx by a disulfide bond between two cysteine: residues 79 and 170 in rOBP1 and 63 and 155 in pOBP1.

Clustering was done using the DBSCAN clustering algorithm [42] with a minimum number of 50 points required to form a cluster and an epsilon of 5, which is the distance cut-off between points required to form a cluster. A minimum of 50 points corresponds to a cluster that persists for 10 ns during simulation. Mass-weighted RMSD was used to define distance between frames. RMSD distance cut-off of about 5 Å was chosen to explore the same binding pose while allowing moderate internal flexibility and various orientations.

dcTMD simulations and MoSAIC-based trajectory clustering

During dissipation corrected target MD (dcTMD), a pulling force is applied along a reaction coordinate (x) using a displacement distance constraint. This method yields both the free energy profile $\Delta G(x)$ and the friction field $\Gamma(x)$. The friction captures dynamical interactions with degrees of freedom that are not directly involved in the free energy landscape and allows monitoring the hydration of the binding site [34].

Dissipation-corrected targeted MD simulations were carried out using the PULL code in Gromacs v2020.3 [43] with the Amber14 forcefield and the TIP3P water model. Ligand parameters were imported from AMBER using acpype [44]. As a starting structure, one of the final simulation structures with bound β -ionone was used, which was first equilibrated with a steepest descent minimization, followed by a 100 ps simulation in the NPT ensemble at 300 K and 1 bar, each with positional restraints of 2 kcal·mol⁻¹·Å⁻² on protein and ligand heavy atoms. The full system was then subjected to another round of steepest descent minimization, followed by another 100 ps in the NPT ensemble with protein and ligand positional restraints, finally followed by 10 ns of fully unbiased equilibration.

From the final simulation structure, 200 statistically independent replicas were generated by the independent attribution of atomic velocities. After 10 ps of heating to 300 K in the NPT ensemble with positional restraints on protein and ligand heavy atoms, each simulation replica was subjected to another 10 ps of short equilibration without positional restraints, but with a constant distance constraint between the centers of masses of the β -ito another 10 ps of short equilibration without positional restraints, but with a constant distance constraint between the centers of masses of the β -ionone heavy atoms and 26 C α -atoms of the protein's central β -sheet. Finally, all 200 simulation replicates were continued for 2.5 ns, replacing the constant distance to a constant velocity constraint with a pulling velocity of 1 m·s⁻¹.

The resulting trajectories were clustered based on a principal component analysis of protein-ligand contacts [36, 45] using the first four principal components to calculate trajectory-pairwise similarities s_{ij} . Unbinding paths were revealed via Leiden community detection [46] using the MoSAIC tool [47], with the constant Potts model as objective function. The constant Potts model requires only the definition of a single parameter, the resolution parameter γ , thereby requiring minimal user input (see the supplementary materials for details). Here, we use a resolution parameter $\gamma = \text{median}(s_{ij})$, which has to be shown to be effective for path separation. Work profiles of trajectories of the largest cluster were calculated using dcTMD [33, 36], yielding free energies and friction factors along the unbinding path.

Network and MoSAIC Analysis of protein dynamics

To assess the global communication profile of our protein, we analyzed the global metapath using webPSN webserver with default settings [48], which represents the shortest communication pathway between all pairs of residues in the network. The global metapath offers a comprehensive view of the overall communication landscape and helps to pinpoint critical residues that contribute to the protein's function and regulation.

We also evaluated the correlated dynamics of the OBP from MD simulations with MoSAIC [47]. As input for the analysis we determined pairwise minimal residue contact distances with a heavy atom distance cut-off of 4.5 Å [45]. Then, we calculated the Pearson correlation ρ between the distances and clustered the resulting pair-wise correlation matrix [47] (see supplementary material for more details). We identified well-connected clusters using a resolution factor $\gamma=0.2$. This threshold was chosen to ensure that the resulting clusters represent functionally meaningful groups of residues. Lower thresholds resulted in clusters with weak internal correlation, which restricted their interpretability in terms of collective motion.

Sequence alignment

Sequences were taken from Pelosi et al. [16]. Alignment was made with Jalview using the Mafft algorithm and used to calculate sequence conservation with Jalview [49] (Supplementary files 1). Weblogo were constructed using Weblogo3 [50] and phylogenetic tree with the online web-server iTOL [51].

rOBP1 production and purification

Recombinant rOBP1-wt (wild-type) and mutants were produced as previously described [52], see fig. S7 for

more details about the sequence and construction. Briefly, the cDNA sequence encoding the mature rOBP1 (Gene name: Obp1f, UniProt ID: P08937) was synthesized with codon usage optimized for expression in *E. coli* (Genewiz, Leipzig, Germany) and cloned into the *Bam*HI and *Hind*III restriction sites of the pQE31 plasmid (Qiagen, Hilden, Germany). The resulting expression plasmid pQE31-rOBP1 encodes a fusion protein composed of an N-terminal His6-tag, followed by rOBP1 (His16-Gln172). Mutants were prepared using customized site-directed mutagenesis by Genewiz (Leipzig, Germany). The plasmids were subsequently introduced into *E. coli* M15 [pREP4] cells (pQE expression system, Qiagen). The expression of rOBP1-wt and mutants were induced with 0.2 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) for 3 h at 37 °C in Terrific Broth medium supplemented with antibiotics.

Recombinant His-tagged rOBP1-wt and mutant proteins were purified as previously described using standard Nickel-affinity chromatography. To remove potentially bound contaminants, the purified rOBP1-wt and mutant proteins were extensively dialyzed for 24 h against a buffer (100 mM sodium phosphate, pH 7.5) containing 5% (v/v) acetonitrile at 4 °C. After an additional 24 h of dialysis against the same buffer in the absence of acetonitrile, fractions containing rOBP1 proteins were analyzed by SDS-PAGE, pooled and stored at -20 °C. All buffers used for purification were prepared with HPLC-grade water. The purity of purified proteins was assessed by SDS-PAGE, and their concentrations were determined with UV spectroscopy.

Recombinant rOBP1 characterization

The purity of rOBP1-wt and the mutants was assessed by SDS-PAGE, and concentrations determined by spectrophotometry, measuring the absorbance at 280 nm and using an extinction coefficient of $\epsilon_{280} = 16,305 \text{ M}^{-1} \cdot \text{cm}^{-1}$. The oligomeric state of rOBP1 was analyzed by size exclusion chromatography using a 24-mL bed volume Superdex 75 10/300 GL column (GE Healthcare). The column was equilibrated in a solution containing 100 mM sodium phosphate and 150 mM NaCl at pH 7.5 at 1 mL·min⁻¹. Bovine serum albumin (67 kDa), chicken egg ovalbumin (43 kDa), bovine ribonuclease A (13.7 kDa), and aprotinin (6.5 kDa) purchased from Sigma, were employed as standards (fig. S8).

Isothermal titration calorimetry

Isothermal titration microcalorimetry (ITC) experiments were carried out at 25 °C using a VP-ITC microcalorimeter (Malvern Instruments, Malvern, UK) as previously described [52, 53]. Protein solutions (25 μM in 50 mM sodium phosphate at pH 7.5) were thoroughly degassed

before the experiments. Odorant solutions (250 μM in 50 mM phosphate buffer, pH 7.5) were injected into 25 successive 10 μL aliquots. Binding parameters were calculated by fitting the values to a single site binding model through nonlinear regression using Microcal Origin[®] software.

Synchrotron radiation circular Dichroism spectroscopy

Electronic Circular Dichroism (ECD) and absorbance spectra of rOBP1 were recorded at the AU-CD beam line on ASTRID2 at Aarhus University (Denmark) [54]. rOBP1-wt and rOBP1 mutant (S34C Y82C) samples were diluted in a 100 mM sodium phosphate solution at pH 7.5 to achieve a concentration of $\sim 1 \text{ mg}\cdot\text{mL}^{-1}$. All spectra were measured using a nominally 0.1 mm path length cell (quartz Suprasil cell, Hellma GmbH & Co., Germany), with the precise length determined using interferometry measurements [55]. The concentration of the samples were determined from the absorbance at 205 nm [56], which is measured simultaneously with the CD spectrum. Spectra were recorded at 25 °C in triplicate in the wavelength range of 170 to 280 nm in steps of 1 nm and a dwell time of 2 s, with a corresponding baseline triplicate scans measured of the buffer alone. Finally, the percentages of secondary structure were obtained by a deconvolution of the spectra using BeStSel [57] webserver.

Temperature scans were carried out in a Starna type 31B cell, with a nominal pathlength of 0.1 mm, with the full CD spectrum measured in triplicate at each set temperature, in steps of 2.5 °C and an equilibration time of 1 min before scanning. Scans of the reference baseline measured with the same protocol were subtracted for each temperature.

A summary of the different methodologies can be found as fig. S9.

Results

Ligand entry mechanism from MD simulations

To clearly identify ligand translocation in OBP, we performed MD simulations to track the entry of three ligands, β -ionone, eugenol and thymol (Fig. 1C), into the calyx of rOBP1-wt. A total of 15 simulations, resulting in an aggregated simulation time of 22 μs , were done. We considered the ligand in the cavity when the distance between its center of mass and the center of mass of residue F100, forming the cradle of the cavity, was less than 5 Å. During these simulations, 10 binding events were observed (fig. S3). The same protocol was applied to pOBP1 and 5 binding events recovered through 10 independent simulations.

The MD simulations revealed a conserved pathway for all the considered ligands. The analysis of a prototypical binding trajectory is presented in Fig. 2. To access the binding cavity, ligands diffuse through a highly dynamic gate formed by the conserved Y82 and the S34 (Fig. 2A). The aromatic tyrosine acts as a hydrophobic cap, occluding the cavity from the solvent. On the basis of the side chain fluctuation and odorant solvation, the binding pathway is divided into five subsequent steps: contact-closed, pocket-closed, entry-open, binding site-open, and binding site-closed (Fig. 2A-C, a, b, c, d, e, respectively).

The first step (contact-closed, a) involves rapid contact with the vicinity of the gate and partial dehydration of the ligand ($\sim 25\%$) and no specific interaction with the protein is clearly observed.

The second step (pocket-closed, b) consists of stabilisation of the ligand in the vicinity of the gate for $\sim 70 \text{ ns}$ and stable solvation at 60%. Ligand migration seems to depend on the conformation of the Y82 side chain, whose closed form is stabilised by a hydrogen bond with the backbone of S34, which is located at the opposite side of the entrance. This interaction must be broken to allow the ligand to enter the protein. Simulations of the apo form of the protein revealed spontaneous opening and closing of the gate, even in the absence of ligand (fig. S10A). This observation suggests that the dynamic of the door opening is independent of the presence of a ligand in the vicinity of the gate and that the binding follows a conformational selection rather than an induced fit process.

The third step (entry-open, c) begins with the breaking of the hydrogen bond between S34 and Y82, increasing the degree of freedom of the loop and causing the switching of Y82, which opens the cavity and allows entry of the ligand. This migration is accompanied by a sharp drop in ligand solvation to less than 20%. Ligand entry into the calyx protein is rapid and appears to be irreversible on the time scale of the simulation as we do not observe any unbinding event.

When the ligand enters the cavity, we observe almost complete solute dehydration (step d in Fig. 2, binding site open). Then, the Y82 sidechain undergoes a rotation that induces the closing of the cavity (step e in Fig. 2A, B and C, binding site-closed). The ligand does not adopt a specific binding position, and appears to be mostly stabilized through hydrophobic contact (table S2) as observed in a recent study [58]. The simulations recover most of the ligand interactions with amino acids previously identified as playing a key role in binding (table S2). Finally, simulations realized on pOBP1 unveil a similar pathway for the dihydromyrcenol binding (fig. S11), suggesting a conserved binding pathway across mammalian OBPs.

Throughout the simulations, we observed a variable duration of the pocket-locked step, which depends on the

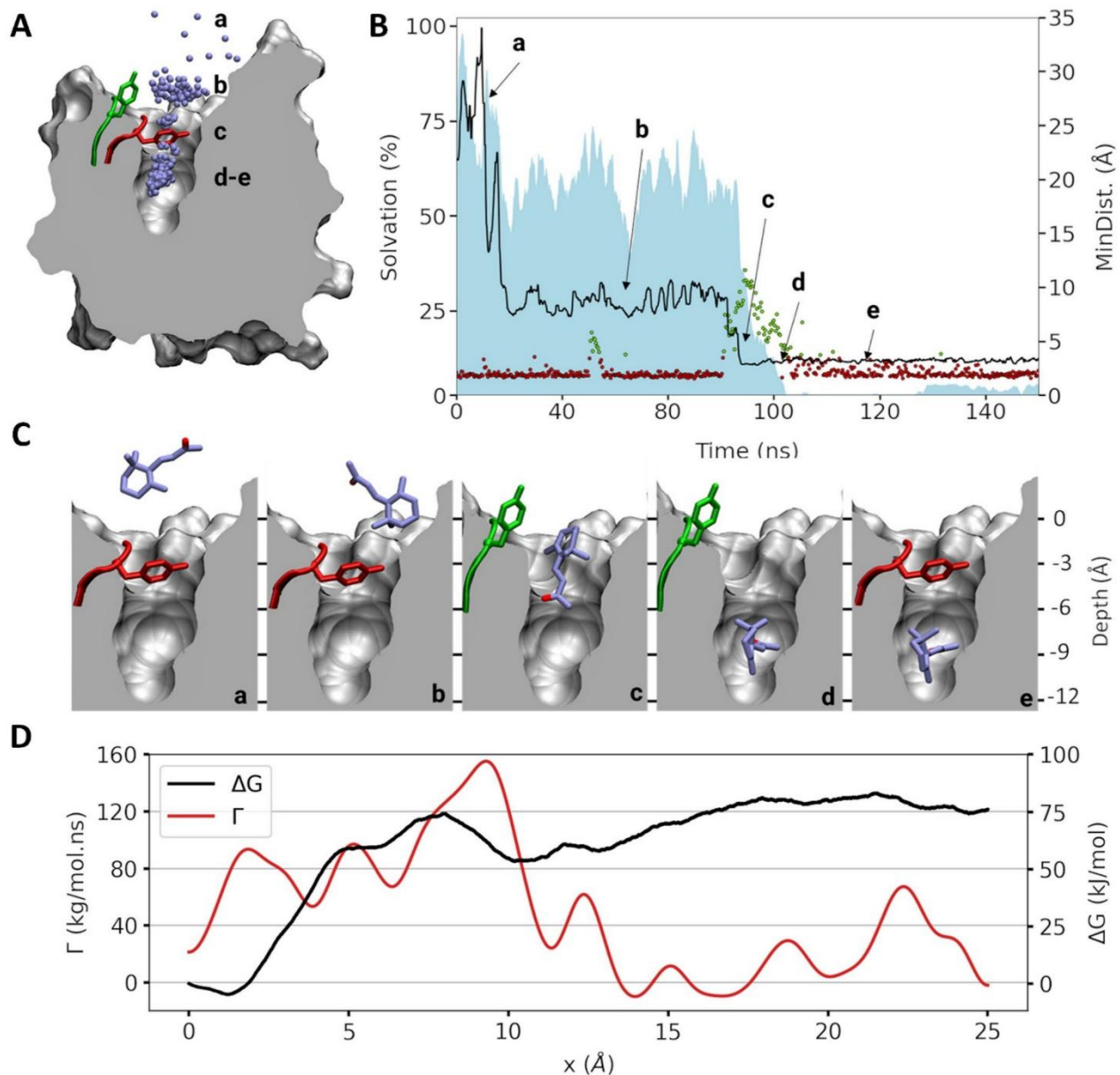


Fig. 2 Translocation mechanism of β -ionone in rOBP1. **(A)** Sliced surface representation showing the center of mass of β -ionone at every timestep in lavender balls. Y82 is shown either in red or green for closed or open state respectively. **(B)** Evolution of the distance between β -ionone centers of mass and the center of mass of the F100 lateral chain drawn as a black curve separated in different areas. Zone (a) represents the binder in water to the contact of rOBP1. Zone (b) represents the partial stabilization in the pocket, (c) the entry, (d) β -ionone in the cavity with the gate open and (e) in the cavity with the gate closed. The blue area shows the percentage of ligand solvation during the binding process normalized to the solvation of the ligand in

water. The distance between the hydroxyl group of Y82 and carbonyl of the backbone of S34 is represented in dots. Red dots correspond to the closed state (< 3.4 Å) and green to the open state. **(C)** Close view of the binding process with representative position for each zone defined in B. Carbons atoms of β -ionone are colored in lavender and oxygen in red. Y82, and its loop are colored according to its opening (red closed, green open). **(D)** Free energy (red) and friction (black) estimate from dcTMD unbinding simulations. The free energy curve exhibits a local minimum at $x = 10$ Å, corresponding to state c in the unbiased simulations shown above. The peak in friction at $x = 9$ Å coincides with the decrease in ligand hydration

gate opening. It appears to be a stochastic event, independent of the ligand presence (fig. S10C). Stabilization next to the gate does not necessarily lead to binding. If the gate does not switch to the open conformation (Fig. 2), the ligand cannot access the calyx of the protein and will dissolve back into the solvent (fig. S10C).

During the simulations, no unbinding event was observed (fig. S3). Thus, in order to sample the unbinding pathway, we performed dcTMD simulations, resulting in 200 unbinding events. Trajectory clustering (fig. S12) revealed that 130 of the 200 trajectories unbind along the same pathway as our unbiased binding path. The resulting free energy profile in Fig. 2D displays a local minimum at a distance of 10 Å from the binding site, which perfectly aligns with the pocket-locked position shown in Fig. 2C. In agreement with this observation, the peak in friction Γ at a distance of 9 Å agrees excellently with the decrease in ligand hydration. The respective free energy difference between the bound state and this local minimum is on the order of ca. 55 kJ·mol⁻¹, i.e., 13.5 kcal·mol⁻¹, which corresponds roughly to a nanomolar affinity for β -ionone, which is in reasonable agreement with our experimental values. We note that the remaining 25 kJ·mol⁻¹ for complete unbinding into the solvent, agree well with β -ionone's logP of 4.

The simulations thus lead to a complete description, at the atomic level, of the ligand binding pathway and guide site-directed mutagenesis experiments combined with ITC assays to assess the role of the gate and amino acids in its local environment.

Site-directed mutagenesis and ITC experiments confirm the translocation pathway

Confirmation of gating via mutagenesis

The simulations revealed the hydrogen bond between Y82 and S34 to be essential in controlling the ligand binding. We sought to strengthen the interaction between these two residues by creating a disulfide bond between these two positions which will lock the protein in the closed state and would prevent the ligand from binding to the calyx. This methodology has already been successful to abrogate ligand binding in lysozyme [59]. The distance between the two residue side-chains is fluctuating between 3 and 5 Å, such that it should be favorable to form a disulfide bond [60, 61]. We first generate a model of the double mutant Y82C-S34C with a disulfide bond in between the two cysteine residues (fig. S10B). No large change in the overall structure appeared during the MD simulations compared to the rOBP1-wt (fig. S4) suggesting that this double mutant should not affect rOBP1 folding.

We further confirmed these *in silico* results by investigating the folding of the double mutants by synchrotron radiation circular dichroism spectrometry [62, 63]. UV CD spectra recorded down to 180 nm for both wt and mutant are similar (Fig. 3A). The deconvolution analysis of the different spectra shows that the percentage of secondary structures are not significantly affected by the mutations (fig. S13) confirming that these double mutations have a minor impact on the protein secondary structure.

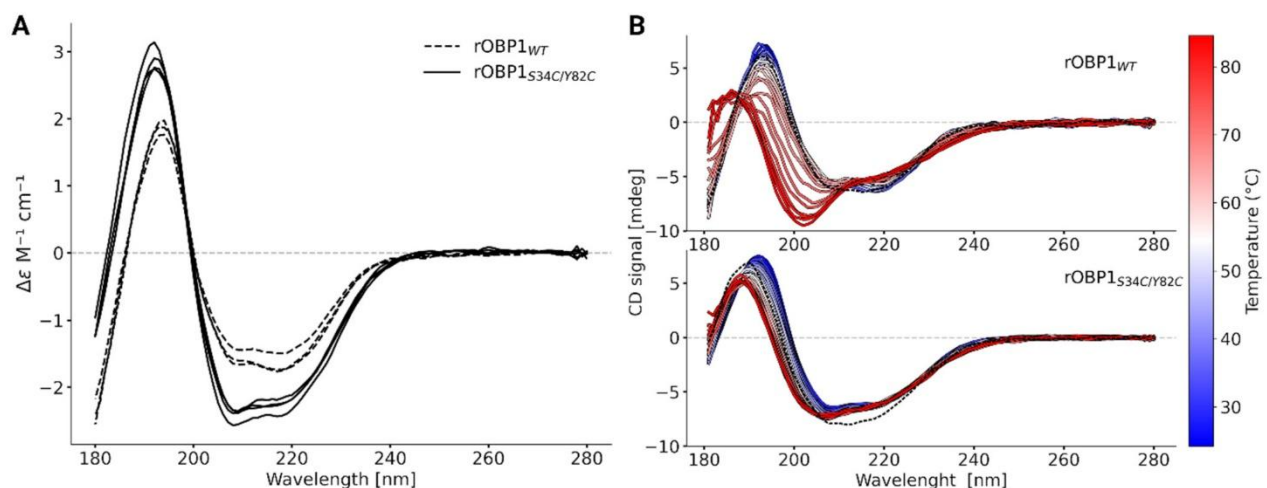


Fig. 3 Analysis of the rOBP1 mutant folding and stability. **(A)** Circular dichroism spectra of rOBP1-wt and double mutant S34C and Y82C in dashed and plain lines respectively. Three and four independent measurements have been made for the wt and double mutant, respectively. **(B)** Temperature scan plots displayed as a function of temperature,

with overlaid CD spectra recorded at temperatures ranging from 25 °C to 85 °C. The dotted black lines represent the spectra at 25 °C, while the series of colored lines correspond to spectra at each incremental temperature

Finally, the stability of both the wt and the double mutant was assessed through measurements of CD spectra with increasing temperature [64]. For rOBP1_{wt} the protein undergoes a clear change with temperature, unfolding and becoming more unordered, with a transition temperature of 70 °C. The unfolding of the wt was reversible, with the original spectrum largely recovered when returning to 25 °C (dotted black line Fig. 3B). The double mutant showed a higher stability than the wt, with a smaller change in the spectrum at higher temperatures, though the folding was not completely recovered when back at 25 °C (Fig. 3B). Overall, these results suggest that the mutations do not affect the folding and that the two cysteines are forming a disulfide bridge that stabilizes the protein structure.

The response of the rOBP1 double mutant response to agonists stimulation was assessed by isothermal titration calorimetry (ITC). This double mutation resulted in a total loss of binding for the three odorants as no heat change was observed (Fig. 4), further emphasizing the crucial role of the two residues in controlling the odorant binding. Moreover, this loss of function demonstrates that this gate is the sole entry point to the calyx and also confirms that there is a unique path for the ligand to access the binding site of the mammalian OBP.

Allosteric network analysis and distal mutation effects

We further investigated the local environment surrounding the gate. During the binding process, we regularly observe ligand transient stabilization at different positions around the protein. Ligands are stabilized in these pockets through weak interactions. Clustering analyses of the odorants' positions revealed two clusters adjacent to the gate across all studied systems (Fig. 4, fig. S14). These two pockets exhibit significant differences in hydrophobicity. The first pocket "3", contains a high proportion of conserved hydrophobic residues. In contrast, the second pocket "4", is characterized by a high proportion of conserved hydrophilic residues. We propose that hydrophobic pocket '4' is involved in partial dehydration and stabilization in step (b) (Fig. 2B). Simulations showed that L57 and V59 interacted significantly with odorants in this pocket.

Mutation of these residues to hydrophilic residues of variable size (V57Q/N and L59E/D) did not result in significant changes in binding affinities for none of the studied ligands (Fig. 4 and fig. S15). In all cases, the K_d remains within the same order of magnitude. However, we observed a compensation in the entropy-enthalpy balance: these quantities

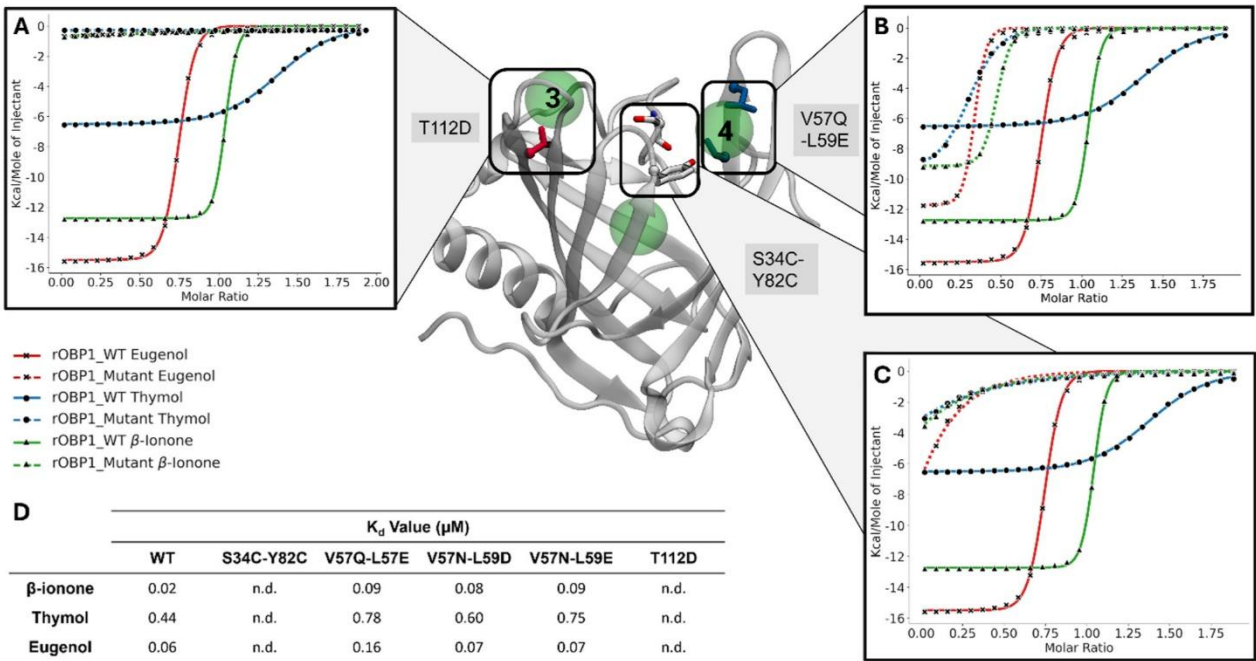


Fig. 4 ITC assays on rOBP1 mutants in the local environment of the gate. Cartoon representation of rOBP1 with cluster position of the ligands near the gate shown as green spheres. Mutated amino acids are represented in different colors. All mutants were tested with the three ligands. (A) Left box, shows ITC assays results from mutation in the hydrophilic pocket 3 with T112 in red. (B) Right box shows experimental results on mutation in the identified hydrophobic pocket

4 with V57 and L59 in blue. For the sake of clarity, results for the mutant V57D/N and L59E/N are not shown. (C) Middle box, shows ITC assays results on the mutant at positions S34 and Y82, represented in grey. All experiments are realized on 3 binders of rOBP1: β -ionone, thymol, and eugenol. (D) Table summarizing the K_d values for all the experiments, n.d. means not determined

increased for β -ionone and eugenol, while they decreased for thymol (table S3). Thus, our results show that the modification of the hydrophobic properties of this pocket does not influence the binding affinity but induces a variation in the thermodynamic binding signature.

We next focused on the opposite hydrophilic pocket. The conserved residue T112 was mutated into aspartic acid to further increase the hydrophilicity [65]. No detectable heat change was measured, even at ligand concentrations in the millimolar range, indicating a complete loss of affinity for all ligands (Fig. 4 and Fig. S15). This suggests that T112 is essential for odorant binding.

As the previously identified pockets are distant from the binding site but still influence binding, we hypothesize that they are involved in allosteric communication with the OBP's gate. To investigate this, we performed an interaction network analysis using PSNtools [48] on the crystallographic structure of rOBP1. This method allowed us to map allosteric communication between distant sites by identifying the shortest communication pathways through the interaction network (Fig. 5A). Residue T112 was found to be part of a communication network involving loops constituting the gate, and the adjacent α -helix. Disruption of this network may impair the dynamics of the gate. On the other hand, residues L59 and V57, which had minimal effect

on binding, were not part of this pathway, highlighting their negligible role in odorant binding to rOBP1.

Additionally, we analyzed correlated motions of the protein during the MD simulations using MoSAIC. This analysis identified small communities of correlated movements, with rOBP1 showing limited large-scale correlations (fig. S16), consistent with previous findings [66, 67]. Notably, the third-largest community includes the gate, the adjacent α -helix, and residue T112, suggesting coordinated movements between the gate and the region surrounding this residue (Fig. 5B). This suggests that T112 would be involved in the allosteric regulation of gate dynamics, influencing the overall binding process. In contrast, residues V57 and L59 were not part of any major motion networks, further confirming their minimal contribution to the protein's dynamics and binding mechanics.

Residue controlling the gating is highly conserved across lipocalin.

In this study, we observe that the binding and unbinding processes rely on the opening of a 'gate' formed by residues Y82 and S34. The gate remains closed thanks to a hydrogen bond between the hydroxyl group of the tyrosine and the carbonyl of the serine backbone. Once this hydrogen bond has been broken, the two opposing loops gain mobility and separate, creating an opening towards the calyx which allows the ligand to enter. This tyrosine residue is highly

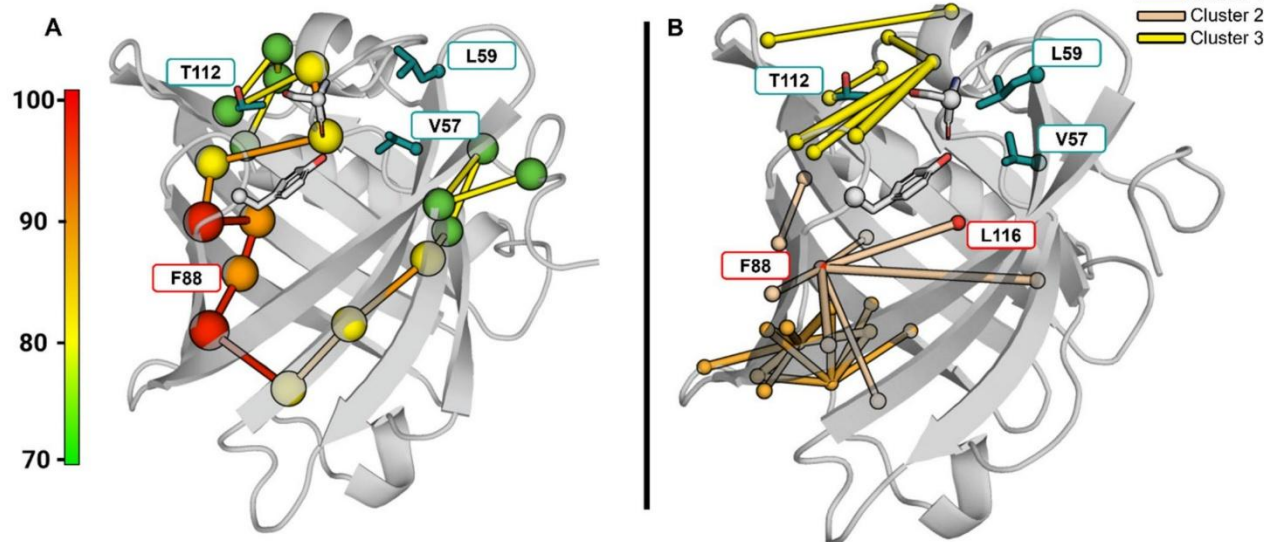


Fig. 5 Structural communication signature in the OBP structure probed by network analysis. Mutated positions (T112, S34, Y82, L59, V57) are represented in licorice. Carbon atoms from the gate (S34 and Y82) are colored white, while V57, L59 and T112 are in cyan. Mutations made on residues boxed in red have been previously shown to abrogate the ligand binding. **(A)** Global metapath obtained on the crystallographic structure of rOBP1 (pdb code 3fiq [27]). Nodes and links are colored with a gradient from green ($I=70$) to red ($I=100$), where $I=100$ is the most interacting node or link. Hubs are shown as spheres centered on

the $\text{C}\alpha$ -atoms and colored according to the average interaction strength of their links. Nodes and link with an interaction frequency I below 60 are not shown. **(B)** MoSAIC analysis results on all apo rOBP1-wt MD simulations. The highly correlated minimal distances appear as colored lines between the involved amino acid $\text{C}\alpha$ -atoms according to the cluster classification. Interestingly, the two previously identified positions F88 and L116 belong to the same cluster (cluster 2 in fig. S16). On the other hand, V57 and L59 are not highly correlated to the rest of the protein, as seen from the analysis of crystallographic structure

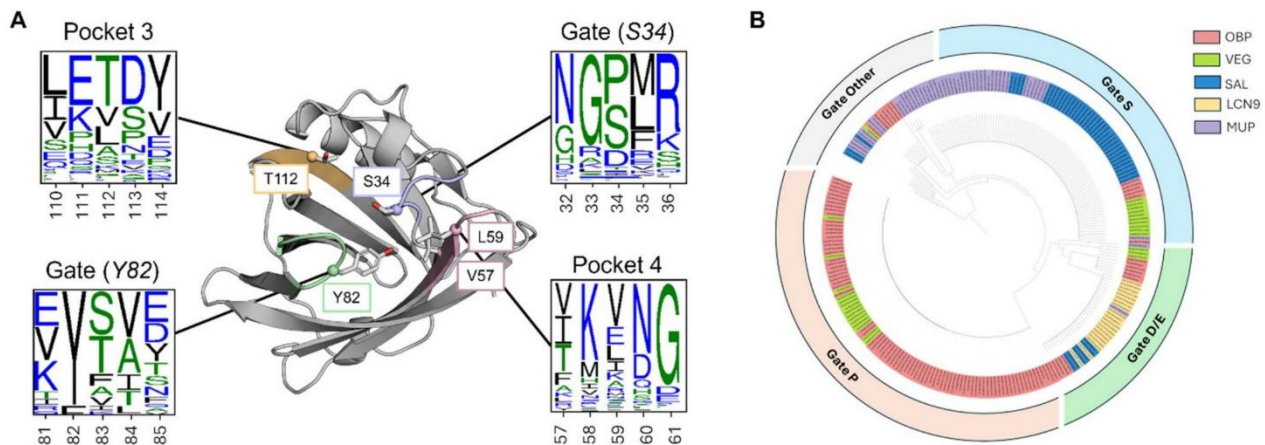


Fig. 6 Lipocalin conservation analysis. **(A)** Residue conservation among 248 lipocalin proteins including 105 OBPs, 29 VEGs, 46 SAL, 21 LCN9 and 47 MUPs. In each frame, the consensus logo is colored according to the side-chain chemistry (blue black and green for charged, hydrophobic and hydrophilic respectively). The height of

conserved (82%, Supplementary files 1) among OBPs and most lipocalins, suggesting that this mechanism may be a common feature of lipocalin folding (Fig. 6).

Tyrosine, with its hydroxyl group capable of acting as both a hydrogen bond donor and acceptor, plays a crucial role in controlling binding. Additionally, its aromatic ring, through hydrophobic interactions, helps occlude the calyx from the surrounding solvent, preventing hydration of the cavity. This feature facilitates the dehydration observed during ligand binding and may explain why tyrosine is highly conserved. Phenylalanine is an alternative to tyrosine in the lipocalin fold, further emphasizing the importance of a bulking residue occluding solvent entry [66]. Interestingly, substituting it with a smaller amino acid, such as alanine, abrogates binding [68]. Previous studies have demonstrated that ligand binding is primarily driven by enthalpic contributions, while the unfavorable entropic loss due to ligand restriction may be offset by the release of highly ordered water molecules within the binding cavity or an increase in the dynamics of protein side chains [69–71]. Substituting tyrosine with alanine, however, disrupts this balance, and might flood the cavity eliminating the favorable entropic gain associated with the release of water molecules [68]. In contrast, replacing the gate with a disulfide bridge enhances steric discrimination by permanently blocking access to the calyx, effectively hindering ligand recognition.

Interestingly, the opposite residue of the Y82, serine in the case of rOBP1, shows variation among the lipocalin family (Fig. 6A). Serine is predominant in most von Ebner's gland proteins (VEGs) or salivary lipocalins (SALs) which are members of lipocalins involved in chemical communication [16], while proline is common to most OBPs (Fig. 6B). However, given that only the backbone is involved in this

symbols indicates the relative frequency of each amino or nucleic acid at that position. **(B)** Hierarchical cluster tree based on the alignment of gate position (S34 in the case of rOBP1). Gates P, S, D/E indicate a proline, a serine, or either an aspartate or glutamate respectively. All other amino-acid types have been grouped under Gate Other

hydrogen-bonding gate, we would expect greater variability at this position. We hypothesize that this residue could have an impact on the dynamics of the gate, with a serine forming an additional hydrogen bond with its side chain, thereby reinforcing a more constrained conformation, and conversely, a proline favoring a more flexible conformation allowing for a broader recognition.

Discussion

While previous studies have shown the role of the gate in protein binding, here we successfully locked the OBP into a closed state. Our findings demonstrate that inserting a disulfide bridge is an effective strategy for engineering non-responsive OBPs by permanently blocking access to the calyx, hindering ligand recognition. This strategy could be applied to any lipocalin fold. Furthermore, we reveal that the gate is the sole entrance to the binding pocket, emphasizing the importance of positioning the OBP with the gate facing the bulk solution in biosensor applications.

Through MD simulations, we also identified two sites of transient stabilization for the binders near the gate region, each exhibiting different polarities (fig. S14). One of these sites, pocket 3, contains a high proportion of conserved hydrophilic amino acids (Fig. 6A), which is involved in a network that includes the gate (Fig. 5). Notably, mutation of a key residue—T112 which is poorly conserved (Fig. 6A)—to aspartic acid abolishes ligand recognition by rOBP1, highlighting the role of this distal residue in the allosteric regulation of binding. This suggests that alterations to the hydrogen bond network may influence the binding of odorants to rOBP1. Network analysis has proven effective

in identifying allosteric sites on protein [72] and highlights T112 as a critical node. Our results recover the observation that mutations at such identified distal sites can affect binding affinity, emphasizing the importance of long-range interactions in protein function [73]. This finding is further supported by Stepanenko et al., [74] who proposed the existence of an ionic network crucial for OBP dynamics. Their work underscores the importance of electrostatic interactions in maintaining the structural flexibility necessary for ligand binding. Moreover, residue T112 might be subject to phosphorylation [24]—a post-translational modification thought to regulate the binding specificity of OBPs. Phosphorylation at this site could provide a mechanism for the regulation of ligand binding.

Conversely, we identified a hydrophobic pocket (pocket 4 in Fig. 6A) where mutations do not directly influence the ligand affinity but modify the thermodynamic signature of binding (Fig. 4B, table S3). The conservation of this region and its high hydrophobic content exposed to the solvent suggest it may play a role in protein-protein interactions. Hydrophobic surface patches are well-known facilitators of such interactions [75], and OBPs are proposed to interact with olfactory receptors [76] or the endocytic receptor megalin [77]. This hydrophobic patch could thus be critical for mediating these interactions.

This patch might facilitate progressive change in the solvation shell of the ligand, affecting the thermodynamic signature [78]. This is further supported by the modeling of unbinding events (Fig. 2D). We observed that the greatest friction between the ligand and the protein coincided with its dehydration, emphasizing the importance of ligand solvation in the binding/unbinding process. These experiments provide a description of the energy along the ligand's entry trajectory, making it possible to quantify the different stages of this migration.

Conclusion

Protein binding pocket dynamics are key to determining interaction specificity and binding processes. Protein flexibility enables adaptation to different ligands and highlights the importance of considering internal motion in predicting binding properties and designing new compounds. Binding pockets are defined by surrounding amino acid residues, and flexibility allows pockets to open, close, or adjust to accommodate ligands. Protein pocket dynamics were categorized in 5 different classes: the formation or disappearance of (1) subpockets and (2) adjacent pockets, (3) pocket breathing due to molecular fluctuations, (4) the opening or closing of channels or gates, and (5) the emergence of allosteric pockets that influence other binding sites [79]. In this study, we

provide an atomic-level description of the odorant-binding mechanism to OBPs and demonstrate that it belongs to class (3) as the fluctuation of amino acids control the binding. The binding process is based on the opening of a gate formed by residues Y82 and S34. The gate remains closed thanks to a hydrogen bond between the hydroxyl group of the tyrosine and the carbonyl of the serine backbone.

We demonstrate that it is possible to design a protein that does not bind to any ligand by generating a disulfide bond through site-directed mutagenesis. The modified S34C, Y82C protein is no longer able to bind three different odorants. This confirms that the ligand accesses the binding site via a unique conserved pathway. Moreover, we also show that the binding could be controlled by distal mutations on residues whose movement correlates with this gate. A genome wide analysis show that this result could be extended to other proteins from the lipocalin family.

The remarkable stability of mammalian OBPs has been demonstrated in numerous studies, highlighting that these proteins retain their odorant-binding properties under extreme temperature conditions [32, 74, 80]. This feature is particularly crucial for using these proteins as olfactory or molecular sensors for electronic noses. Therefore, this fine description of ligand binding will open the way to rationally modulate their recognition spectrum to target specific molecules.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00018-025-05777-8>.

Acknowledgements The authors acknowledge support by the bwUniCluster computing initiative, the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC. The authors acknowledge the Ministry for Education and research (BMBF). The authors acknowledge support by the Université Côte d'Azur's Center for High-Performance Computing (OPAL infrastructure) for providing resources and support.

Author contributions Conceptualization: LB, JT. Methodology: MJ, LM, CM, SH, NJ, CB, ML, SW, LB, JT. Investigation: ML, MJ, LM, CB, CM, SW, LB, JT. Visualization: MJ, MB, CB, ML, JT. Supervision: LB, JT. Writing—original draft: ML, JT. Writing—review & editing: ML, MJ, CB, NJ, SH, CM, SF, SW, LB, JT.

Funding Fondation Roudnitska under the aegis of Fondation de France (PhD fellowship to M.L.). Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the grant WO 2410/2 –1 within the framework of the Research Unit FOR 5099 “Reducing complexity of non-equilibrium systems” (Project No. 431 945 604) for financial support (M.J. and S.W.). Deutscher Akademischer Austausch Dienst– DAAD-PROCOPE project “Molecular kinetics in chemical senses” (DAAD, grant agreement ID 57702382) (M.J. and S.W.). European Union's Horizon 2020 research and innovation programme

under grant agreement No 101004806 (MOSBRI-2024-293) (S.V.H., N.C.J., C.M. and J.T.). French government through the France 2030 investment plan managed by the National Research Agency (ANR), as part of the Initiative of Excellence Université Côte d'Azur under reference number ANR-15-IDEX-01 (S.F., M.L., J.T.).

Data availability All data are available in the main text or the supplementary materials. All MD trajectories are available upon request.

Declarations

Competing interests Authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Breustedt DA, Schönfeld DL, Skerra A (2006) Comparative ligand-binding analysis of ten human lipocalins. *Biochim Biophys Acta, Proteins Proteomics* 1764(2):161–173
2. Stopková R et al (2021) Biological roles of lipocalins in chemical communication, reproduction, and regulation of microbiota. *Front Physiol* 12:740006
3. Flower DR (1995) Multiple molecular recognition properties of the Lipocalin protein family. *J Mol Recognit* 8(3):185–195
4. Yang H-H et al (2023) Lipocalin family proteins and their diverse roles in cardiovascular disease. *Pharmacol Ther* 244:108385
5. Ganfornina MD, Gutiérrez G, Bastiani M (2000) A phylogenetic analysis of the Lipocalin protein family. *Mol Biol Evol* 17(1):114–126
6. Schlehuber S, Skerra A (2005) Lipocalins in drug discovery: from natural ligand-binding proteins to 'anticalins'. *Drug Discovery Today* 10(1):23–33
7. Flower DR, North AC, Sansom CE (2000) The lipocalin protein family: structural and sequence overview. *Biochim Biophys Acta, Protein Struct Mol Enzymol* 1482(1–2):9–24
8. Schiefner A, Skerra A (2015) The menagerie of human lipocalins: a natural protein scaffold for molecular recognition of physiological compounds. *Acc Chem Res* 48(4):976–985
9. Flower DR (1996) The Lipocalin protein family: structure and function. *Biochem J* 318(1):1–14
10. Kalikka J, Akola J (2011) Steered molecular dynamics simulations of ligand–receptor interaction in lipocalins. *Eur Biophys J* 40:181–194
11. Gonçalves F et al (2021) Biotechnological applications of mammalian odorant-binding proteins. *Crit Rev Biotechnol* 41(3):441–455
12. Kim HJ, Eichinger A, Skerra A (2009) High-affinity recognition of lanthanide (III) chelate complexes by a reprogrammed human Lipocalin 2. *J Am Chem Soc* 131(10):3565–3576
13. El Kazzy M et al (2024) Study and optimization of the selectivity of an odorant binding protein-based bioelectronic nose. *Biosens Bioelectron* 116879
14. Fukuhara A et al (2012) Drug delivery system for poorly water-soluble compounds using lipocalin-type prostaglandin D synthase. *J Controlled Release* 159(1):143–150
15. Miyamoto Y et al (2023) Structural and interaction analysis of human lipocalin-type prostaglandin D synthase with the poorly water-soluble drug NBQX. *FEBS J* 290(16):3983–3996
16. Pelosi P, Knoll W (2022) Odorant-binding proteins of mammals. *Biol Rev* 97(1):20–44
17. Bianchet MA et al (1996) The three-dimensional structure of bovine odorant binding protein and its mechanism of odor recognition. *Nat Struct Biol* 3(11):934–939
18. Briand L et al (2000) Ligand-binding properties and structural characterization of a novel rat odorant-binding protein variant. *Eur J Biochem* 267(10):3079–3089
19. Lazar J et al (2002) Molecular and functional characterization of an odorant binding protein of the Asian elephant, *Elephas maximus*: implications for the role of lipocalins in mammalian olfaction. *Biochemistry* 41(39):11786–11794
20. Bianchi F et al (2019) Vertebrate odorant binding proteins as antimicrobial humoral components of innate immunity for pathogenic microorganisms. *PLoS ONE* 14(3):e0213545
21. Grolli S et al (2006) Odorant binding protein has the biochemical properties of a scavenger for 4-hydroxy-2-nonenal in mammalian nasal mucosa. *FEBS J* 273(22):5131–5142
22. Vincent F et al (2000) Complexes of Porcine odorant binding protein with odorant molecules belonging to different chemical classes. *J Mol Biol* 300(1):127–139
23. Tegoni M et al (2000) Mammalian odorant binding proteins. *Biochim Biophys Acta, Protein Struct Mol Enzymol* 1482(1–2):229–240
24. Meillour N-L (2014) Proteomic analysis of pig (*Sus scrofa*) olfactory soluble proteome reveals O-linked-N-acetylglucosamylation of secreted odorant-binding proteins. *Front Endocrinol* 5:202
25. Brimau F et al (2010) Binding specificity of Recombinant odorant-binding protein isoforms is driven by phosphorylation. *J Chem Ecol* 36:801–813
26. Nespoulous C et al (2004) Odorant binding and conformational changes of a rat odorant-binding protein. *Chem Senses* 29(3):189–198
27. White SA et al (2009) Structure of rat odorant-binding protein OBPI at 1.6 Å resolution. *Acta Crystallogr Sect D: Biol Crystallogr* 65(5):403–410
28. Shenol A et al (2024) Molecular dynamics-based identification of binding pathways and two distinct high-affinity sites for succinate in succinate receptor 1/GPR91. *Mol Cell* 84(5):955–966 e4
29. Pacalon J et al (2023) Elucidation of the structural basis for ligand binding and translocation in conserved insect odorant receptor co-receptors. *Nat Commun* 14(1):8182
30. Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. *Neuron* 99(6):1129–1143
31. Golebiowski J et al (2012) Interaction between odorants and proteins involved in the perception of smell: the case of odorant-binding proteins probed by molecular modelling and biophysical data. *Flavour Fragr J* 27(6):445–453
32. Burova TV et al (1999) Conformational stability and binding properties of Porcine odorant binding protein. *Biochemistry* 38(45):15043–15051
33. Wolf S, Stock G (2018) Targeted molecular dynamics calculations of free energy profiles using a nonequilibrium friction correction. *J Chem Theory Comput* 14(12):6175–6182
34. Wolf S et al (2020) Multiscale ligand dissociation dynamics from atomistic simulations. *Nat Commun* 11(1):2918

35. Wolf S, Post M, Stock G (2023) Path separation of dissipation-corrected targeted molecular dynamics simulations of protein–ligand unbinding. *J Chem Phys*, 158(12)
36. Taniel V, Jager M, Wolf S (2024) Learning Protein-Ligand unbinding pathways via Single-Parameter community detection. *J Chem Theory Comput* 20(12):5058–5067
37. Wallace BA et al (2011) Synchrotron radiation circular dichroism (SRCD) spectroscopy: an emerging method in structural biology for examining protein conformations and protein interactions. *Nuclear Instruments and methods in physics research section A: accelerators, spectrometers, detectors and associated equipment* 649(1):177–178
38. Freyer MW, Lewis EA (2008) Isothermal Titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol* 84:79–113
39. Case DA, Ben-Shalom KBIY, Brozell SR, Cerutti DS, Cheatham TE III, Cruzeiro VWD, Darden TA, Duke RE, Giambasu G, Gilson MK, Gohlke H, Goetz AW, Harris R, Izadi S, Izmailov SA, Kasavajhala K, Kovalenko A, Krasny R, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Man V, Merz KM, Miao Y, Mikhailovskii O, Monard G, Nguyen H, Onufriev A, Pan F, Pantano S, Qi R, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Skrynnikov NR, Smith J, Swails J, Walker RC, Wang J, Wilson L, Wolf RM, Wu X, Xiong Y, Xue Y, D.M. York and P.A., Kollman (2020) *AMBER*. University of California, San Francisco, 2020.
40. Anandakrishnan R, Aguilar B, Onufriev AV (2012) H++3.0: automating p K prediction and the Preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* 40(W1):W537–W541
41. Frisch MJ, Schlegel GWTHB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr., Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, Fox DJ (2009) *Gaussian 09*. Gaussian, Inc., Wallingford CT
42. Ester M et al (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. in *kdd*
43. Abraham MJ et al (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1:19–25
44. da Sousa AW, Vranken WF (2012) ACPYPE-Antechamber python parser interface. *BMC Res Notes* 5:1–8
45. Ernst M, Sittel F, Stock G (2015) Contact-and distance-based principal component analysis of protein dynamics. *J Chem Phys* 143(24)
46. Traag VA, Waltman L, van Eck NJ (2019) From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 9:1–12. <https://doi.org/10.1038/541598-019-41695-z>
47. Diez G, Nagel D, Stock G (2022) Correlation-based feature selection to identify functional dynamics in proteins. *J Chem Theory Comput* 18(8):5079–5088
48. Felline A, Seeber M, Fanelli F (2022) PSNtools for standalone and web-based structure network analyses of conformational ensembles. *Comput Struct Biotechnol J* 20:640–649
49. Waterhouse AM et al (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191
50. Crooks GE et al (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188–1190
51. Letunic I, Bork P (2024) Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* gkae268
52. Moitrier L et al (2022) Ligand binding properties of Odorant-Binding protein OBP5 from *Mus musculus*. *Biology* 12(1):2
53. Barou E et al (2015) Electrochemical detection of the 2-isobutyl-3-methoxypyrazine model odorant based on odorant-binding proteins: the proof of concept. *Bioelectrochemistry* 101:28–34
54. Miles AJ et al (2007) Synchrotron radiation circular dichroism (SRCD) spectroscopy: new beamlines and new applications in biology. *Spectroscopy* 21(5–6):245–255
55. Hoffmann SV, Fano M, van de Weert M (2016) Circular dichroism spectroscopy for structural characterization of proteins. *Anal Techniques Pharm Sci* 223–251
56. Anthis NJ, Clore GM (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 Nm. *Protein Sci* 22(6):851–858
57. Micsonai A et al (2018) BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res* 46(W1):W315–W322
58. Paesani M et al. Odorant binding proteins facilitate the gas-phase uptake of odorants through the nasal mucus. *Chemistry–A Eur J* e202403058
59. Matsumura M, Matthews BW (1989) Control of enzyme activity by an engineered disulfide bond. *Science* 243(4892):792–794
60. Dombkowski AA, Sultana KZ, Craig DB (2014) Protein disulfide engineering. *FEBS Lett* 588(2):206–212
61. Shimaoka M et al (2001) Reversibly locking a protein fold in an active conformation with a disulfide bond: integrin α L I domains with high affinity and antagonist activity in vivo. *Proc Natl Acad Sci* 98(11):6009–6014
62. Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* 1(6):2876–2890
63. Matsuo K et al (2009) Comprehensive secondary-structure analysis of disulfide variants of lysozyme by synchrotron-radiation vacuum-ultraviolet circular dichroism. *Proteins Struct Funct Bioinform* 77(1):191–201
64. Greenfield NJ (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat Protoc* 1(6):2527–2535
65. Hoffmann W et al (2019) An intrinsic hydrophobicity scale for amino acids and its application to fluorinated compounds. *Angew Chem Int Ed* 58(24):8216–8220
66. Vincent F et al (2004) Crystal structures of bovine odorant-binding protein in complex with odorant molecules. *Eur J Biochem* 271(19):3832–3842
67. Hajjar E et al (2006) Odorant binding and conformational dynamics in the odorant-binding protein. *J Biol Chem* 281(40):29929–29937
68. Meillour N-L et al (2009) Phenylalanine 35 and tyrosine 82 are involved in the uptake and release of ligand by porcine odorant-binding protein. *Biochim Biophys Acta, Proteins Proteomics* 1794(8):1142–1150
69. Barratt E et al (2005) Van der Waals interactions dominate ligand–protein association in a protein binding site occluded from solvent water. *J Am Chem Soc* 127(33):11827–11834
70. Portman KL et al (2014) Enthalpy/entropy compensation effects from cavity desolvation underpin broad ligand binding selectivity for rat odorant binding protein 3. *Biochemistry* 53(14):2371–2379
71. Gómez-Velasco H, Rojo-Domínguez A, García-Hernández E (2020) Enthalpically-driven ligand recognition and

- cavity solvation of bovine odorant binding protein. *Biophys Chem* 257:106315
72. Wu N, Barahona M, Yaliraki SN (2024) Allosteric communication and signal transduction in proteins. *Curr Opin Struct Biol* 84:102737
73. Kazan IC et al (2022) Design of novel cyanovirin-N variants by modulation of binding dynamics through distal mutations. *Elife* 11:e67474
74. Stepanenko OV et al (2008) Hydrophobic interactions and ionic networks play an important role in thermal stability and denaturation mechanism of the Porcine odorant-binding protein. *Proteins Struct Funct Bioinform* 71(1):p35–44
75. Rego N.B., Xi E, Patel AJ (2021) Identifying hydrophobic protein patches to inform protein interaction interfaces. *Proc Natl Acad Sci* 118(6):e2018234118
76. Vidic J et al (2008) On a chip demonstration of a functional role for odorant binding protein in the preservation of olfactory receptor activity at high odorant concentration. *Lab Chip* 8(5):678–688
77. Strotmann J, Breer H (2011) Internalization of odorant-binding proteins into the mouse olfactory epithelium. *Histochem Cell Biol* 136:357–369
78. Peccati F, Jiménez-Osés G (2021) Enthalpy–entropy compensation in biomolecular recognition: A computational perspective. *ACS Omega* 6(17):11122–11130
79. Stank A et al (2016) Protein binding pocket dynamics. *Acc Chem Res* 49(5):809–815
80. Paolini S et al (1999) Porcine odorant-binding protein: structural stability and ligand affinities measured by Fourier-transform infrared spectroscopy and fluorescence spectroscopy. *Biochim Biophys Acta, Protein Struct Mol Enzymol* 1431(1):179–188

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Nature invented olfaction twice: the intriguing structure of insect olfaction



Reste l'odeur. Dans le domaine de nos sens, des émanations odorantes, mieux que tout autre chose, expliqueraient à peu près les papillons accourus, et ne trouvant néanmoins qu'après certaines hésitations l'appât qui les attire. Y aurait-il, en effet, des effluves analogues à ce que nous appelons odeur, effluves de subtilité extrême, absolument insensibles pour nous, et néanmoins capables d'impressionner un odorat mieux doué que le nôtre ?⁵

The scientific exploration of insect olfaction can be traced back to the pioneering observations of Jean-Henri Fabre in the late nineteenth century, who used meticulous naturalistic experiments to demonstrate that insects rely heavily on chemical cues to locate food, mates, and hosts. Fabre's behavioral studies provided the first compelling evidence that smell, rather than vision alone, governs many insect actions. Throughout the twentieth century, this behavioral foundation was progressively complemented by physiological and electrophysiological approaches, leading to the identification of olfactory sensilla on antennae and maxillary palps, and the characterization of odor-evoked neuronal responses. The molecular era, initiated in the 1990s with the cloning of insect odorant receptors, transformed the field by revealing a large and diverse receptor repertoire and a unique ionotropic mode of signal transduction distinct from vertebrates. Today, insect olfaction is investigated through an integrated framework combining genetics, structural biology, neurophysiology, and computational modeling, enabling a mechanistic understanding of how chemical signals are encoded, processed, and ultimately translated into behavior.

Role of olfaction in insects

The main difference between intercellular communication and animal communication lies in the distance separating the emitter from the receiver. Consequently, the signalling molecules perceived by an insect are weakly hydrophilic and, above all, volatile (e.g., DEET, PMD) so that they can be transported through the air¹. They therefore have relatively simple molecular structures. For the message to be precise and intelligible, olfactory reception must be both diversified, so as to recognize a wide range of molecule, and highly specific, in order to recognize only the molecule(s) actually present.

The [Figure 10](#) illustrates the organization of chemosensory pathways in the insect brain and highlights the segregation and integration of olfactory and gustatory information. Odorant signals detected by sensory neurons in the antennae and maxillary palps converge onto the antennal lobe, the primary olfactory processing center, before being relayed by projection neurons to higher-order brain regions such as the mushroom body and the lateral protocerebrum. These structures are key sites for sensory integration, learning, and memory. In parallel, gustatory inputs follow partially distinct neural routes, underscoring the functional specialization of chemosensory circuits while emphasizing their coordinated contribution to odor-guided behaviors².

⁵ Souvenirs entomologiques, Jean-Henri FABRE, 1900, VIIème Série, Chapitre 23.

Insect Olfaction

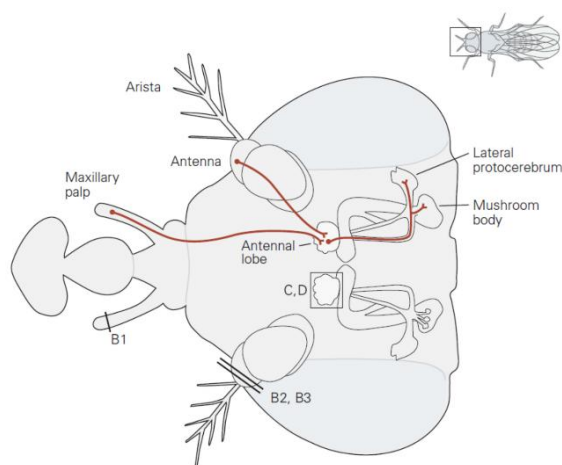


Figure 10. Neural pathways of olfactory and gustatory processing in the insect brain. Sensory inputs from the antennae (arista) and maxillary palps project to the antennal lobe, where olfactory information is first processed, while gustatory inputs (B1–B3) follow distinct pathways. Projection neurons relay olfactory signals from the antennal lobe to higher brain centers, including the mushroom body and the lateral protocerebrum, which are involved in learning, memory, and integration of sensory information.

The diversity of recognized molecules is made possible by a repertoire of olfactory receptors (Ors) expressed in the insect's olfactory sensory organs. The number of ORs varies widely among insects, from 4 in the damselfly

to over 350 in some ants, due to rapid diversification during insect evolution and adaptation to different ecological niches³. These receptors have a high discriminative power for small molecular structures that may differ by only a few carbon atoms, thanks to the following principle: a single odorant molecule is recognized by several receptors, and a single receptor can recognize several odorant molecules. Recognition thus involves a process based on relative affinities with multiple receptors, and the resulting information is combinatorial in nature. An odor therefore corresponds to an olfactory image, defined as the combination of the responses of several receptors to the same molecule⁴. This explains the ability of insects to recognize more odors than the number of olfactory receptors encoded in their genome.

Structure of Olfactory Receptors in Insect

Insect ORs have evolved to recognize an extremely large panel of ligands, endowing each insect species with the ability to selectively detect volatile chemicals associated with its specialized habitat and lifestyle⁵. This extraordinary diversity of insect ORs contrasts with a conserved structural organization and common activation mechanisms. Our knowledge of insect ORs at the molecular level is recent and surprising. It has been discovered that these receptors have an inverted topology compared to the vertebrate olfactory receptors⁶ and that they belong to the family of ligand-activated ion channels (LGICs)⁷.

Insect olfactory signal transduction relies on a molecular architecture that is distinct from that of vertebrates. Instead of G protein-coupled receptors, insects express odorant receptors (ORs) that form heteromeric complexes composed of an odorant-specific subunit (OrX) and a highly conserved co-receptor (Orco). Functional insect ORs assemble as tetrameric complexes, in which Orco is an essential structural and functional component, required for receptor trafficking, membrane (Figure 11).

Major breakthroughs in structural biology have recently enabled the resolution of three-dimensional structures of insect ORs, providing a molecular framework for understanding their function. The first milestone was the determination of the tetrameric structure of the olfactory receptor co-receptor (Orco), revealing its architecture as a ligand-gated ion channel⁸. This was followed by the structure of an odorant-specific receptor, *MhOR5*, solved in complex with odorant ligands, which offered the first direct view of how odorants are accommodated within an insect OR binding pocket⁹. Subsequently, several structures of Orco–OR heteromeric complexes have been reported, revealing distinct stoichiometries and assembly modes¹⁰. Together, these studies establish a structural basis for receptor diversity, ligand recognition, and channel gating in insect olfaction.

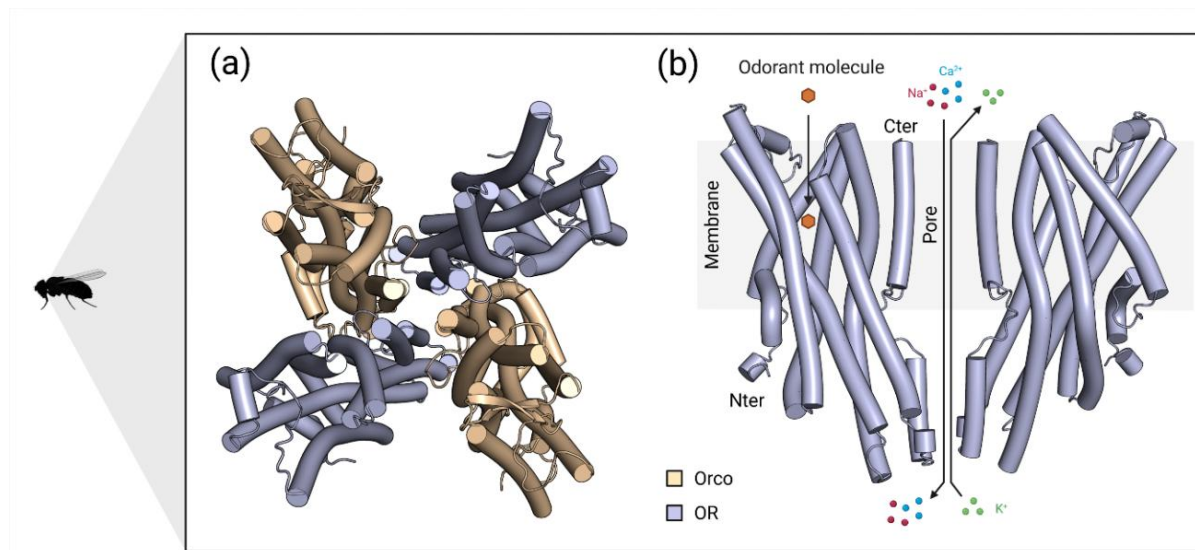


Figure 11. Structural organization of the insect odorant receptor. (a) Top view of a modeled *DmelOrco*-*DmelOR5* ion channel complex (Orco in flesh color and OR5 in blue), with a 2:2 subunit stoichiometry. Note that a 3:1 stoichiometry appears to be more probable. (b) Side view of the channel, centered on the OR subunits embedded within the membrane (gray). Insect OR and Orco subunits adopt a common fold composed of seven transmembrane helices, with the N-terminus facing the cytosol and the C-terminus oriented extracellularly. Binding of an odorant to the OR subunits allosterically triggers opening of the channel pore, permitting the flux of Na⁺, Ca²⁺, and K⁺ ions.

Insect Olfactory Signal Transduction

These tetrameric OrX/Orco complexes are expressed in the dendritic membranes of olfactory sensory neurons within antennal sensilla. Odorant binding to the OrX subunit directly activates the receptor complex, which functions primarily as a ligand-gated, non-selective cation channel⁷. Channel opening induces a rapid influx of cations, including Ca²⁺, leading to depolarization of the sensory neuron and initiation of the olfactory signal (Figure 12). This ionotropic mechanism accounts for the fast kinetics and high temporal resolution characteristic of insect olfactory responses .

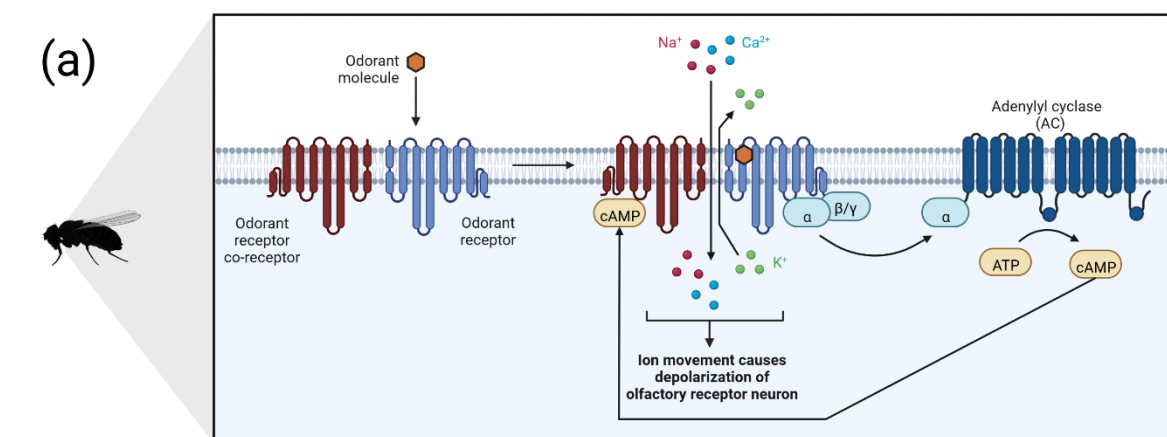


Figure 12. Ionotropic transduction mechanism in insect olfaction. An odorant activate the heteromeric ion channel by binding to the OR part. The flux of ions causes a depolarization of the neuron. Recent work highlights the activation of the G protein by the OR, activating adenylyl cyclase and producing the second messenger cAMP. The cAMP then binds to the odorant co-receptor (Orco), increasing the activation of the channel further.

In addition to this direct ion channel activity, several studies indicate that second-messenger signaling pathways, notably those involving cyclic AMP (cAMP), contribute to olfactory transduction and modulation. Odorant stimulation has been associated with increases in intracellular cAMP levels, suggesting the involvement of adenylate cyclase-dependent pathways. cAMP signaling may modulate channel gating,

neuronal excitability, or adaptation processes, either directly through Orco or indirectly via Ca²⁺-dependent mechanisms. These observations support a model in which insect ORs integrate ionotropic signaling with metabotropic modulation, rather than relying exclusively on one mode of transduction.

The conserved Orco subunit appears central to this dual functionality. Beyond its role in tetramer assembly and channel formation, Orco has been implicated in sensitivity regulation and signal amplification, acting as a molecular hub that couples odorant detection to intracellular signaling dynamics. This organization enables insects to combine rapid odor detection with adaptive modulation of sensory input.

Taken together, insect olfactory transduction can be described as a hybrid signaling system in which tetrameric OrX/Orco complexes operate as ligand-gated ion channels, supplemented by cAMP-dependent modulatory pathways. This distinctive mechanism underlies the efficiency, flexibility, and robustness of olfactory perception in insects and highlights fundamental evolutionary differences between insect and vertebrate chemosensory systems.

References

3.3 – INSECT OLFACTORY RECEPTOR

- J. Pacalon, G. Audic, J. Magnat, M. Philip, J. Golebiowski, C. J. Moreau*, J. Topin*. Elucidation of the structural basis for ligand binding and translocation in conserved insect odorant receptor co-receptors, *Nature Communications*, **2023**.

In this article, for which I am the corresponding author (together with my collaborator C. Moreau from IBS, Grenoble), we highlighted the molecular mechanisms underlying the migration of one of the few known artificial ligands of an insect olfactory co-receptor. This work was communicated in the INSB newsletter⁶.

Uncategorized References

1. Conchou, L.; Lucas, P.; Meslin, C.; Proffit, M.; Staudt, M.; Renou, M., Insect odorscapes: from plant volatiles to natural olfactory scenes. *Frontiers in physiology* **2019**, *10*, 972.
2. Tanaka, N. K.; Awasaki, T.; Shimada, T.; Ito, K., Integration of chemosensory pathways in the *Drosophila* second-order olfactory centers. *Current biology* **2004**, *14* (6), 449-457.
3. Robertson, H. M., Molecular evolution of the major arthropod chemoreceptor gene families. *Annual review of entomology* **2019**, *64* (1), 227-242.
4. Wicher, D.; Miazzi, F., Functional properties of insect olfactory receptors: ionotropic receptors and odorant receptors. *Cell and tissue research* **2021**, *383* (1), 7-19.
5. Andersson, M. N.; Löfstedt, C.; Newcomb, R. D., Insect olfaction and the evolution of receptor tuning. *Frontiers in Ecology and Evolution* **2015**, *3*, 53.
6. Benton, R.; Sachse, S.; Michnick, S. W.; Vosshall, L. B., Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. *PLoS biology* **2006**, *4* (2), e20.
7. Sato, K.; Pellegrino, M.; Nakagawa, T.; Nakagawa, T.; Vosshall, L. B.; Touhara, K., Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* **2008**, *452* (7190), 1002-1006.
8. Butterwick, J. A.; Del Marmol, J.; Kim, K. H.; Kahlson, M. A.; Rogow, J. A.; Walz, T.; Ruta, V., Cryo-EM structure of the insect olfactory receptor Orco. *Nature* **2018**, *560* (7719), 447-452.
9. Del Marmol, J.; Yedlin, M. A.; Ruta, V., The structural basis of odorant recognition in insect olfactory receptors. *Nature* **2021**, *597* (7874), 126-131.
10. Zhao, J.; Chen, A. Q.; Ryu, J.; Del Marmol, J., Structural basis of odor sensing by insect heteromeric odorant receptors. *Science* **2024**, *384* (6703), 1460-1467.

⁶ <https://www.insb.cnrs.fr/fr/cnrsinfo/olfaction-des-insectes-speleologie-moleculaire-dans-un-co-recepteur>



Elucidation of the structural basis for ligand binding and translocation in conserved insect odorant receptor co-receptors

Jody Pacalon^{1,4}, Guillaume Audic^{2,4}, Justine Magnat², Manon Philip², Jérôme Golebiowski³, Christophe J. Moreau²✉ & Jérémie Topin¹✉

¹Université Côte d'Azur, Institut de Chimie de Nice UMR7272, CNRS, Nice, France.

²Université Grenoble Alpes, CNRS, CEA, IBS, Grenoble, France.

³Department of Brain & Cognitive Sciences, DGIST, 333, Techno JungAng, Daero, HyeonPoong Myeon, Daegu, Republic of Korea.

⁴These authors contributed equally: Jody Pacalon, Guillaume Audic.

✉ e-mail: christophe.moreau@ibs.fr; jeremie.topin@univ-cotedazur.fr

Received: 16 June 2023

Accepted: 28 November 2023

Published online: 11 December 2023

<https://doi.org/10.1038/s41467-023-44058-5>

Abstract. In numerous insects, the olfactory receptor family forms a unique class of heteromeric cation channels. Recent progress in resolving the odorant receptor structures offers unprecedented opportunities for deciphering their molecular mechanisms of ligand recognition. 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. By combining molecular modeling with electrophysiological recordings, we identified amino acids involved in the dynamic entry pathway and the binding of VUAA1 to *Drosophila melanogaster's* odorant receptor co-receptor (Orco). Our results provide evidence for the exact location of the agonist binding site and a detailed and original mechanism of ligand translocation controlled by a net-work of conserved residues. These findings would explain the particularly high selectivity of Orcos for their ligands.

Among all living multicellular organisms, insects represent more than half of all identified species on Earth, thus forming the most diverse group of animals¹. 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 adaptation², 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 for pest control^{3,4}. 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 1c summarizes the position of different residues that showed a functional impact when mutated^{5,23–26}. These studies have highlighted the central role of residues from helix S7 in forming a hydrophobic gate that contributes to cation selectivity. Moreover, the structures of MhOr5 in complex with two agonists, eugenol and DEET, revealed the ligand binding cavity of this receptor (Fig. 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. In this work, in order to identify the binding pocket and the translocation pathway of VUAA1 from the extracellular space to the Orco binding site, we combine molecular modeling approaches with site-directed mutagenesis and functional

characterization by the TEVC technique.

Groundbreaking 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 receptors⁸. 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 OR⁹, which defines the neuron's response spectrum¹⁰, even if non-canonical co-expressions have been observed in a mosquito¹¹. 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 (Fig. 1a), which have a different recognition spectrum than ORs^{12,13}.

Orcos seem 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 diverged to recognize specific ligands^{14,15}. 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^{16–22}. On the other hand, the “ancestral” OR5 receptor from *Machilis hrabei* (MhOr5) is activated by a large set of odorants, but not by VUAA1⁶. DmelOrco and MhOr5 share 18.3% sequence identity and adopt the same tertiary fold (Fig. 1b). However, the origin of the differences in the recognition spectra of the two receptors is still not fully understood.

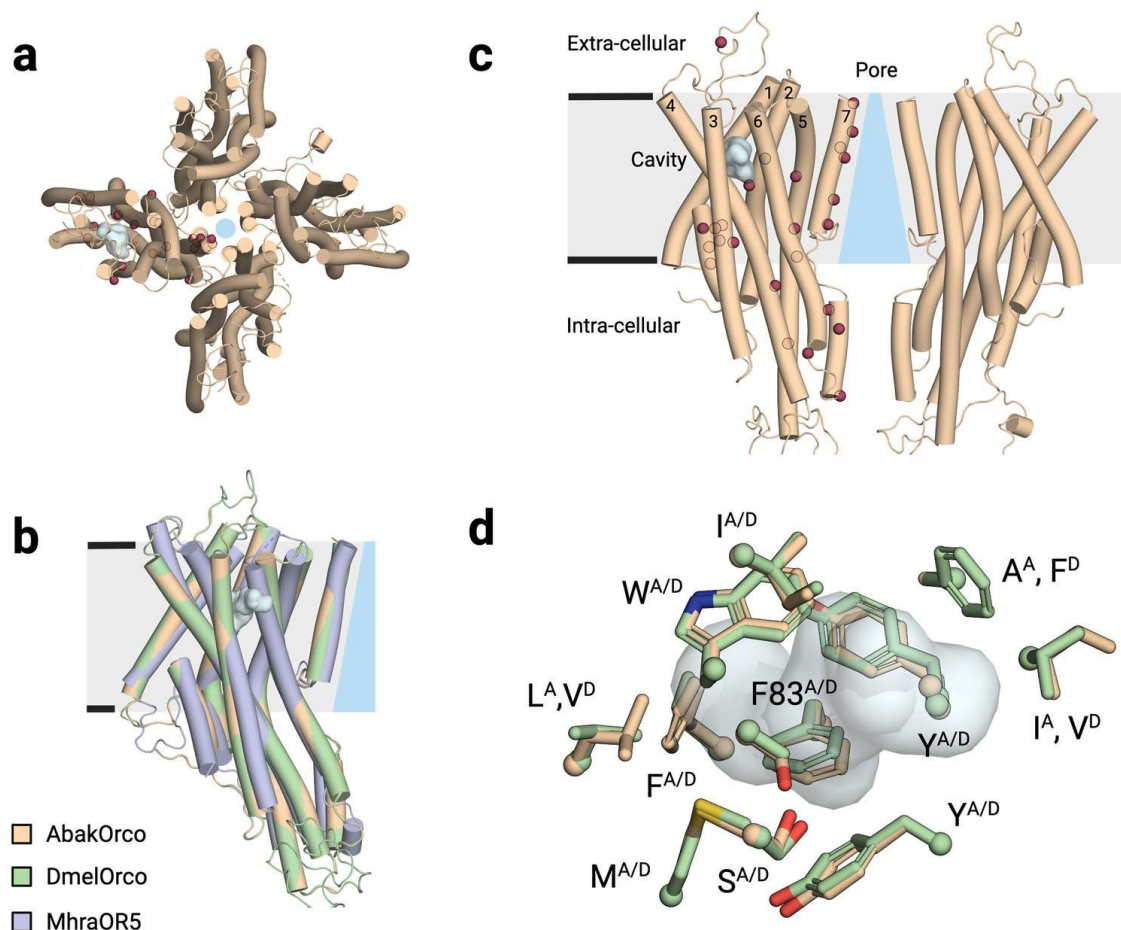


Fig. 1 | Architecture of the DmelOrco homotetramer model and comparative representations of cavities and mutants. a Extracellular view of the cryo-electronic microscopy structure of the homotetramer of *Apocrypta bakeri* Orco (AbakOrco) (pdb: 6C70). A ligand-binding pocket (in cyan) of related Orco receptors shown on the left subunit. The central pore is symbolized by a blue circle. b AbakOrco (pdb: 6C70, beige) membrane view, superposed on *Machilis hrabei* odorant receptor 5 (MhraOR5, 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. c Side view of two Orco subunits with a diagram 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*. 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, respective.

Results

Determination of the optimal Orco

Olfactory receptors are notorious for weakly expressing in heterologous systems, which impedes their functional characterization. Before initiating molecular dynamics (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 (Supplementary Fig. 1) 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.

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 template⁵. The two protein sequences are highly similar (76% of sequence identity) prefiguring a high confidence in the accuracy of the model of DmelOrco²⁷. The full protocol is detailed in the Methods section. After the release of AlphaFold2, we compared our model to the one extracted from the Alpha Fold Protein database²⁸. 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 2 (IL2) (Supplementary Fig. 2). This loop is not resolved on the cryoEM

structure of DmelOrco, which suggests a high flexibility. The structure of AbakOrco⁶ and the homology model of DmelOrco, revealed a cavity between helices S1 to S4 and S6 that could play the role of the ligand binding site for VUAA1 and its analogues (Fig. 1). Interestingly, this cavity has a position similar to the ligand binding site found in the structure of MhraOR5⁶ (Fig. 1b). The amino acids lining the two cavities are highly conserved between DmelOrco and AbakOrco with 73% identity (Fig. 1d). Notably, the cradle of this pocket would be formed by the residue F83^{Dmel}, which is critical for activation by VUAA1²⁶. 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 highlight a stepwise mechanism of VUAA1 entry to the embedded binding cavity

To reach the deeply embedded binding site, residing in the core of the transmembrane helices, the ligand must transit through a path that is assumed to be dynamic since it is closed in the structures of AbakOrco and MhraOR5. Consequently, while traditional docking approach was successfully used on human olfactory receptor²⁹, it did not allow to observe ligand entry in DmelOrco. To identify this path, multiple MD simulations were performed with several ligands to enhance the sampling of rare events such as ligand migration^{30,31} and protein conformational changes^{32,33}.

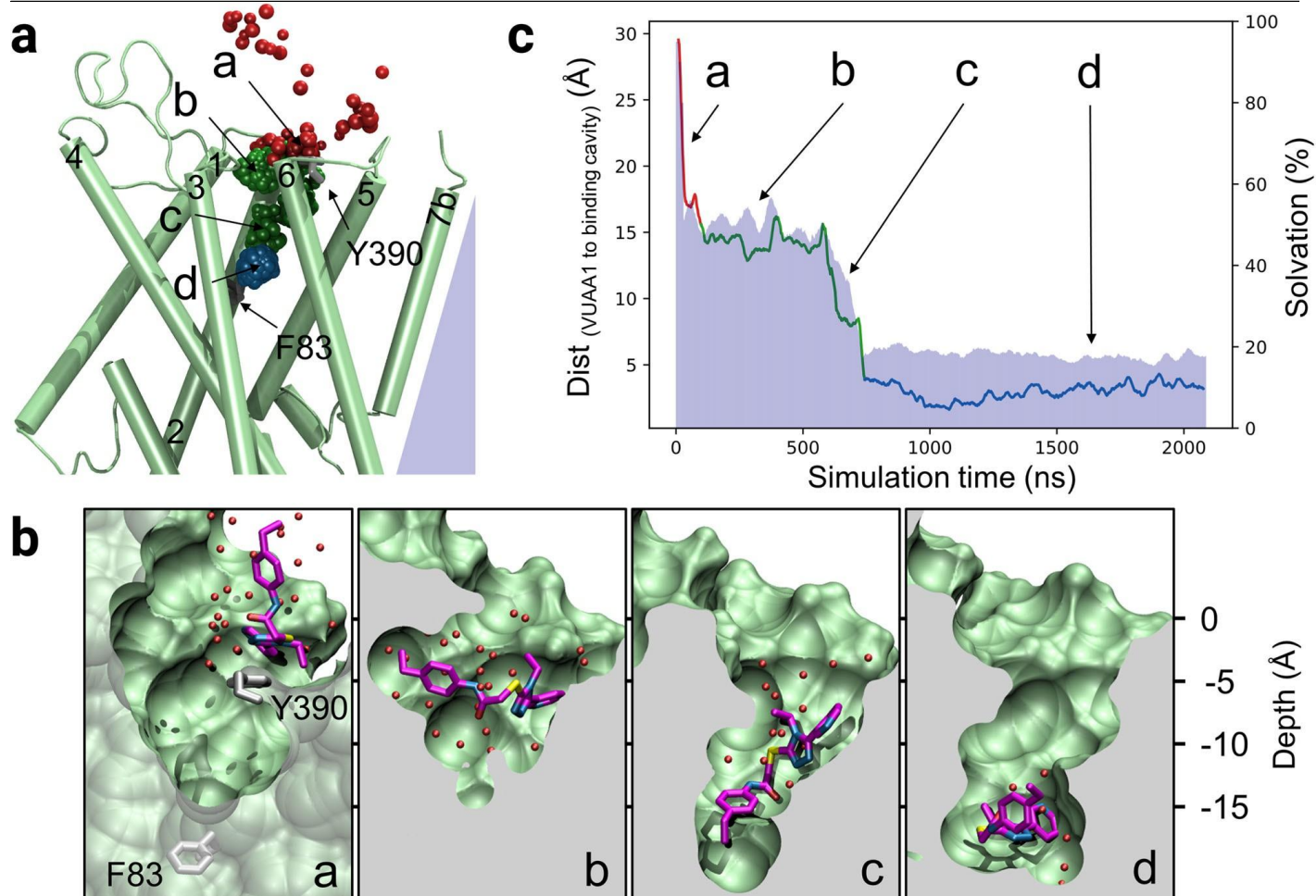


Fig. 2 | Stepwise translocation of VUAA1 into DmelOrco model in MD simulations. a Prototypical trajectory of VUAA1 binding event. The Orco monomer is shown in green. F83 and Y390 labels give their localization. The VUAA1 center of mass is represented by beads colored from red to blue according to the simulation time. a–d are sequential positions of VUAA1 progressing toward the binding site, from contact to Orco, entry, vestibule and binding to the internal cavity, respectively. b Close view of the progression of VUAA1 inside the Orco monomer corresponding to the positions a–d. Carbon atoms of VUAA1 are colored purple and the sulfur atom in yellow. Carbon atoms of F83S2 and Y390S6 are in grey and water molecules found less than 3 Å away from VUAA1 are represented by red spheres. c 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 MharaOR5, pdb: 7LID). The red curve represents the positions outside of the receptor to the contact a. The green part of the curve represents the entry and vestibule events b, c and the blue one the sample of the binding cavity d. The blue area shows the percentage of ligand solvation during the binding process (normalized to the solvation of the ligand outside the protein).

We constructed a system containing 4 DmelOrco monomers 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 for a total simulation time of ~31 μ s (see Methods). The migration of VUAA1 through the protein core was evaluated by the evolution of the distance between its center of mass and the center of mass of the binding cavity (defined as the center of mass of the eugenol molecule in MharaOR5, pdb: 7LID). We thus identified 4 distinct steps: contact (Fig. 2a, area a), entry (Fig. 2a, area b), vestibule (Fig. 2a, area c) and binding site (Fig. 2a, area d). 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 vestibule. Out of these 19 trajectories, 7 full entries into the binding pocket were observed. The other trajectories resulted in a partial binding event, where VUAA1 remains within a vestibular site, half-way to the pocket cradle (Supplementary Fig. 3). 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 (Supplementary Dataset S1) in line with the similar action of VUAA1 observed in the majority of insect Orcos²⁶. 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 (Fig. 2b, c). The first step (a) is a rapid contact (few ns) of VUAA1 with the extracellular side of DmelOrco and a rapid partial desolvation. The second step (b) is a stabilization of the position of VUAA1 during ~500 ns and a solvation stable at ~50%. The third step (c) is a rapid progress (less than 200 ns) of the molecule toward the cavity and a decrease of solvation up to ~20%. The fourth and last step (d) is a position of the molecule in the cavity with a stable solvation of around 20%. In steps (a) and (c), the desolvation of VUAA1 significantly increases, playing an essential role in the progression of the molecule toward the binding site.

The hydrophobicity and electrostatic complementarities of VUAA1 with DmelOrco in the different areas (a to d) have been evaluated (Supplementary Table 1). Analogues of VUAA1 (VUAA2, VUAA3, and VUAA4, which display greater potency, and VUAA0.5, which is less potent than VUAA1) were also incorporated and ranked by their EC_{50} ¹⁸. For all ligands considered, we noticed an increase in hydrophobic and electrostatic complementarity when the ligand was located deeper in the protein. Furthermore, the trend in hydrophobic complementarity approached that of ligand strength. Although the differences in ligand EC_{50} were minor, these observations suggest a correlation between hydrophobic complementarity in the ligand translocation pathway and ligand strength of VUAA1 analogues.

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 (Supplementary Dataset S2). The initial binding event occurs through contact between VUAA1 and Y390^{S6} at the extracellular end of S6 (Fig. 2b, area a). Starting from this position, VUAA1 makes regular contacts with the residue side chains (Fig. 2b, area b) and undergoes a large desolvation process upon its entry into the receptor bundle (Fig. 2b, area 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 (Fig. 2b, area c). The ligand finally enters the cavity (Fig. 2b, area d) that was previously identified in the structures of AbakOrco and MharaOR5, and in the model of DmelOrco (Fig. 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 MharaOR5 structure (Supplementary Fig. 4).

During its progression toward the binding site from the area b to d (Fig. 2b), VUAA1 is mostly orthogonal to the membrane plane (area c). In addition to the desolvation process, the flexibility of the molecule appears to greatly facilitate its migration. Thus, 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. Because of its general shape, the cavity could only accommodate VUAA1 in two directions, one of which is the opposite of the one observed in the simulations (Supplementary Fig. 5). Therefore, we manually flipped VUAA1 into the cavity. VUAA1 analogues identified in a previous structure-activity relationship study¹⁸ were also tested and considered in the two poses (MD and reverted). Scores of both electrostatic and hydrophobic matches for the “reverted” pose were inferior for all ligands to those of the initial MD pose (Supplementary Table 1, locations (d)),

suggesting that the initial orientation from MD simulations is preferred.

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

Site directed mutagenesis and electrophysiological characterization support 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 were 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 serine (small hydrophilic residue) or in phenylalanine (aromatic residue without hydroxyl group), 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.

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 (Fig. 3). 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).

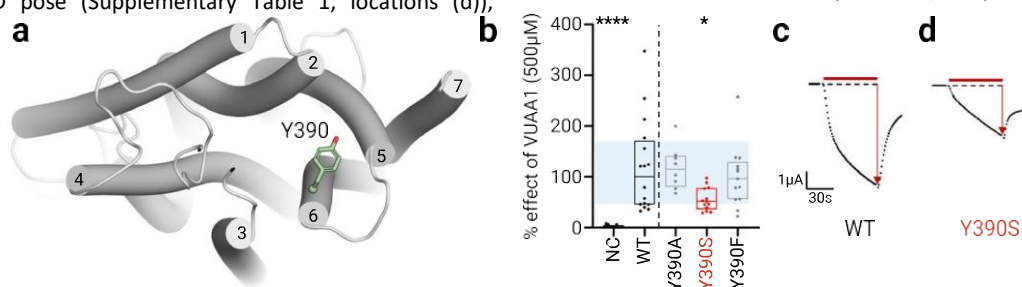


Fig. 3 | Electrophysiological characterization of Orco mutants in the contact region. **a** Top view of the entry highlighting the position of Y390. **b** Boxplot showing the current induced by 500 μ M of the ligand VUAA1 and measured by TEVC recordings on DmelOrco WT and mutants expressed in *Xenopus* oocytes. The lower bound of the box plot represents the first quartile, while the upper bound represents the third quartile. The centre corresponds to the median. The blue area is a projection of the first and third quartiles of the WT box plot. $n = 13$ recordings from different oocytes for NC; $n = 16$ (WT); $n = 8$ (Y390A); $n = 13$ (Y390S and Y390F); NC: Negative Control (non-injected oocytes). Data are analysed with one-way ANOVA with α -error = 0.05 followed by Dunn's post-hoc test, with WT used as reference for the multiple comparison test. **** $p < 0.0001$, * $p = 0.0187$ (Y390S). Results from mutant showing a statistical decrease from WT are coloured in red. **c, d** Representative current measured on WT and mutant with statistical difference from DmelOrco WT.

A Western-blot has been performed to verify that the expression level of the Y390S mutant was similar to the WT (Supplementary Fig. 6), and the semi-quantitative analysis indicated no significant differences between both constructs. This result supported the role of Y390 in VUAA1 activation. In particular, the differences observed between the mutants emphasize the impact of the hydrophilic character of position 390^{S6}. The introduction of a serine in place of tyrosine generates a hydrophilic environment³⁴ that would hamper the first step of desolvation process that is crucial for the entry of VUAA1, as observed in the simulations (Fig. 2c).

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 (Fig. 4a) at frequencies of 0.44, 0.56, 0.68, 0.17, 0.11, 0.36, and 0.70, respectively (averaged over all entry trajectories). Mutations into alanine of all seven residues did not significantly affect the amplitude of activation induced by VUAA1 (Fig. 4b),

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 tryptophan significantly reduced or abolished the activation by VUAA1 (Fig. 4b–i, red dots). Western-blot results (Supplementary Fig. 6) showed a decrease in the expression of T80W and V206W. 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 supporting that the ability of VUAA1 to access this region is critical for the channel response.

As these residues are pointing into the ligand pathway observed during simulations, these functional results support their implication in the entry of VUAA1. Interestingly, inserting the

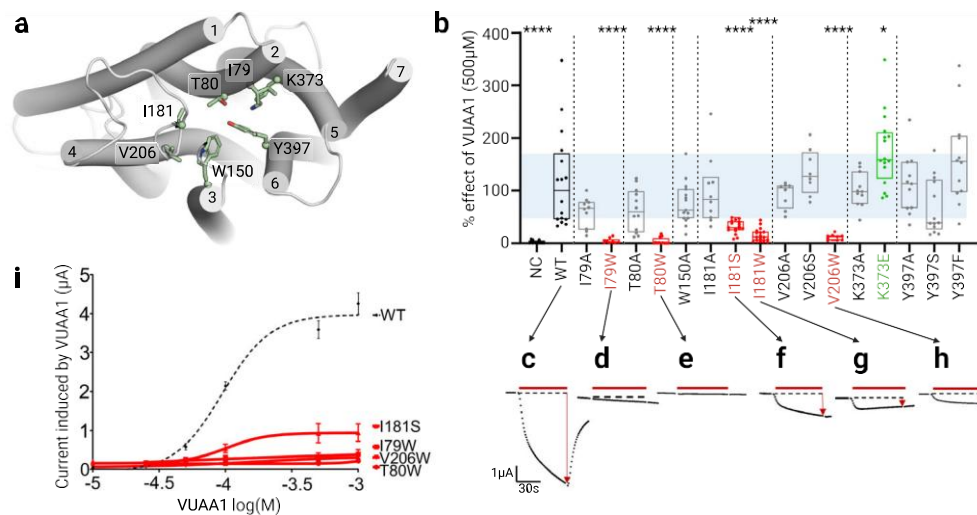


Fig. 4 | Electrophysiological characterization of Orco mutants in the entry and vestibule region. **a** Close view of the vestibule region highlighting the position of the different mutation sites. **b** Boxplot showing the current induced by 500 μM of the ligand VUAA1 and measured by TEVC recordings on DmelOrco WT and mutants expressed in *Xenopus* oocytes. The bounds of the box plot represent the first quartile (lower) and the third quartile (upper), while the centre is the median. The blue area is a projection of the first and third quartiles of the WT box plot. $n = 13$ (NC and Y397S), $n = 16$ (WT; K373E), $n = 10$ (I79A and I181A), $n = 9$ (I79W and V206W), $n = 12$ (T80A, T80W, K373A and Y397F), $n = 15$ (W150A), $n = 17$ (I181S and I181W), $n = 8$ (V206A and V206S), $n = 11$ (Y397A) 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, with WT used as reference for multiple comparisons test. **** $p < 0.0001$, * $p = 0.0269$ (K373E). Results of mutants with statistically significant decrease or increase compared to 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. EC_{50} are 94.5 μM for WT, 101.9 μM for I181S, not determined for the others. Error bars correspond to SEM from $n = 20$ (WT), $n = 8$ (T80W, I181S and V206W), $n = 10$ (I79W) recordings from different oocytes.

hydrophilic and shorter serine residue in place of the hydrophobic I181^{EL2} (I181S), significantly reduced the amplitude of activation (1.40 vs 4.71 μA for the WT) (Fig. 4b) 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 (Fig. 4b). 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 (Fig. 4b). Inversion of the charge by mutation of K373^{S5} into glutamate (K373E) did not abolish the response but increased it (7.47 μA), potentially by decreasing the polarity of the binding cavity (Supplementary Fig. 7). Western-blot results confirmed that the K373E was not overexpressed. All mutations made at position Y397^{S6} did not significantly change the amplitude of activation induced by VUAA1 (Supplementary Fig. 6). 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 support 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 (Supplementary Dataset S2). 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 define more precisely the cradle of the VUAA1 binding cavity (Fig. 5). Using the site-directed mutagenesis approach, all the five residues were mutated in alanine and tryptophan (Fig. 5b) to reduce or increase the steric hindrance of the side chains, respectively.

In contrast to previous results, the mutation in alanine of two phenylalanine residues (F83A and F84A) decreased the response to VUAA1 (Fig. 5b–d) with a greater extent for F84A (medians: 2.33, 0.68 μA for F83A, F84A respectively vs 4.71 μA for WT). Mutation in tryptophan induced the same phenotype in position 83 (F83W), while the mutation in serine had the same impact in position 84 (F84S). Finally, the mutation F84W did not induce a significant change compared to the WT (Fig. 5b). 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 tryptophan, while position 84 is more tolerant to replacement by tryptophan 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 Corcoran et al.²⁶, 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 μA vs 4.71 μA for WT). This effect is strengthened by the introduction of the bulkier and more hydrophobic residue, valine³⁴ (median: 11.68 μA) (Fig. 5b, 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 (5.72 and 6.19 μA , respectively vs 4.71 μA for WT), while the mutation in the bulkier tryptophan induced a significant decrease in the amplitude of activation (2.12 μA) (Fig. 5b, 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 tryptophan 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 (Fig. 5g and Supplementary Table 2) showed a change in I_{max} that was either negative (for F84A and M210W, 1.01 and 1.50 μA , respectively vs 3.97 μA for WT) or positive (for S146V, 6.99 μA vs 3.97 μA for the WT), without affecting the EC_{50} . 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 and that the

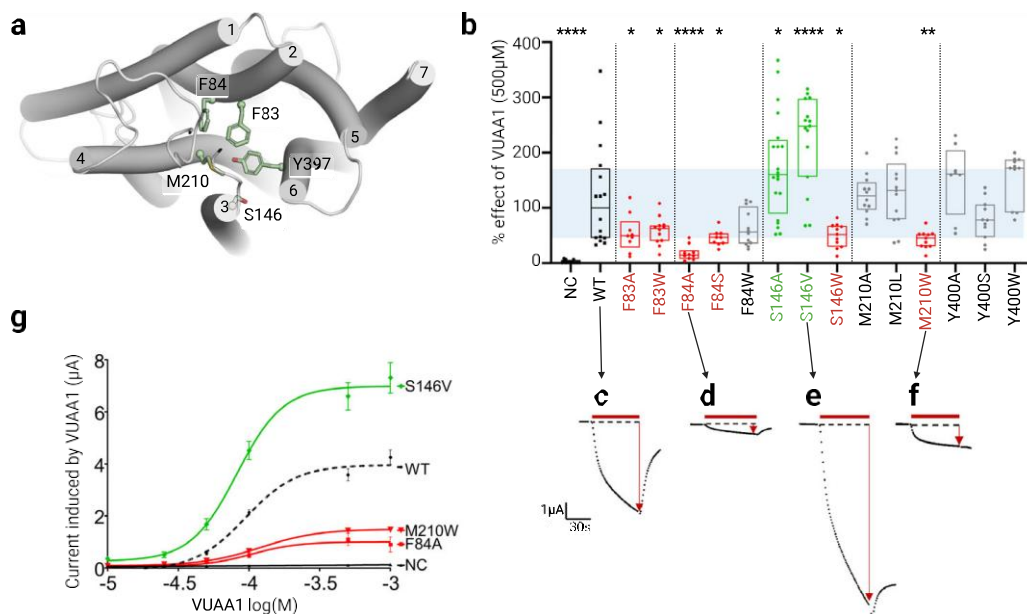


Fig. 5 | Electrophysiological characterization of Orco mutants in the binding cavity. a Close view of the binding region highlighting the position of the different mutation sites. b Boxplot showing the current induced by 500 μM of the ligand VUAA1 and measured by TEVC recordings on DmelOrco WT and mutants expressed by *Xenopus* oocyte. The lower bound of the box plot represents the first quartile, while the upper bound represents the third quartile. The centre is the median. The blue area is a projection of the WT box plot. n = 13 (NC and M210A), n = 16 (WT), n = 11 (F83A, S146W, Y400S, and Y400W), n = 12 (F83W, F84W, and M210A), n = 10 (F84A), n = 9 (F84S), n = 20 (S146A), n = 15 (S146V), n = 13 (M210L), n = 8 (Y400A) 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, with WT used as reference for multiple comparisons test. ****p < 0.0001, **p = 0.0033 (M210W), *p = 0.0421 (F83A), *p = 0.0365 (F83W), *p = 0.0115 (F84S), *p = 0.0176 (S146A), *p = 0.0118 (S146V). Results from mutant showing a statistical decrease or increase from DmelOrco WT. c–f Representative current measured on mutants with statistical differences from DmelOrco WT. g Dose-response curves for the mutants considered. EC50 are 94.5 μM for WT, 82.2 μM for S146V, 120.2 μM for M210W, 105.0 μM for F84A, and not determined for NC. Errors bars correspond to SEM from n = 20 (WT), n = 11 (NC and S146V), n = 8 (F84A), n = 7 (M210W) recordings from different oocytes.

level of expression is not correlated with the mutant response to VUAA1 (Supplementary Fig. 6).

Discussion

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 (Fig. 6). 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 Orcos 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^{35–37} 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^{38–40}. 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^{41,42}. In particular, it has been shown that water plays a crucial role in the binding kinetics⁴³. 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 Y397⁵⁶ 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 MhraOR5.

Comparative analysis of the eugenol-bound MhraOR5 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) (Fig. 6). However, the ORs show a remarkable diversity in the binding site composition. This particularity is also found in mammalian ORs, allowing for broad detection of chemicals^{44–46}.

The polarity and volume of the ligand binding cavity influence 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 (Supplementary Fig. 7). 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.²⁶ 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 to be a good indicator to predict the response to VUAA1 of a given Orco or mutant (Supplementary Fig. 7).

The volume of the cavity also influenced the response of Orco to VUAA1 (Supplementary Fig. 9). A substantial reduction of the volume (such as the introduction of a tryptophan 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 induce a change in the response to VUAA1. An exception was the mutation F84A, which abolished the response to VUAA1, potentially due to a direct interaction with the ligand. These results suggest that the protein could fluctuate to accommodate bulky ligands such as VUAA1, as it has already been shown for olfactory receptors^{6,44,45}.

The architecture of the ligand binding site is conserved among 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.

Interestingly, decreasing the ethyl moiety to a methyl almost abolished the response of Orco¹⁸. In contrast, the replacement of the ethyl group with an isophenyl group improved the potency of the VUAA1-derivative. Altogether, these observations show that increasing hydrophobicity by mutations or ligand modifications

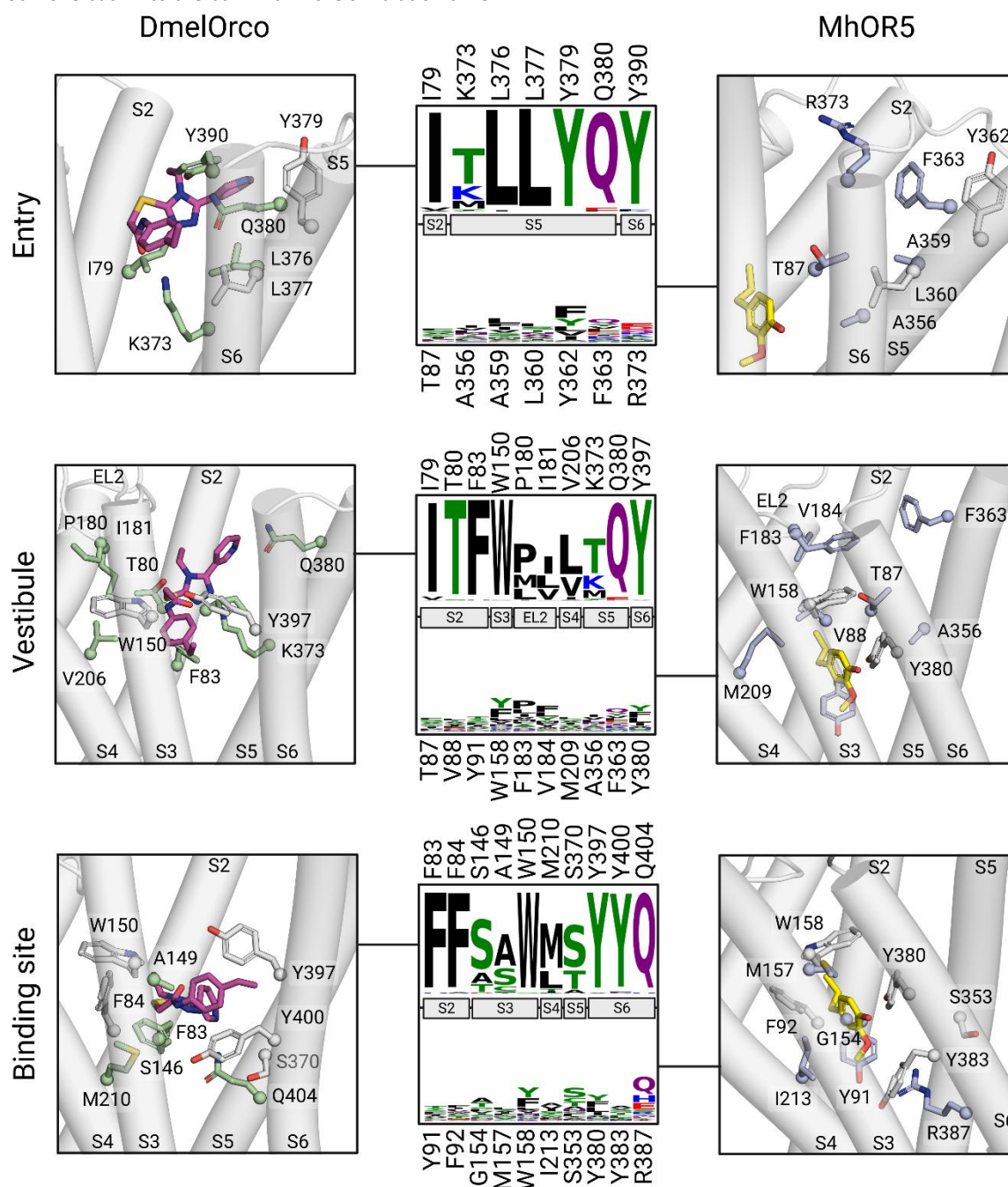


Fig 6 | Structural details and sequence conservation of the entry, vestibule, and binding site in Orcos and ORs. The structures for DmelOrco are extracted from MD simulations, while the structure for MhOR5 is taken from pdb: 7LID. The amino acids were selected according to their frequencies of interaction with VUAA1 during MD simulations. Carbon atoms from amino acids conserved between Orcos and MhOR5 are coloured in white. Carbon atoms from amino acids specific to Orcos or ORs are coloured in green and blue, respectively. Carbon atoms from VUAA1 and Eugenol are shown in purple and yellow, respectively. Residue conservation among 176 Orcos from 174 species and 361 ORs from 4 species are coloured according to their side-chain chemistry. In each frame, the upper con-sensus logo account for Orcos and the lower one for ORs.

We compared our electrophysiological results with already published data on mutants of two ORs: MhOR1 and MhOR5 (Supplementary Table 3). In particular, 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^{S2A}, I107^{S2A}) and MhOR5 (Y91^{S2A} and F92^{S2A}) result in non-responsive channels. In the 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 (Fig. 6 and Supplementary Fig. 4). This conclusion is strengthened by the increased sensitivity of S146^{S3} mutants (alanine and valine).

increases the response of Orco to its ligands.

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 in 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 pocket conservation of the translocation pathway and binding pocket observed in Orco

compared to the high diversity in ORs. The conservation and the variability are then shared between the two subunits forming the heteromer. This combination of the highly con-served Orcos subunit with the more versatile ORs provides the insect with very 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^{47,48}. Our results provide a fine description of the particular binding process, opening the way to a rational design of orthosteric and allosteric modulators.

Methods

The research fully complies with European regulations for animal handling and experiments 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).

In silico modelling

Alignment between Orcos and ORs with MhraOR5. Alignment between MhraOR5 and Orcos was based on the alignment files for 176 Orcos and 361 ORs from Butterwick et al.⁵. MhraOR5 was realigned with the Orcos using ClustalO⁴⁹ 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 pipeline⁵⁰ using PDB 6C70 as a template with default settings of the Alignment Mode. DmelOrco alphafold model (version 07.01.2021) was retrieved from AlphaFold Protein structure database⁵¹. RMSD between the SWISS-model and AlphaFold model was calculated using ccpptraj⁵² 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 MhraOR5. Detection of the pockets of the 176 Orcos plus DmelOrco mutants (SWISS-model), AbakOrco (pdb: 6C70), and MhraOR5 (in APO form, pdb: 7LIC) cavities was carried out using fpocket3⁵³ 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. Propka⁵⁴ 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 server⁵⁵. Five VUAA1 molecules were added in different orientations on the extra-cellular side. The system was embedded into a POPC-only model membrane using PACKMOL-memgen⁵⁶. 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 286,736 atoms, in a $3.4 \times 10^6 \text{ \AA}^3$ 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 membrane⁵⁷. VUAA1 parameters were generated by calculating partial atomic charges with the HF/6-31G* basis set using Gaussian 09⁵⁸. The obtained electrostatic potential was fitted by the RESP program⁵⁹. 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 $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ 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

$200 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ harmonic potential). Each system underwent 50000 steps of equilibration in the NPT ensemble and 1 bar (restraining the protein and ligands with a $15 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ 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} . 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. 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 $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). 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 $5 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$. 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 310 K 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 (Supplementary Fig. 10). 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 MhraOR5 (pdb: 7LID) on each DmelOrco monomer using the cealign pymol command⁶⁰, then calculating the center of mass distance between eugenol and VUAA1 on each trajectory using the mindist pytraj module⁵².

Selection of representative frames for contact, vestibule, and binding. Representative frames of the diffusion were obtained by dividing the prototypical trajectory into 4 steps according to the curve shown in Fig. 2c. 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 ccpptraj⁵² 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 VUAA4¹⁸. 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 and 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 server⁶¹ with default settings. Electrostatic complementarity scores for each system were calculated using the Flare electrostatic complementarity tool⁶².

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 optimized⁶³ 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 electro-phoretic restriction profile and external sequencing (Genewiz). DNAs of positive clones were amplified with Qiagen Midiprep Kit and the ORF was 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 described⁶⁴. Briefly, oocytes were defolliculated after surgical retrieval by type 1 A collagenase over 2–3 h 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. Different batches of oocytes have been tested per construct. The results of the mutants of interest have been confirmed by a second set of experiments comparing the amplitudes with WT in the same day and from the same batch of oocytes (Supplementary Fig. 11). Whole cell currents were recorded with the two-electrode voltage-clamp (TEVC) technique with the HiClamp robot (MultiChannel System). Microelectrodes were filled with 3 M KCl. The high K⁺ buffer used for recordings was composed of 91 mM KCl, 1 mM MgCl₂, 1.8 mM CaCl₂, 5 mM HEPES, pH 7.4. Membrane voltage was clamped to –50 mV and VJAA1 was applied for 60 s. Data were extracted with M. Vivaudou's programs⁶⁵ and statistically analyzed with Prism 8 (Graphpad).

Western Blots

All expression experiments were assessed on 4–20% mini-Protean TGX SDS-PAGE gels (Bio-Rad). 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 (18 g then 27 g) into a solubilization buffer (PBS 1X, protease inhibitor cocktail tablets) and stored at –80 °C. Western blots were performed by transferring proteins onto a nitro-cellulose 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 Genscript) and the secondary antibody Goat anti-rabbit IgG HRP conjugate (1:5000 Sigma-Aldrich) for 1 h 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 (order# U439YGB120-7) and designed against the peptide

sequence SSIPVEIPRLPIKSFYPW (Genscript ref: U5526GC070-1) in the second extracellular loop (ECL2). Anti-Orco was produced in rabbit and purified by antigen affinity (Genscript).

Reporting summary

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

Data availability

The data that support the findings of this study are available from the corresponding authors upon request. The PDB database (www.rcsb.org) has been used to get access to the following entries: 6C70, 7LIC, and 7LID. Source data are provided in this paper.

References

- Mora C, Tittensor DP, Adl S, Simpson AG, Worm B. How many species are there on Earth and in the ocean? *PLOS Biol* **9**, (2011).
- Hansson BS, Stensmyr MC. Evolution of insect olfaction. *Neuron* **72**, 698–711 (2011).
- Bohbot J, Fu L, Le T, Chauhan K, Cantrell C, Dickens J. Multiple activities of insect repellents on odorant receptors in mosquitoes. *Med Vet Entomol* **25**, 436–444 (2011).
- Ray A. Reception of odors and repellents in mosquitoes. *Curr Opin Neurobiol* **34**, 158–164 (2015).
- Butterwick JA, et al. Cryo-EM structure of the insect olfactory receptor Orco. *Nature* **560**, 447–452 (2018).
- Del Marmol J, Yedlin MA, Ruta V. The structural basis of odorant recognition in insect olfactory receptors. *Nature* **597**, 126–131 (2021).
- Sato K, Pellegrino M, Nakagawa T, Nakagawa T, Vosshall LB, Touhara K. Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* **452**, 1002–1006 (2008).
- Wicher D, Miazzi F. Functional properties of insect olfactory receptors: ionotropic receptors and odorant receptors. *Cell Tissue Res*, 1–13 (2021).
- Vosshall LB, Wong AM, Axel R. An olfactory sensory map in the fly brain. *Cell* **102**, 147–159 (2000).
- Dobritsa AA, van Naters WvdG, Warr CG, Steinbrecht RA, Carlson JR. Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* **37**, 827–841 (2003).
- Herre M, et al. Non-canonical odor coding in the mosquito. *Cell* **185**, 3104–3123. e3128 (2022).
- Pitts RJ, Liu C, Zhou X, Malpartida JC, Zwiebel LJ. Odorant receptor-mediated sperm activation in disease vector mosquitoes. *Proc Natl Acad Sci USA* **111**, 2566–2571 (2014).
- Jones PL, Pask GM, Rinker DC, Zwiebel LJ. Functional agonism of insect odorant receptor ion channels. *Proc Natl Acad Sci USA* **108**, 8821–8825 (2011).
- Robertson HM, Warr CG, Carlson JR. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **100**, 14537–14542 (2003).
- Brand P, et al. The origin of the odorant receptor gene family in insects. *elife* **7**, e38340 (2018).
- Chen S, Luetje CW. Identification of new agonists and antagonists of the insect odorant receptor co-receptor subunit. *PLOS ONE* **7**, e36784 (2012).
- Jones PL, et al. Allosteric antagonism of insect odorant receptor ion channels. *PLOS ONE* **7**, e30304 (2012).
- Taylor RW, et al. Structure–activity relationship of a broad-spectrum insect odorant receptor agonist. *ACS Chem Biol* **7**, 1647–1652 (2012).
- Romaine IM, et al. Narrow SAR in odorant sensing Orco receptor agonists. *Bioorg Med Chem Lett* **24**, 2613–2616 (2014).
- Chen S, Luetje CW. Phenylthiophenecarboxamide antagonists of the olfactory receptor co-receptor subunit from a mosquito. *PLOS ONE* **8**, e84575 (2013).
- Pask GM, Bobkov YV, Corey EA, Ache BW, Zwiebel LJ. Blockade of insect odorant receptor currents by amiloride derivatives. *Chem Senses* **38**, 221–229 (2013).
- Bobkov Y, Corey E, Ache B. An inhibitor of Na⁺/Ca²⁺ exchange blocks activation of insect olfactory receptors. *Biochem Biophys Res Commun* **450**, 1104–1109 (2014).
- Nakagawa T, Pellegrino M, Sato K, Vosshall LB, Touhara K. Amino acid

residues contributing to function of the heteromeric insect olfactory receptor complex. *PLOS ONE* **7**, e32372 (2012).

24. Kumar BN, Taylor RW, Pask GM, Zwiebel LJ, Newcomb RD, Christie DL. 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).

25. Turner RM, *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).

26. Corcoran JA, Sonntag Y, Andersson MN, Johanson U, Löfstedt C. 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).

27. Jaiteh M, Rodríguez-Espigares I, Selent J, Carlsson J. Performance of virtual screening against GPCR homology models: Impact of template selection and treatment of binding site plasticity. *PLoS Comput Biol* **16**, e1007680 (2020).

28. Jumper J, *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

29. Yuan S, Dahoun T, Brugarolas M, Pick H, Filipek S, Vogel H. Computational modeling of the olfactory receptor Olfr73 suggests a molecular basis for low potency of olfactory receptor-activating compounds. *Communications biology* **2**, 141 (2019).

30. Dror RO, *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci U S A* **108**, 13118–13123 (2011).

31. Topin J, Diharce J, Fiorucci S, Antonczak S, Golebiowski J. O₂ migration rates in [NiFe] hydrogenases. A joint approach combining free-energy calculations and kinetic modeling. *J Phys Chem B* **118**, 676–681 (2014).

32. Dror RO, Arlow DH, Borhani DW, Jensen MØ, Piana S, Shaw DE. 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).

33. de March CA, *et al.* Odorant receptor 7D4 activation dynamics. *Angew Chem Int Ed* **130**, 4644–4648 (2018).

34. Hoffmann W, *et al.* An intrinsic hydrophobicity scale for amino acids and its application to fluorinated compounds. *Angew Chem Int Ed* **58**, 8216–8220 (2019).

35. Wang G, Carey AF, Carlson JR, Zwiebel LJ. Molecular basis of odor coding in the malaria vector mosquito *Anopheles gambiae*. *Proc Natl Acad Sci USA* **107**, 4418–4423 (2010).

36. Kreher SA, Mathew D, Kim J, Carlson JR. Translation of sensory input into behavioral output via an olfactory system. *Neuron* **59**, 110–124 (2008).

37. Boyle SM, McNally S, Ray A. Expanding the olfactory code by in silico decoding of odor-receptor chemical space. *elife* **2**, e01120 (2013).

38. Schiebel J, *et al.* Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nat Com* **9**, 1–15 (2018).

39. Gajdos L, Blakeley MP, Haertlein M, Forsyth VT, Devos JM, Imberty A. Neutron crystallography reveals mechanisms used by *Pseudomonas aeruginosa* for host-cell binding. *Nat Com* **13**, 1–9 (2022).

40. Mondal J, Friesner RA, Berne B. Role of desolvation in thermodynamics and kinetics of ligand binding to a kinase. *J Chem Theory Comput* **10**, 5696–5705 (2014).

41. Davis AM, Teague SJ. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew Chem Int Ed* **38**, 736–749 (1999).

42. de Freitas RF, Schapira M. A systematic analysis of atomic protein–ligand interactions in the PDB. *Medchemcomm* **8**, 1970–1981 (2017).

43. Setny P, Baron R, Kekenus-Huskey PM, McCammon JA, Dzubiella J. Solvent fluctuations in hydrophobic cavity–ligand binding kinetics. *Proc Natl Acad Sci USA* **110**, 1197–1202 (2013).

44. Charlier L, *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).

45. Topin J, de March CA, Charlier L, Ronin C, Antonczak S, Golebiowski J. Discrimination between Olfactory Receptor Agonists and Non-agonists. *Chem Eur J* **20**, 10227–10230 (2014).

46. Zhang X, Firestein S. The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* **5**, 124–133 (2002).

47. Yan H, *et al.* An engineered orco mutation produces aberrant social

behavior and defective neural development in ants. *Cell* **170**, 736–747. e739 (2017).

48. Fandino RA, *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).

49. Sievers F, Higgins DG. Clustal omega. *Curr Protoc Bioinform* **48**, 3.13. 11–13.13. 16 (2014).

50. Waterhouse A, *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296–W303 (2018).

51. Varadi M, *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).

52. Roe DR, Cheatham III TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* **9**, 3084–3095 (2013).

53. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinform* **10**, 1–11 (2009).

54. Søndergaard CR, Olsson MH, Rostkowski M, Jensen JH. 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).

55. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res* **40**, D370–D376 (2012).

56. Schott-Verdugo S, Gohlke H. PACKMOL-memgen: a simple-to-use, generalized workflow for membrane-protein–lipid-bilayer system building. *J Chem Inf Model* **59**, 2522–2528 (2019).

57. D.A. Case, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. 672 Darden, *et al.*, AMBER 2018, University of California, San Francisco. (2018).

58. Frisch M, *et al.* Gaussian 09, Revision D. 01, Gaussian, Inc., Wallingford CT. See also: URL: <http://www.gaussian.com>, (2009).

59. Wang J, Cieplak P, Kollman PA. 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).

60. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* **11**, 739–747 (1998).

61. Pyrkov TV, Chugunov AO, Krylov NA, Nolde DE, Efremov RG. PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes. *Bioinformatics* **25**, 1201–1202 (2009).

62. Bauer MR, Mackey MD. Electrostatic complementarity as a fast and effective tool to optimize binding and selectivity of protein–ligand complexes. *J Med Chem* **62**, 3036–3050 (2019).

63. Roberts RE, Yuvaraj JK, Andersson MN. Codon optimization of insect odorant receptor genes may increase their stable expression for functional characterization in HEK293 cells. *Front Cell Neurosci* **15**, (2021).

64. Moreau CJ, Dupuis JP, Revilloud J, Arumugam K, Vivaudou M. Coupling ion channels to receptors for biomolecule sensing. *Nat Nanotechnol* **3**, 620–625 (2008).

65. Vivaudou M. eeFit: a Microsoft Excel-embedded program for interactive analysis and fitting of experimental dose–response data. *Biotechniques* **66**, 186–193 (2019).

Author contributions

J.T., J.G., and C.J.M. designed research; J.P., G.A., J.M., and M.P. performed research; J.P., G.A., J.T., and C.J.M. analyzed data; J.T. and C.J.M. wrote the paper.

Competing interests

The authors declare no competing interest.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-44058-5>.

Correspondence and requests for materials should be addressed to Christophe J. Moreau or Jérémie Topin.

Peer review information *Nature Communications* thanks Ren Lai and Shuguang Yuan for their contribution to the peer review of this work. A peer review file is available.

Article

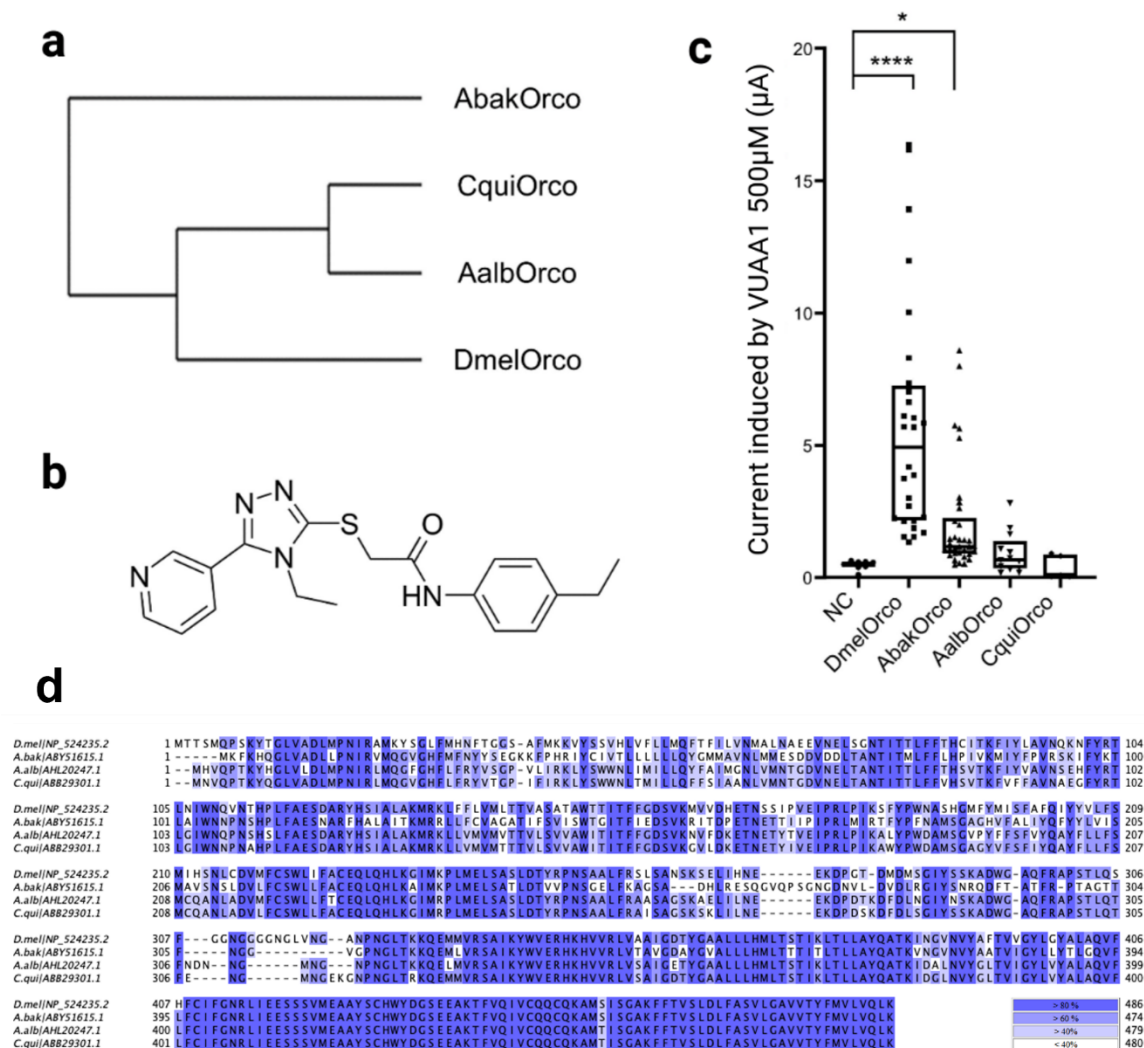
Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

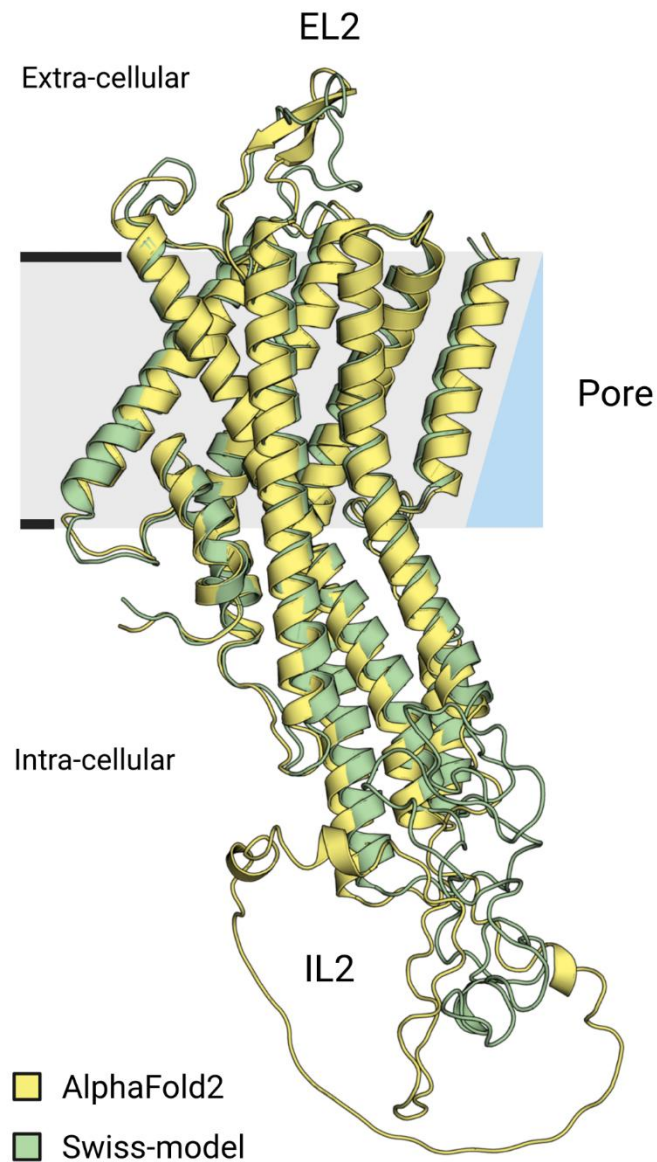
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons

licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

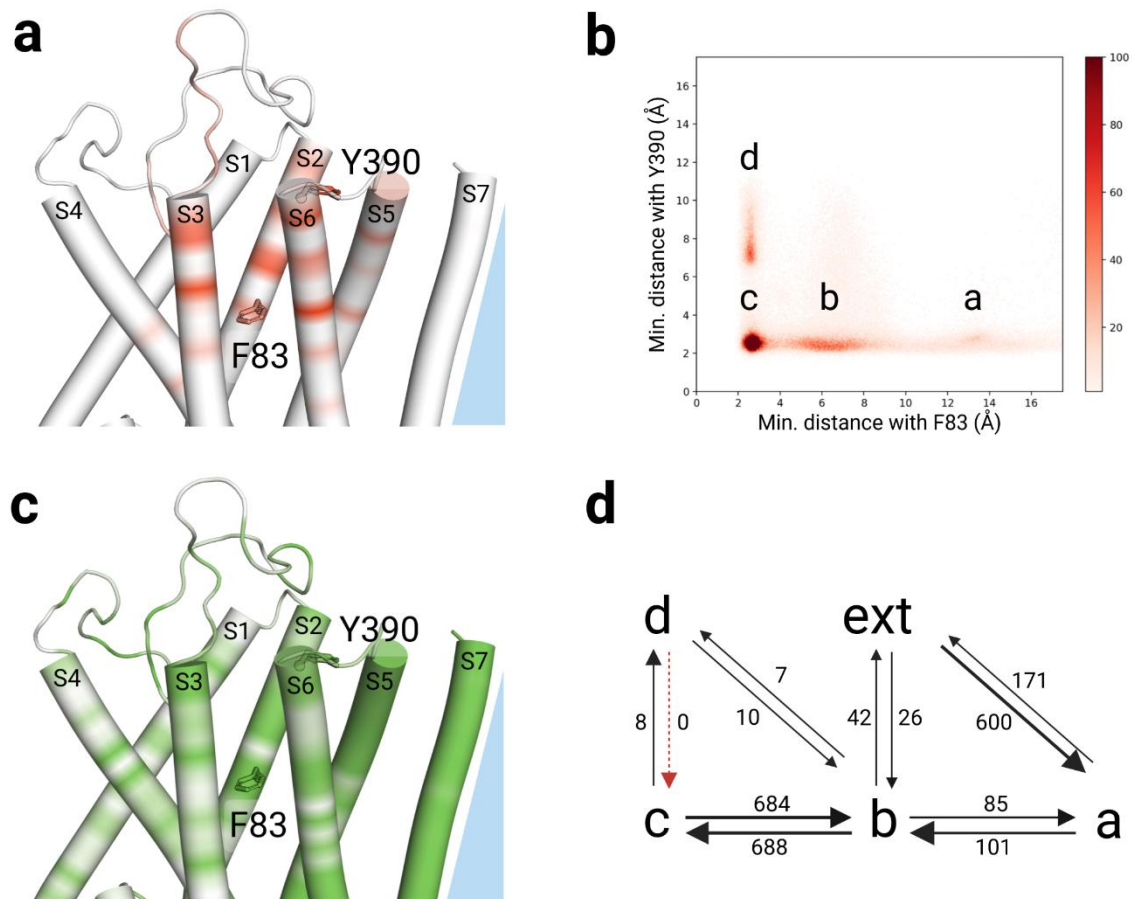
© The Author(s) 2023



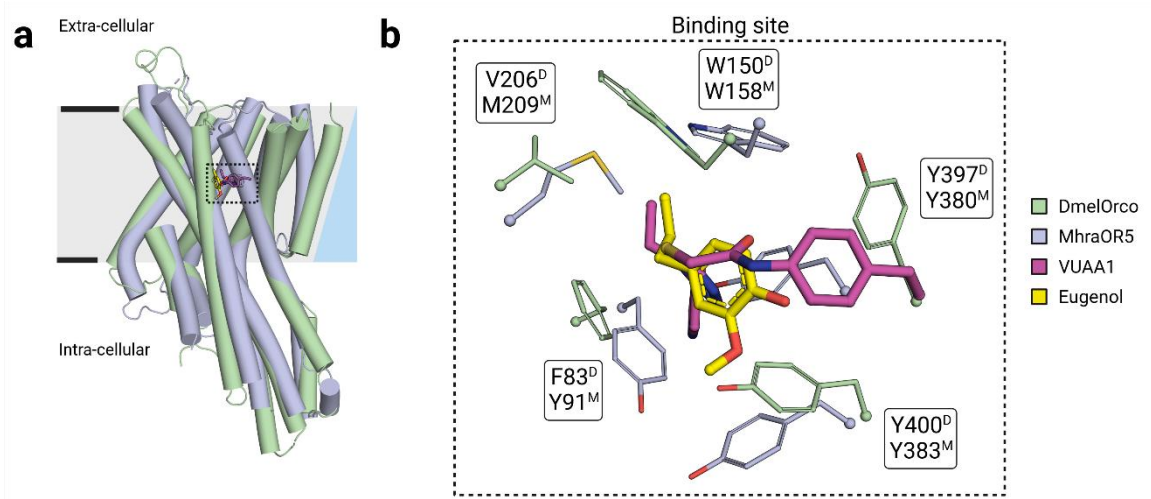
Supplementary Fig. 1 | Determination of the Orco species generating the highest response to VUAA1 in TEVC recordings. (a) Phylogenetic tree of Orcos from different species. *Abak*: *Apocrypta bakeri*; *Cqui*: *Culex quinquefasciatus*, *Aalb*: *Aedes albopictus*, *Dmel*: *Drosophila melanogaster*. (b) Structure of VUAA1. (c) Current amplitudes induced by 500 μ M VUAA1 on Orcos from different species expressed in *Xenopus* oocytes and recorded by TEVC. Results are median $\pm$ SEM. P values are <0.0001 (Dmel, ****), 0.0438 (Abak, *) and 0.92 (Aalb), with NC as reference in the Kruskal-Wallis test. $n = 8$ Negative control, $n = 28$ DmelOrco, $n = 34$ AbakOrco, $n = 13$ AalbOrco, $n = 5$ CquiOrco; NC: negative control (water-injected oocytes). (d) Alignment of the 4 Orco species tested in this study. The amino acids are colored according to the percentage of the residues in each column that agree with the consensus sequence. Only the residues that agree with the consensus residue for each column are colored.



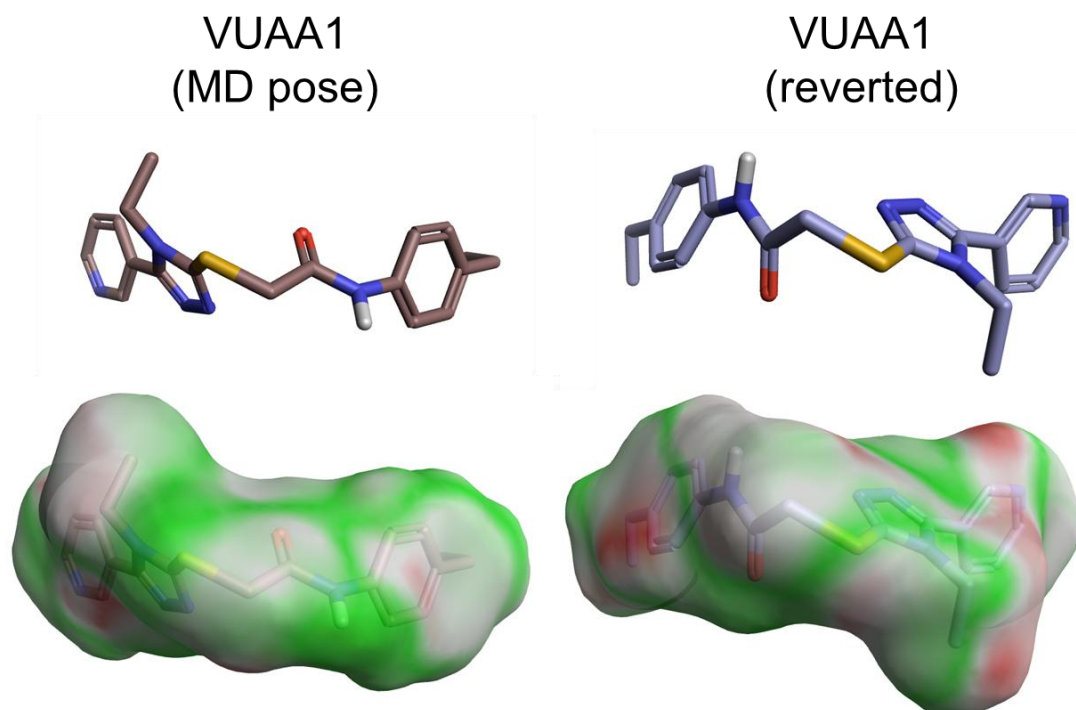
Supplementary Fig. 2 | Superposition of the AlphaFold2 DmelOrco model with the AbakOrco structure. Superposition of the experimental structure of AbakOrco homotetramer (pdb ID: 6C70) superposed on the AlphaFold2 model retrieved from the Alpha Fold Protein database (in yellow). The two structures have a 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.



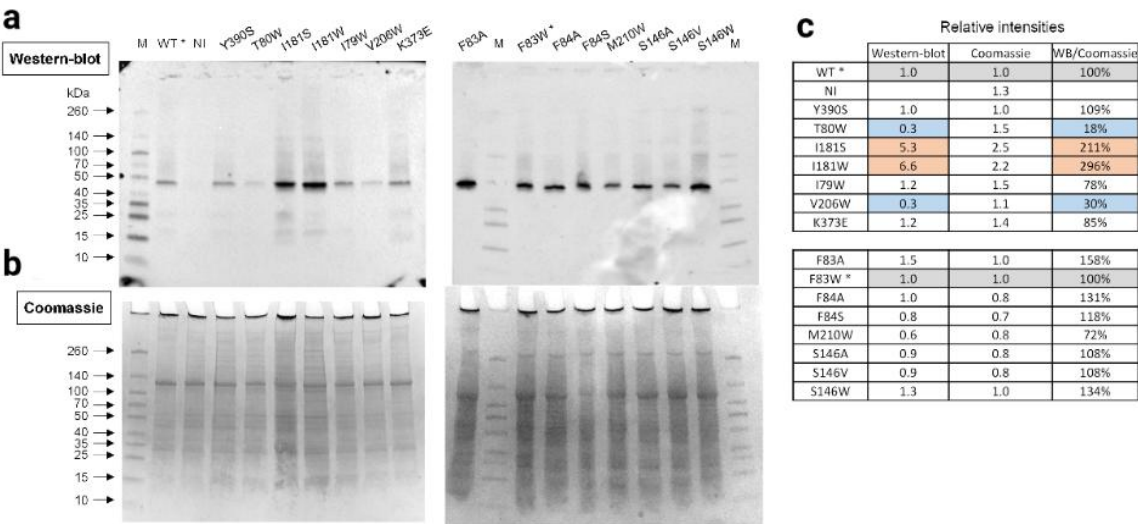
Supplementary Fig. 3 | Summary of the interactions observed during the MD simulations between VUAA1 and DmelOrco. (a) Contact frequencies of VUAA1 mapped onto the structure of DmelOrco. The color gradient account for high frequency interaction (in red) to no interaction (in white). (b) Contour map of 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 basin allow identifying four states in the VUAA1 binding: contact a, entry b, vestibule c and binding d. (c) Residue conservation among 176 Orcos from 174 species mapped onto the structure of DmelOrco using a colour scale from green high conservation, to white low conservation. (d) Proposed mechanism of VUAA1 binding. The numbers indicate the transition from one state to another. ext: extracellular.



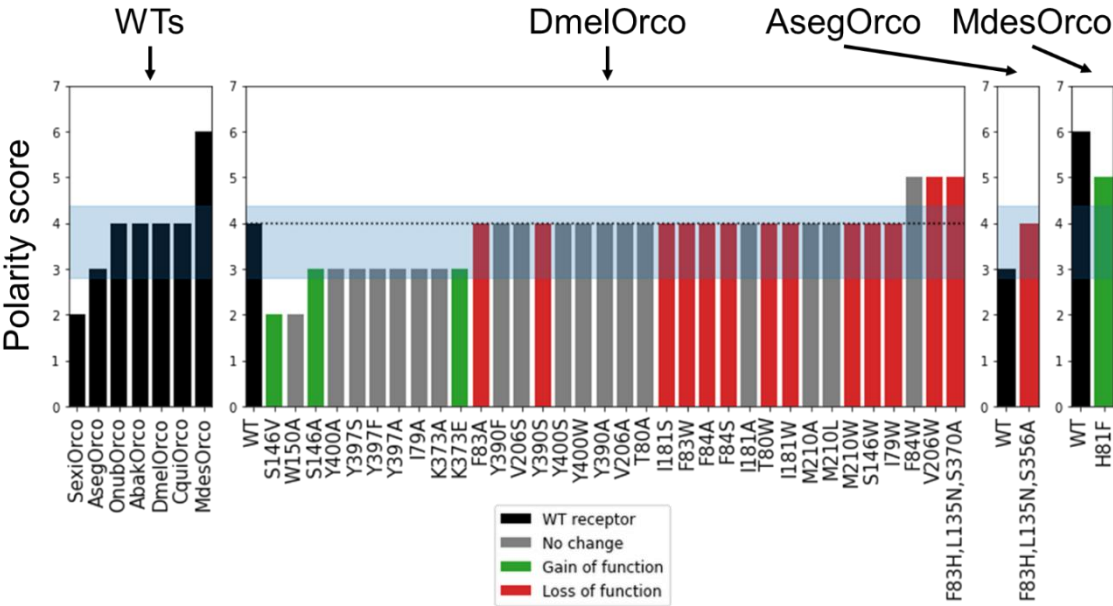
Supplementary Fig. 4 | Comparison of the binding site location predicted in DmelOrco and observed in MhraOR5 structure. (a) Superposition of DmelOrco (green tubes) and MhraOR5 (blue tubes), with their respective ligands VUAA1 (magenta sticks) and eugenol (yellow sticks). (b) Close-up view of the binding site of DmelOrco (green licorice) and MhraOR5 (blue licorice) with their respective ligands VUAA1 (magenta sticks) and eugenol (yellow sticks). The superscripts D or M on the amino acids names represent DmelOrco and MhraOR5, respectively.



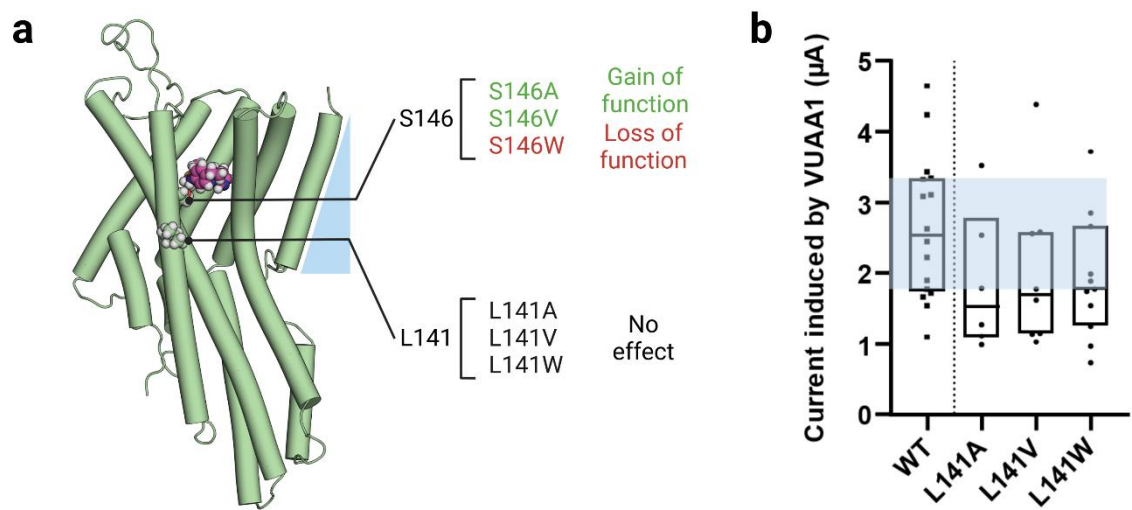
Supplementary Fig. 5 | Comparison of the orientation of VUAA1 in the binding site obtained during the MD simulations (MD pose) and the manually reversed orientation. The electrostatic complementarity of the ligand to the protein appears as a surface. The areas where the protein-ligand electrostatics are favorable or unfavorable are colored from green to red.



Supplementary Fig. 6 | Western-blot analysis of DmelOrco WT and mutant expression in *Xenopus* oocytes. (a) Western blots 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: ThermoScientific 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.

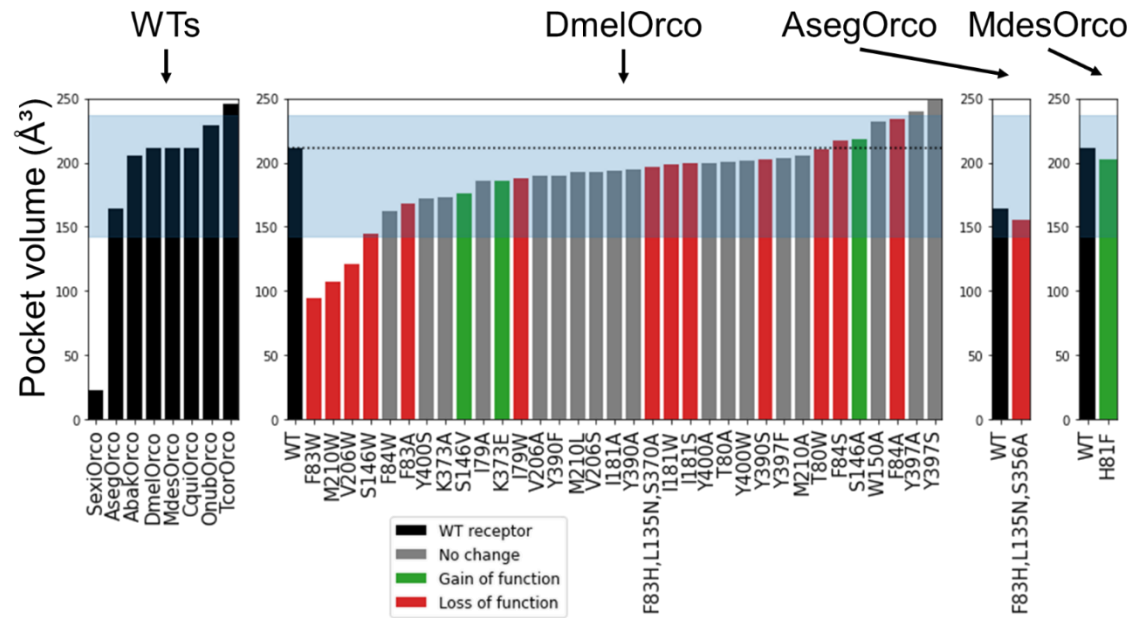


Supplementary Fig. 7 | Evolution of the polarity score of DmelOrco's mutants and from other species computed with fpocket. On the left, the results for WT Orco; for clarity, only the VUAA1-responsive and VUAA1-non-responsive Orcos are shown. SexiOrco and MdesOrco are the receptors with the lowest and 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.

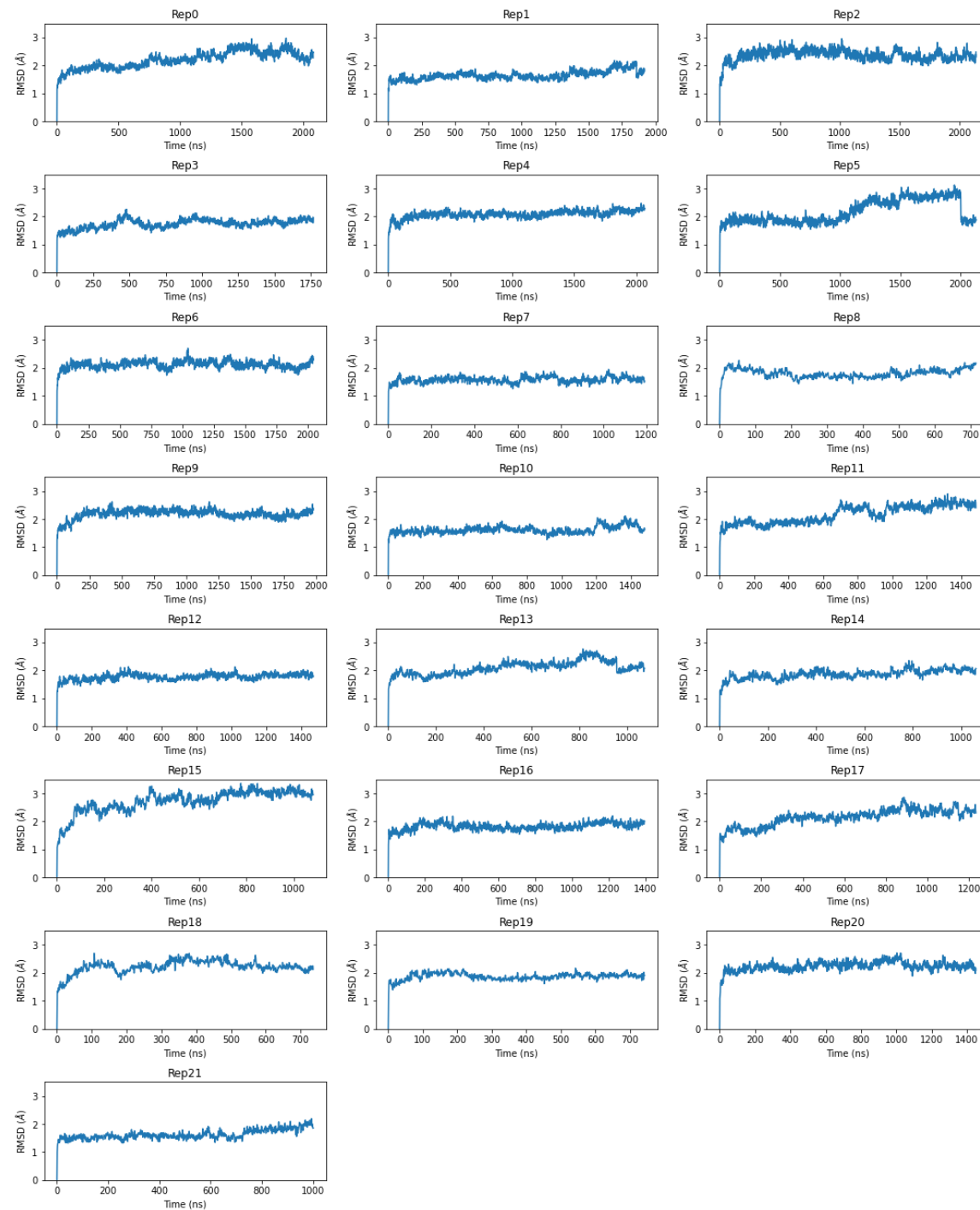


Supplementary Fig. 8 | Characterization of mutations at the position L141 that is not involved in the binding site nor in the translocation pathway of VUAA1. (a) Location of the L141 in the model of DmelOrco. For comparison, S146 was also depicted. The carbon atoms from the residue are in pale green, while the ones of VUAA1 are in purple. (b) Boxplot showing the current induced by 500 μM of the ligand VUAA1 and measured by TEVC recordings on DmelOrco WT and mutants expressed in *Xenopus* oocytes. The lower limit of the box plot represents the first quartile, while the upper limit represents the third quartile. Center bar correspond to the median. The colored blue area is a projection of the WT box plot. $n=16$ (WT), $n=6$ (L141A), $n=8$ (L141V), $n=11$ (L141W). P values are 0.290, 0.396, 0.192 for L141A, L141V and L141W, respectively. Data are analysed with one-way ANOVA with α -error=0.05 followed by Dunn's post-hoc test, with WT used as reference for multiple comparisons test.

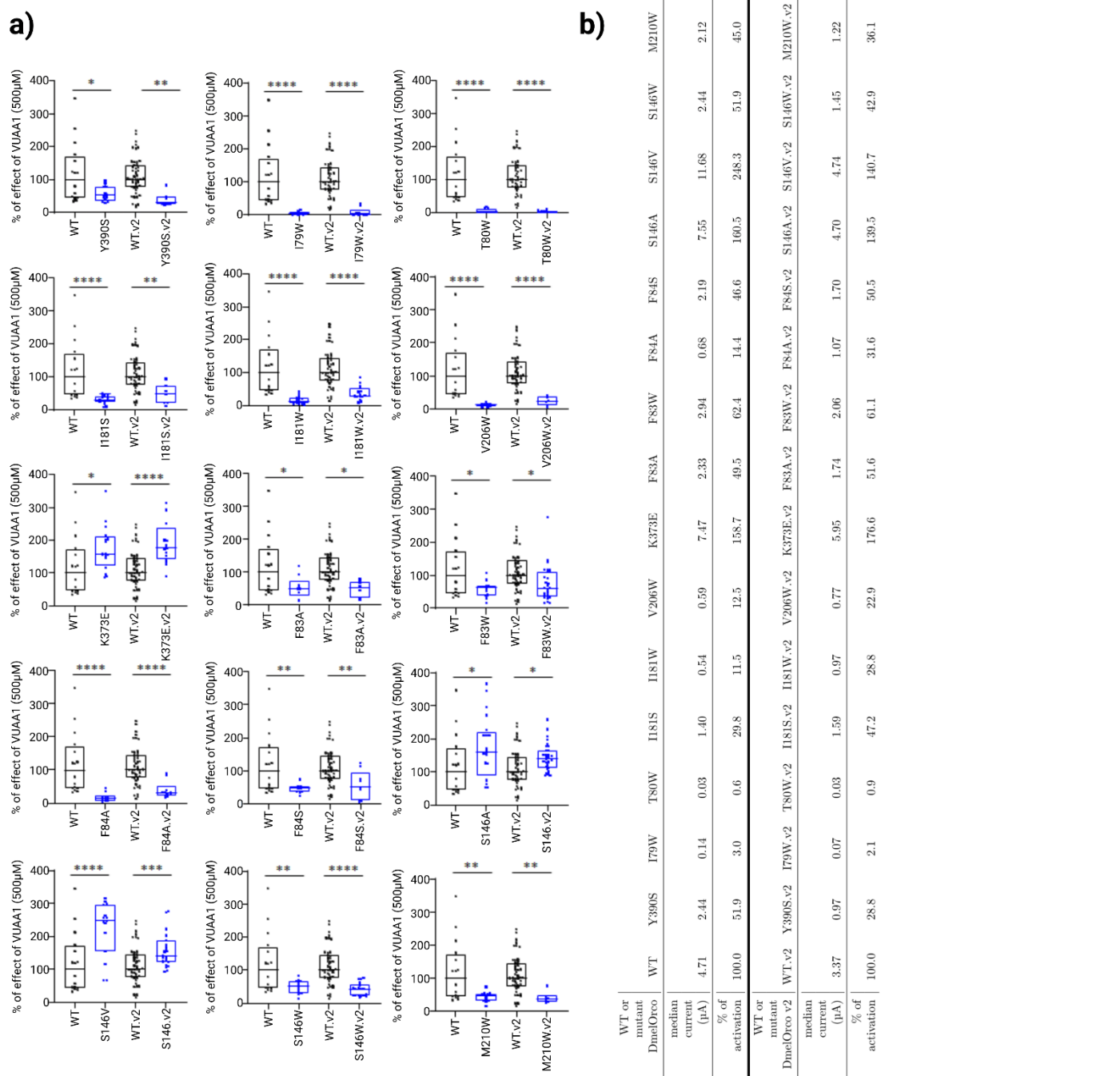
Supplementary Fig. 9 | Evolution of the binding cavity volume of DmelOrco's mutants and from other species computed with fpocket. On the left, the



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.



Supplementary Fig. 10 | Stability of the DmelOrco model in 22 replicas. The RMSD is computed for the receptor backbone (CA, C, N atoms) with respect to the initial model structure.



Supplementary Fig. 11 | Second set of experiments. a) Relative responses to 500 μ M VUAA1 of mutants (blue bars) compared to WT (black bars) in 2 different sets of experiments. Percentages are calculated based on the median of WT current amplitudes as 100%. The experiments in the 2nd set (noted – v2) have been performed in the same day and with the same batch of oocytes as WT.v2. n=49 (WT.v2), n=7 (Y390S.v2; I181S.v2), n=9 (I79W.v2; F84S.v2), n=14 (T80W.v2), n=15 (I181W.v2), n=6 (V206W.v2; F83A.v2; M210W.v2), n=16 (K373E.v2), n=24 (F83W.v2), n=10 (F84A.v2), n=31 (S146A.v2), n=23 (S146V.v2), n=20 (S146W.v2) recordings from different oocytes. P values are symbolized over bars, each star being one decimal. Data are analysed with one-way ANOVA with α -error=0.05 followed by Dunn’s post-hoc test, with WT used as reference for multiple comparisons test. ****p<0.0001, **p=0.0010 (Y390S.v2), *p=0.0145 (F83A.v2), *p=0.0218 (F83W.v2), **p=0.0071 (F84S.v2), *p=0.0165 (S146A.v2), ***p=0.0007 (S146V.v2), **p=0.0032 (M210W.v2). b) Table gathering the median current and the percentage of activation relatively to the WT.

Supplementary Table 1 | Electrostatic and hydrophobic complementary between the series of VUAA1 analogues and different stable locations in the protein. The locations are defined according to Figure 3: entry (b), vestibule (c) and binding site (d). The analogues have been ranked according to the EC50 measured by Tailor et al. on AgamOrco/AgamOR651. The reversed ligand accounts for the reversed orientation described in Supplementary Fig. 9.

Location	Ligand (EC ₅₀)	Hydrophobic complementarity	Electrostatic complementarity
Entry (b)	VUAA0.5	0.37	0.26
	VUAA1	0.41	0.29
	VUAA2	0.44	0.31
	VUAA3	0.46	0.33
	VUAA4	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 (110 μM)	0.65	0.33
	VUAA1 (37 μM)	0.64	0.34
	VUAA2 (9.2 μM)	0.64	0.36
	VUAA3 (8.4 μM)	0.71	0.36
	VUAA4 (2.1 μM)	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

The electrostatic complementarity of the VUAA series in the binding site show little differences in EC values, although it shows a correct trend (Pearson correlation of -0.93). These small differences are in the same order of magnitude with a previous study from Bauer *et al.*² In their study, the 6 ligands are discriminated based on an EC score with subtle EC variations (lig18: IC₅₀ = 2nM, EC = 0.425; lig23: IC₅₀ = 447nM, EC = 0.385).

Supplementary Table 2 | EC50 and I_{max} of VUAA1 from WT and mutated DmelOrco. ND means Not Determined due to the absence of plateau. Separation lines differentiate WT from mutants in the vestibule (Figure 4) and the binding site (Figure 5), in this respective order. Within separation lines, mutants are ranked in descending order of I_{max}, which is measured in concentration-effect curves and slightly lower than the I_{max} measured in single-concentration experiments.

	EC ₅₀ (μM)	I _{max} (μA)
WT	94.5	3.97
I181S	101.9	0.94
I79W	ND	0.65
V206W	ND	0.33
T80W	ND	0.23
S146V	82.2	6.99
M210W	120.2	1.50
F84A	105.0	1.01

Supplementary Table 3 | Comparison of the mutation effects on MhraOR5, MhraOR1 eugenol-induced channel response and DmelOrco VUAA1-induced channel response. A, D, I, na and nt mean Abolished, Decreased, Increased, non-affected and not tested, respectively. Results for MhraOR5 and MhraOR1 are taken from Del Marmol et al.³ Each line corresponds to the same position in the three considered proteins MhraOR5, MhraOR1 and DmelOrco. Alignment between the sequences was taken from ref³.

	MhraOR5	Effect on Eugenol	Effect on DEET	MhraOR1	Effect on Eugenol	DmelOrco	Effect on VUAA1
Ligand diffusion	T87A	D		L102A	na	I79A	na
	L379A	na		L398A	na	G399	nt
Binding site	V88A	D		T103	A	T80A	na
	Y91A	D		Y106A	A	F83A	D
	F92A	A		I107A	A	F84A	D
	S151A	A		S166A	A	S146A	I
	G154A	A		G169A	I	A149	nt
	W158A	A		W173A	A	W150A	na
	M209A	A	D	L227A	D	V206A	na
	M209V	D	I			V206S	n.a
	M209L	D	D			V206W	A
	I213A	A	A	M231A	A	M210A	na
	I231M	D	A			M210W	D
	L379A	na		L398A	na	G399	nt
	Y380A	A		Y339A	A	Y400A	na
	Y383A	A		C402A	na	A403	nt

SI References

1. R. W. Taylor et al., Structure–activity relationship of a broad-spectrum insect odorant receptor agonist. ACS Chem. Biol. 7, 1647-1652 (2012).

2. M. R. Bauer, M. D. Mackey, Electrostatic complementarity as a fast and effective tool to optimize binding and selectivity of protein–ligand complexes. Med. Chem. 62, 3036-3050 (2019).

3. J. Del Marmol, M. A. Yedlin, V. Ruta, The structural basis of odorant recognition in insect olfactory receptors. Nature 597, 126-131 (2021).

Why Life Is Homochiral: Origins and Early Chiral Selection Mechanisms



People, they come together

People, they fall apart

No one can stop us now

'Cause we are all made of stars⁷

Life on Earth is fundamentally defined by molecular asymmetry: both nucleic acids and proteins possess an intrinsic structural “twist.” Elucidating how this asymmetry first emerged remains a central question of my research.

Several hypotheses have been proposed to explain the origin of life, and many emphasize the role of extraterrestrial chemistry in bringing key prebiotic ingredients to the early Earth. Astrochemical observations reveal that complex organic molecules are abundant in interstellar molecular clouds and can survive incorporation into comets or asteroids during the formation of planetary systems. During the early stages of Earth’s history, intense bombardment by such bodies likely delivered a diverse inventory of organic compounds, including amino acids, nucleobases, and other molecules of prebiotic relevance, as evidenced by the analysis of carbonaceous meteorites such as the Murchison meteorite^{1, 2}. These findings suggest that the emergence of life on Earth may have been strongly influenced by chemical processes occurring in space, linking the origin of life to the broader chemical evolution of the cosmos.

Several research groups propose circularly polarized light (CPL) as one of the most plausible astrophysical mechanisms capable of generating the initial symmetry breaking required for biomolecular homochirality. In astrochemical environments such as star-forming regions CPL can arise from scattering processes involving aligned dust grains or magnetically induced polarization fields. Because CPL is itself a chiral physical force, it interacts differently with enantiomers through circular dichroism, enabling asymmetric photochemical reactions³.

Two main mechanisms have been proposed. The first is asymmetric photolysis, in which CPL selectively photodegrades one enantiomer of a racemic mixture, gradually generating an enantiomeric excess (ee). The second is absolute asymmetric synthesis, where CPL directly induces the formation of chiral products from achiral or prochiral precursors⁴. Laboratory experiments have demonstrated that CPL irradiation of amino acids, sugars, and other prebiotic molecules can indeed produce small but measurable enantiomeric excesses. Although the ee values generated by CPL are typically low due to the weak chiral interaction between photons and molecules, they may nevertheless represent the initial bias required for subsequent amplification processes in prebiotic chemistry⁵.

⁷ We are all made of stars, Moby, 2002.

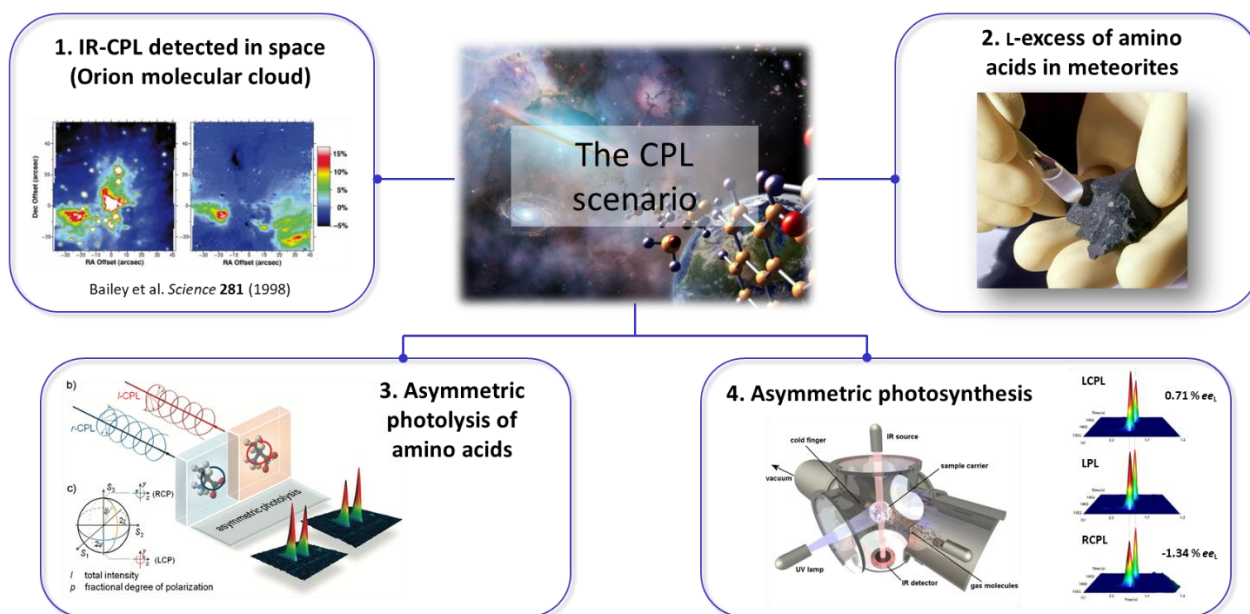


Figure 13. Illustration of the CPL scenario for the origin of building block of life. Astrophysical observations reveal the presence of circularly polarized light in star-forming regions (top left), providing a potential chiral physical force capable of inducing asymmetric photochemical reactions in interstellar or protostellar environments. The prebiotic molecules present in interstellar ices or proto-planetary materials may be exposed to CPL, leading to small enantiomeric excesses through asymmetric photolysis or synthesis. These molecules can subsequently be incorporated into meteorites and delivered to the early Earth (top right), where they represent potential precursors of biological homochirality. Laboratory experiments reproduce this process by irradiating organic molecules with controlled CPL (bottom left), while advanced spectroscopic setups allow the measurement of the resulting chiral signatures and enantiomeric excesses (bottom right).

Altogether, these observations support the possibility that the asymmetry of life originated in interstellar environments, where CPL may have induced an initial chiral bias in prebiotic molecules. Such processes could have selectively enriched specific amino acids and sugars, ultimately contributing to the emergence of biological homochirality.

Uncategorized References

1. Ehrenfreund, P.; Charnley, S. B., Organic molecules in the interstellar medium, comets, and meteorites: a voyage from dark clouds to the early Earth. *Annual Review of Astronomy and Astrophysics* **2000**, 38 (1), 427-483.
2. Irvine, W. M., Extraterrestrial organic matter: a review. *Origins of Life and Evolution of the Biosphere* **1998**, 28 (4), 365-383.
3. Jorissen, A.; Cerf, C., Asymmetric photoreactions as the origin of biomolecular homochirality: a critical review. *Origins of Life and Evolution of the Biosphere* **2002**, 32 (2), 129-142.

Origin of Life

4. Myrgorodska, I.; Meinert, C.; Hoffmann, S. V.; Jones, N. C.; Nahon, L.; Meierhenrich, U. J., Light on chirality: absolute asymmetric formation of chiral molecules relevant in prebiotic evolution. *ChemPlusChem* **2017**, 82 (1), 74-87.
5. Garcia, A. D.; Meinert, C.; Sugahara, H.; Jones, N. C.; Hoffmann, S. V.; Meierhenrich, U. J., The astrophysical formation of asymmetric molecules and the emergence of a chiral bias. *Life* **2019**, 9 (1), 29.

2.3 – PUBLICATIONS RELATED TO RESEARCH ACTIVITIES ON ORIGIN OF LIFE

- C. Meinert, A. D. Garcia[‡], J. Topin[‡], N. C. Jones, M. Diekmann, R. Berger, L. Nahon, S. V. Hoffmann, U. J. Meierhenrich. Amino acid gas phase circular dichroism and implications for the origin of biomolecular asymmetry, [Nature Communications](#), **2022**.

We investigated how chiral photons can induce molecular asymmetry in the gas phase, offering insights into the origin of life's preference for L-amino acids. Using gas-phase circular dichroism (CD) spectroscopy and *ab initio* calculations, we measured intrinsic chiroptical properties of amino acids without environmental distortions. Our results support the idea that asymmetric photochemical reactions in space may have contributed to the emergence of biological homochirality. In this work, I performed QM calculations and participate to experimental measurements.

- Garcia[‡], J. Topin[‡], J. Bocková, N. C. Jones, S. V. Hoffmann, & C. Meinert. Chiroptical activity of gas-phase propylene oxide predicting the handedness of interstellar circular polarization in the presolar nebula. [Science Advances](#), **2022**.

We report the first experimental measurements of the anisotropy spectra of gas-phase propylene oxide in the vacuum ultraviolet range, a key step toward understanding chiral asymmetry in the interstellar medium. As the first chiral molecule detected in space, propylene oxide raises important questions about the cosmic origins of biomolecular chirality. Our results provide indirect measurement of the polarized light which irradiated the early stages of our star and planet formation. I performed all quantum mechanical calculations and also contributed to the experimental measurements, helping to produce high-resolution electronic circular dichroism data that will support future advances in computational spectroscopy.

ARTICLE


<https://doi.org/10.1038/s41467-022-28184-0>

OPEN

Amino acid gas phase circular dichroism and implications for the origin of biomolecular asymmetry

Cornelia Meinert ^{1✉}, Adrien D. Garcia^{1,5}, Jérémie Topin ^{1,5}, Nykola C. Jones ², Mira Diekmann³, Robert Berger³, Laurent Nahon ⁴, Søren V. Hoffmann ² & Uwe J. Meierhenrich ^{1✉}

Life on Earth employs chiral amino acids in stereochemical L-form, but the cause of molecular symmetry breaking remains unknown. Chiroptical properties of amino acids – expressed in circular dichroism (CD) – have been previously investigated in solid and solution phase. However, both environments distort the intrinsic charge distribution associated with CD transitions. Here we report on CD and anisotropy spectra of amino acids recorded in the gas phase, where any asymmetry is solely determined by the genuine electromagnetic transition moments. Using a pressure- and temperature-controlled gas cell coupled to a synchrotron radiation CD spectropolarimeter, we found CD active transitions and anisotropies in the 130–280 nm range, which are rationalized by *ab initio* calculation. As gas phase glycine was found in a cometary coma, our data may provide insights into gas phase asymmetric photochemical reactions in the life cycle of interstellar gas and dust, at the origin of the enantiomeric selection of life's L-amino acids.

¹Institut de Chimie de Nice, Université Côte d'Azur, UMR 7272 CNRS, 06108 Nice, France. ²ISA, Department of Physics and Astronomy, Aarhus University, 8000 Aarhus, Denmark. ³Fachbereich Chemie, Philipps-Universität Marburg, 35032 Marburg, Germany. ⁴Synchrotron SOLEIL, L'Orme des Merisiers, 91192 Gif-sur-Yvette, France. ⁵These authors contributed equally: Adrien D. Garcia, Jérémie Topin. ✉email: cornelia.meinert@univ-cotedazur.fr; uwe.meierhenrich@univ-cotedazur.fr

Evolution has led to proteins with an incredible array of structures and functions. All proteins have in common that their molecular structure exclusively uses amino acids in stereochemical L-configuration. However, the origin of the molecular asymmetry and the conformational landscape of the 21 left-handed proteinogenic amino acids that orchestrate the assembly of proteins into their unique, biologically relevant three-dimensional structures¹, is still not properly understood. The importance of studying isolated pure amino acids in the gas phase becomes apparent, not only in terms of changes to their conformation due to (i) the absence of the polarizability of the surrounding medium², (ii) intermolecular hydrogen bonding or coordination with adjacent molecules, but also (iii) because amino acids exist as zwitterions in solution and solid state³, whereas they are in the non-ionic molecular form when isolated in the gas phase^{4,5}—the configuration amino acids adopt in proteins and polypeptides⁶ as well as in interstellar gas clouds.

In the last few decades, experimental spectroscopic approaches using electronic or vibrational spectroscopy^{7–11}, microwave and millimeter wave spectroscopy^{12–15}, as well as photoelectron circular dichroism (PECD)^{16,17} for probing the structural properties of amino acids in the gas phase collectively aimed to map the conformational behavior of protein-embedded amino acids. However, absorption-based chiroptical measurements in the gas phase have so far been inaccessible for amino acids because the target molecules show insufficient vapor pressure leading to low gas phase density and therefore to poor signal-to-noise ratios. The very few efforts made, for instance via infrared Cavity-Ring Down Spectroscopy (CRDS)¹⁸ to probe circular birefringence and circular dichroism in the gas phase, have so far been limited to volatile organic species not including amino acids¹⁹. First attempts in the VUV on alanine, even with the very-sensitive technique of ion yield, provided only upper limits without succeeding in providing the absolute configuration²⁰. With a specifically constructed gas cell (Fig. 1) coupled to a synchrotron spectropolarimeter to lower the beam divergence that occurs with increased path length (Supplementary Methods), we have now recorded the CD and anisotropy spectra for enantiomeric pairs of gas phase amino acids including alanine (Fig. 2), 2-aminobutyric acid, proline, valine, norvaline, isovaline, and leucine (Supplementary Note 1) in the vacuum-ultraviolet and far UV regions.

Results

Gas phase chiroptical spectra. The gas cell with an optical path of 500 mm (Fig. 1) allowed constant elevated temperatures (423–473 K) to be maintained, at which sufficient vapor pressures of our target molecules enabled reproducible measurements of the chiroptical properties (Supplementary Note 2), while minimizing thermal decomposition (Supplementary Note 3). We obtained highly consistent data on pure gas phase molecules without chiroptical secondary effects such as those linked to the first polarized solvation shell, chiral induction effects, and local micro-crystallization, as observed in condensed phases. Solvent cut-off effects did not limit the energetic range of these spectra. The L-alanine enantiomer shows negative CD states at $\lambda = 134, 174$, and 236 nm; positive CD states at 152 and 205 nm. D-alanine shows the expected mirroring effect (Fig. 2).

The anisotropy spectrum, which is the absorbance (A) normalized CD spectrum, $g(\lambda) = CD(\lambda)/A(\lambda)$, is directly related to the enantiomeric excess (ee) inducible by enantioselective photolysis²¹. Compared to the CD spectrum, the anisotropy spectrum has enhanced intensity at long wavelength, despite the much more intense CD states at shorter wavelengths. Our data show that highest enantiomeric excesses can be photochemically induced into alanine at $\lambda = 214$ nm and $\lambda = 244$ nm.

An overview of gas phase CD and anisotropy spectra of the investigated D-amino acids and L-amino acids is given in Table 1 and Supplementary Table 1, with full spectra shown in the supplementary information. Both the sign and the magnitude of the gas phase CD and anisotropy bands are shown to be sensitive to the lateral side chain of amino acids (Fig. 2c). Highest correlation in the position and magnitude of CD bands are observed for the three straight-chain amino acids alanine, 2-aminobutyric acid, and norvaline. Methyl group substitution at the alpha carbon, as in the non-coded amino acid isovaline, led to sign inversion of almost all transitions. The unique five-membered nitrogen heterocycle moiety in proline resulted in hypsochromic shifts of the two lowest energy transitions, as well as different spectral features in the high-energy region compared to all other amino acids investigated. Chiroptical spectra of serine could not be measured because enantiomers decomposed upon sublimation, while phenylalanine showed very high absorbance but very weak CD bands and consequently weak anisotropies in the lower 10^{-4} range.

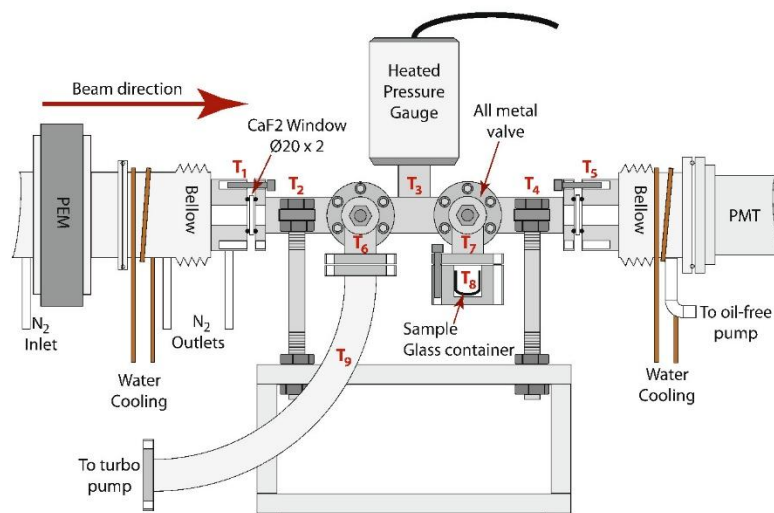


Fig. 1 Temperature and pressure-controlled gas cell providing sufficient molecular density to perform gas phase CD and anisotropy spectroscopy. The photo elastic modulator (marked PEM) produces alternating left-handed and right-handed polarized light. The detector is a photomultiplier tube (marked PMT). The nine different temperature monitor and control points are labeled T1–T9.

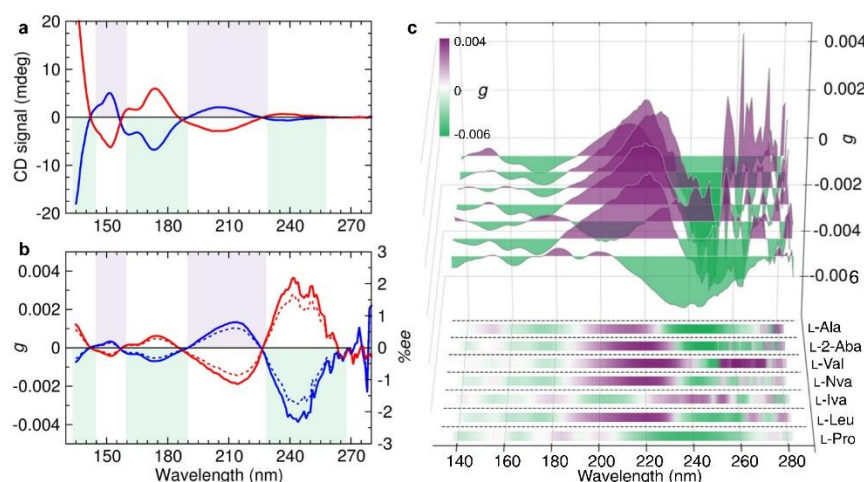


Fig. 2 Chiroptical gas phase spectra of amino acid enantiomers. **a** CD spectra of D-alanine (red); L-alanine (blue). Green and purple areas are indicative of the sign of the L-alanine CD spectrum for comparison with other amino acids. **b** Anisotropy (g) spectra of D-alanine (red) and L-alanine (blue) (thick lines) revealing the %enantiomeric excess (%ee) inducible by either left or right circularly polarized light at the extent of reaction $\xi = 0.9999$ (dotted lines). **c** Anisotropy spectra (g) of seven L-amino acids. Green/purple color and brightness indicate negative/positive and magnitude, respectively, of g values. Source data of CD and g spectra are provided for all amino acid enantiomers as a Source Data file.

Table 1 Amino acid gas phase chiroptical properties and inducible ee values for an extent of reaction $\xi = 0.9999$.

Compound	Transition A			Transition B			Transition C		
	λ/nm	g	ee (%) ^a	λ/nm	g	ee (%) ^a	λ/nm	g	ee (%) ^a
L-Alanine	244	−0.0039	[1.77]	214	0.0013	[0.61]	173	−0.0007	[0.32]
L-2-Aminobutyric acid	244	−0.0042	[1.90]	220	0.0029	[1.35]	176	−0.0005	[0.25]
L-Proline	236	−0.0021	[1.62]	200	0.0002	[0.20]	192	−0.0001	[0.08]
L-Valine				226	0.0031	[1.42]	176	−0.0004	[0.17]
L-Norvaline	245	−0.0016	[0.75]	219	0.0025	[1.14]	175	−0.0006	[0.27]
L-Isovaline	246	0.0017	[0.77]	203	−0.0008	[0.38]	179	0.0003	[0.12]
L-Leucine	247	−0.0019	[0.88]	221	0.0027	[1.22]	174	−0.0007	[0.30]

^aThe lower limit of inducible ee is calculated using²¹: $[\%ee] \geq (1 - (1 - \xi)^2) \times 100\%$, with $\xi = 0.9999$.

The newly recorded CD and anisotropy spectra of gas phase amino acids greatly differ in band position and intensity from those recorded in solution^{22,23} and their isotropic solid state^{21,24}. Obviously, the available spectral region in solution is limited due to solvent absorption. The CD and anisotropy spectra in aqueous solution of all amino acids are dominated by a single broad band at about 205 ± 10 nm corresponding to the $n_{\text{O}} \rightarrow \pi^*_{\text{CO}}$ electronic transition of the carboxylate anion chromophore²⁵. Whereas solid-state absorption and CD spectra of amino acids provided new insights into the intrinsic amino acid structures in the absence of any solvent and enabled chiroptical measurements in the far UV region^{21,24}, it still probes the amino acid zwitterion structure rather than the inherent properties of the amino acids' non-ionic forms as in the gas phase. Moreover, interactions of adjacent molecules in the isotropic condensed phase, leading to conformational changes and spectral congestion, are expected.

Conformer-specific CD and band assignment from quantum chemistry. Quantum chemical ab initio calculations were used to interpret the spectroscopic results. Amino acids are flexible molecules due to multiple torsional degrees of freedom. Their conformational landscape in the gas phase is strongly influenced by stabilization through intramolecular hydrogen bonding, which results in numerous low-energy conformers that contribute

independently to the overall electronic spectrum and challenges the prediction of the conformational distribution. Several basis sets have been selected for an in-depth analysis of gas-phase CD and anisotropy spectra of alanine and details are provided in the Supplementary Methods. The best agreement between theory and experiment for the excitation energies and transition moments on the time-dependent density functional theory (TDDFT) level was obtained applying the M06-2X/aug-cc-pVQZ basis set (Supplementary Fig. 12). High level structural data and relative energies were obtained on the explicitly correlated coupled cluster level (CCSD(T)-F12), with the corresponding CD signals being computed on the approximate second order coupled cluster level CC2 (Supplementary Figs. 13 and 14 and Supplementary Data 1 and 2).

Gas phase alanine reportedly has five energetically close-lying conformers in the ground electronic state, which are connected by low energy barriers^{14,15,26,27}. To sample the most abundant conformers, we ran a typical flowchart combining molecular dynamics simulations and quantum calculation²⁸. Seven stable conformations with relative energies of up to 8 kJ mol^{-1} (Fig. 3a) were considered for TDDFT calculations and a Boltzmann distribution was applied to the set of conformers at different temperatures (300 and 460 K). Comparison of these conformers with previous work is reported in Supplementary Table 3. Temperatures in the range of 433–458 K were necessary to

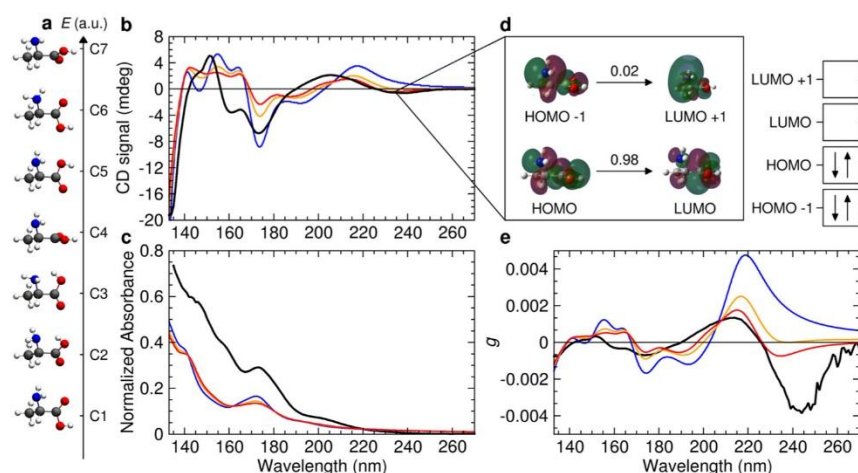


Fig. 3 Time-dependent density functional theory derived chiroptical gas phase spectra of L-alanine. **a** The seven low-energy conformers considered for theoretical calculations (M06-2X/aug-cc-pVQZ). **b, c, e**, Comparison of experimental gas phase (black line) and theoretical spectra of L-alanine (M06-2X/aug-cc-pVQZ) calculated at three different temperatures, $T = 0$ K, lower energy (blue), Boltzmann weighted at $T = 300$ K (yellow), and Boltzmann weighted at $T = 460$ K (red). **b** CD spectra, **c** absorbance spectra and **e** anisotropy spectra. **d** TD-DFT calculations allow for the attribution of the first chiroptical transition with the oscillator strengths f given for the involved natural transition orbitals.

provide adequate vapor pressures for CD and anisotropy measurements. TDDFT calculations show that the energy and intensity of the chiroptical transitions do not vary significantly in between $T = 0$ K and $T = 460$ K, importantly suggesting that the experimental conditions are consistent with those expected for the interstellar medium. Differences between measured and DFT-calculated absorbance below 200 nm (Fig. 3c) is explained by both the presence of photodegradation products, such as water and ammonia in the gas cell (Supplementary Figs. 9 and 10), and due to deficiencies of the quantum chemical methods when approaching the ionization threshold. The photodegradation products (Supplementary Note 3 and Supplementary Fig. 11) were kept to a minimum by operating the gas cell under a flow condition, where the sample was constantly replenished via simultaneous inlet and pumping (Supplementary Methods). The experimental anisotropy values represent therefore a lower limit, especially in the far UV.

The long-wavelength band only appears when considering alanine conformers at higher temperature (Fig. 3b, e). The g band observed at 244 nm, with comparatively broad vibronic profile characteristic of a $n_O \rightarrow \pi^*_{CO}$ (HOMO-LUMO) transition in carbonyl compounds (Fig. 3d), results predominantly from the higher energy conformer 4 (Supplementary Figs. 12 and 14). This conformer adopts an unusual conformation with the amine and carboxylic group found in two orthogonal planes (Fig. 3a), while it is also adopting a planar *cis* arrangement of the carboxyl functional group, which increases its stability²⁷. Previous experimental work allowed the detection of a total of four individual neutral alanine conformers including conformers 1 and 2 by electron diffraction²⁹, conformers 1 and 3 by microwave^{14,15} and PECED spectroscopy¹⁶, and conformer 5 by jet-cooled Raman spectroscopy³⁰. To the best of our knowledge, our gas phase chiroptical analyses, as conducted in combination with quantum chemical calculations, provide the first experimental evidence of conformer 4 of neutral alanine. Gas phase CD and anisotropy spectroscopy can therefore offer crucial information for elucidating conformation-specific configurations, as CD signs and magnitudes are very sensitive to conformations.

Whereas the CD spectrum at 460 K estimated on the TDDFT level remains nearly featureless in the 140–160 nm spectral

region, both the CD spectrum on the coupled cluster level (Supplementary Fig. 13) as well as its experimental counterpart show pronounced structure. The latter displays inversion of the CD signal from positive to negative with a zero crossing near 157 nm, which is partially reproduced on the CC2 level, but found to be more expanded reaching down to 150 nm. This type of CD signal is typical for an exciton coupling between two close identical chromophores³¹ and could therefore indicate the presence of a dipeptide. Investigations of polymerization and cyclization of alanine at elevated temperatures is found in the Supplementary Note 5 and shows that sublimation does not lead to the formation of significant amounts of dipeptides or diketopiperazines (Supplementary Fig. 15). We found that commercially available L-alanine contains less than 1% of L-alanyl-L-alanine; hence, dipeptides are not a priori expected to influence the alanine spectrum significantly. However, since exciton bands are often more intense than CD bands from the monomer^{32,33} even small amounts of dipeptides could potentially contribute to the difference from the calculations.

While we can confirm the assignments of previously proposed states of the electronic spectrum of neutral alanine³⁴, our experimental and theoretical data extend our understanding of hitherto unassigned bands of neutral amino acids. Most of the primary low-lying singlet transitions in gas phase amino acids can be attributed to the valence excitation of the highest lying non-bonding molecular orbital localized at the carboxylic oxygen atom (n_O). Thus, besides the broad $n_O \rightarrow \pi^*_{CO}$ band, a series of transitions belonging to the $n_O \rightarrow 3sp$ Rydberg excitations are predictable (Table S1). Moreover, our gas phase data allow for the measurement of the first Rydberg excitations from the non-bonding lone pair on the nitrogen atom ($n_N \rightarrow 3sp$) of the neutral amine group, inaccessible in solution and solid state. Theoretical approaches to assign the observed CD bands of higher lying excited states to specific orbital transitions become less accurate as significant mixing occurs. We tentatively assign the valence $\pi_{C=O} \rightarrow 3s$ and $\pi_{C=O} \rightarrow \pi^*_{CO}$ states of neutral amino acids to be located below 155 nm. Upon quantitative and qualitative examination, important differences between proline and all other amino acids come to the fore as the five-membered pyrrolidine ring introduces significant distortions to the conformational

space, so that band positions, signs, and magnitudes of all electronic excitations differ considerably.

Discussion

Gas phase optical activity as the origin of extra-terrestrial symmetry breaking. Life's amino acid preference for left-handedness is still a central puzzle in modern biochemistry, with several theoretical and experimental causes discussed³⁵, refs therein such as molecular parity violation, selective adsorption on minerals surfaces, and enantioselective interactions with circularly polarized light or spin-polarized electrons. The detection of L-enriched amino acids in the Murchison meteorite^{36,37} suggested the possibility of an extra-terrestrial contribution to the missing link between purely racemic molecules on prebiotic earth and homochirality of life. Substantial progress has been made in the understanding of the photochemical formation of life's molecular building blocks in interstellar molecular clouds based on the identification of amino acids³⁸, aldehydes³⁹, and sugar molecules⁴⁰ in simulated interstellar pre-cometary ices. Obtained results were confirmed by in situ data of ESA's cometary Rosetta mission and its COSAC⁴¹ and Ptolemy⁴² instruments on board the Philae Lander. Most of the current models of comet nuclei presume that to a major extent they are the aggregates of interstellar icy grains in their evolved state in the collapsing molecular cloud. Gas and solid-state chemical reactions induced by UV photolysis (unpolarized and polarized), cosmic rays or thermal cycling form and process the complex organic inventory of interstellar dust that is later integrated in the presolar nebula from which planets and small solar system bodies formed⁴³. Besides, Rosetta's ROSINA instrument detected the achiral amino acid glycine in the cometary coma⁴⁴ allowing the assumption that comets may contain more complex amino acids that can desorb into the comet's atmosphere at sufficiently high temperatures.

As several sources of naturally polarized light have been detected in interstellar star-forming regions^{45,46} and given the strong evidence that the solar system originated in a high-mass star-forming region⁴⁷, this has been considered to further strengthen the hypothesis of a photochemical origin of life's homochirality. Most relevant astronomical circularly polarized light (CPL) sources are reflection nebulae in high-mass star forming regions that show high degrees of circular polarization due to dichroic extinction of linearly polarized scattered light^{45,46}. All CPL sources so far exhibit a quadrupolar pattern of left-handed and right-handed CPL which is explained by the scattering of the bipolar outflow lobes where the right-handed or left-handed CP regions are situated at the opposite sides of the outflow axis as well as at the opposite sides of the central illuminating source. Each quadrant of single-handed CPL extends on very large spatial scales of up to 0.65 pc⁴⁶, which is hundreds of times the size of most planetary-forming systems, including our solar system. Assuming our solar system formed in a similar high-mass star-forming region, all organic molecules in the gas phase or condensed on icy grains would have been illuminated by CPL of one helicity only during protoplanetary disk evolution^{48,49}.

Anisotropy spectroscopy of chiral molecules is a direct measure of the selectivity and efficiency of circularly polarized light to induce optical activity and enrich one enantiomer over the other via asymmetric photolysis²¹. Corresponding experiments to understand the transfer of asymmetry from chiral photons to racemic starting material demonstrated that CPL can induce asymmetric photochemistry in alanine⁵⁰ and proline¹⁷, as well as being capable of asymmetric photochirogenesis of several amino acids in interstellar analog ices^{48,49}. Due to an interstellar dust cycling process⁴³, by which organic molecules such as amino acids can reversibly and continuously desorb and condense back

onto the surface of cometary and pre-cometary dust particles, chiral photon-molecule interactions are assumed to have—at least partly—taken place in the gas phase of interstellar environments. Such desorption–condensation cycles may have been crucial for an *ee* amplification via asymmetric photolysis of interstellar ices intimately connected with the surrounding gas, i.e., starting from 10^{-3} g values, as reported here and 10^{-3} to 10^{-2} g values in the solid state²¹, to a few %*ee* as identified in carbonaceous chondrites. In such a scenario, the origin and evolution of biomolecular asymmetry would be considered as a two-step process, with an initial mirror-symmetry breaking event in the life cycle of interstellar gas and dust during cloud evolution followed by nonequilibrium reaction networks on the early Earth driving prebiotic molecular systems toward a homochiral state^{51,52}. Our gas phase chiroptical data on neutral amino acids therefore complement previous CPL-based investigations on zwitterionic amino acids in the solid state⁵⁰ and the absolute asymmetric photosynthesis of amino acids in interstellar analog ices^{48,49}.

Stellar CPL, which is polychromatic, exhibits a dominant UV emission⁴⁵ matching the wavelength range of the anisotropy extrema of the investigated amino acids studied here. Given the decline in UV photon flux of most stars in the far UV range (<200 nm)^{45,53}, the broad anisotropy $n_O \rightarrow \pi^*_{CO}$ band (220–250 nm) would dictate the inducible *ee* in almost all gas phase amino acids, except valine, where the $n_N \rightarrow 3s$ band (190–220 nm) would dominate the interstellar asymmetric photolysis processes. Because stellar CPL would induce a net *ee* of the same handedness in almost all amino acids, such gas phase asymmetric photochemical reactions might have originally contributed to the evolution towards an enantiomeric selection of life's L-amino acids. The gas phase chiroptical and absorption spectra reported here are moreover crucial to trigger the UV-spectroscopic identification of chiral molecules in interstellar and circumstellar space, including their enantiomeric excess determination.

Methods

Gas-phase circular dichroism and anisotropy measurements. Circular dichroism spectra CD(λ), anisotropy spectra $g(\lambda)$, and absorption spectra $A(\lambda)$ of seven pairs of amino acid enantiomers were recorded using a temperature and pressure-controlled gas cell at the ASTRID2 synchrotron radiation source, Aarhus University, Denmark (Supplementary Methods). The cell consists of a stainless-steel tube, 500 mm in length with an 18 mm inner diameter, mounted on the AU-CD beam line. The pressure inside the gas cell is monitored using a type 631D Baratron heated manometer (MKS instruments), while the cell is evacuated using a Varian V70 turbo pump. Two all-metal valves are mounted on the main gas cell tube: one connects, via a flexible hose, to the turbo pump, while the other valve connects to the heated sample reservoir. Heating tapes and cartridge heaters were used to heat the entire gas cell up to the required temperature using PID computer control, with a maximum possible temperature of 200 °C. Differential heating of the windows to 5–10 °C higher temperatures compared to the tube avoids local condensation of analytes.

Typically, less than 1 g of a given amino acid enantiomer was placed into a glass container ($\varnothing$ 12 mm, length 15 mm) and inserted into the sample reservoir heater. Absorbance and CD spectra of each amino acid were recorded starting at $T = 160$ °C with gradually increased temperatures until sufficient vapor pressures with optimal optical densities were reached. The gas cell was typically operated under flow conditions, where the valves to the sample and the turbo pump were partially open, to ensure a continuous renewal of the gas in the cell.

Gas chromatographic measurements. The GC $\times$ GC-TOF-MS Pegasus IV D system from operated at a storage rate of 100 Hz, with a 50–400 amu mass range and a detector voltage of 1.45 kV (Supplementary Methods). Data were processed using the LECO Corp. ChromaTOFTM software. The source and transfer temperature were kept at 230 and 240 °C, respectively. The column set consisted of a Chirasil-L-Val column (25 m $\times$ 0.25 mm inner diameter, 0.12 mm film thickness) in the first-dimension modulator-coupled to a DB Wax secondary column (1.5 m $\times$ 0.1 mm inner diameter, 0.1 mm film thickness). Helium was used as carrier gas at a constant flow of $\dot{u} = 1$ mL min⁻¹. Sample volumes (1 μ L) were injected in splitless mode at an injector temperature of 230 °C. The GC primary oven was operated as follows: 40 °C (1 min), warm up to 80 °C (10 min) at 10 °C min⁻¹, and warm up to 190 °C (15 min) at 4 °C min⁻¹. The secondary oven used the same temperature program with a constant temperature offset of 20 °C. A modulation period of $P_M = 5$ s was applied.

Liquid chromatography measurements. The HPLC Agilent 1200 system was composed of a quaternary pump (Agilent G1311 A) equipped with a Diode Array Detector (DAD Agilent G1315D) and an ELSD (Evaporative Light Scattering Detector). A Luna C18 column (Phenomenex, 150 × 4.6 mm, 5 µm) was used at 25 °C with an injection volume set at 10 µL and a flow rate set at 1.0 mL min⁻¹. The HPLC was used in isocratic mode with 95% chromatography grade water (A) (acidified with 1% formic acid) and 5% methanol (B) during 40 min.

Computational details. A thorough conformational search was achieved by a biased molecular dynamics simulation following an Umbrella Sampling protocol. A single alanine residue was described in the neutral state using the ff14SB force field. All 22 selected alanine structures were refined with Gaussian16 using a combination of different methods and a large basis set quadruple- ζ with polarization and diffuse functions: aug-cc-pvQZ. A single-point energy calculation with zero-point corrected energies including frequency calculations was done to verify their nature of true minima. Three methods were evaluated: CAM-B3LYP, ω B97X-D, and M06-2X. Finally, seven low energy conformers were further considered for calculations. TD-DFT calculations were done on the seven optimized structures using the three previous functionals combined with the aug-cc-pvQZ basis set. Two hundred excited states were calculated for each optimized conformer.

For the coupled cluster calculations, Csaszar's²⁷ study of alanine conformers was used as a starting point, from which all reported 13 local minimum structures were obtained and subsequently Boltzmann-weighted. A total of nine alanine conformers were included in the computation of the CD spectrum. All 13 conformer structures were energy optimized with the program package Molpro on the df-CCSD(T)-F12 level with the aug-cc-pVDZ-F12 basis set, specifically with the F12b method and the 3* ζ (FIX, HY1) ansatz as implemented in Molpro and their cartesian coordinates in Å are provided in the Supplementary Data 1. Harmonic vibrational wavenumbers were calculated on the same level (Supplementary Data 2). Relative energies were determined by single point calculation with CCSD(T)-F12 and aug-cc-pVTZ-F12 basis set on the structures optimized with the double zeta basis set. Electronic excitation energies and intensities of the CD and the one-photon absorption spectra were calculated with the program package Turbomole with the RI-CC2 method and the aug-cc-pVQZ basis set.

Data availability

The authors declare that all data supporting this study are available within the main text and Supplementary Information file. A detailed experimental material and method section as well as Supplementary Figs. 1–15 and Tables 1–3 can be found in the Supplementary Information file. Raw data of the individual circular dichroism and anisotropy data are provided in the Source data file. Source data are provided with this paper.

Received: 29 July 2021; Accepted: 13 January 2022;

Published online: 26 January 2022

References

- Fersht, A. R. et al. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**, 235–238 (1985).
- Chen, Y. & Wallace, B. A. Secondary solvent effects on the circular dichroism spectra of polypeptides in non-aqueous environments: influence of polarisation effects on the far ultraviolet spectra of alamethicin. *Biophys. Chem.* **65**, 65–74 (1997).
- Donohue, J. The crystal structure of DL-alanine. II. Revision of parameters by three-dimensional Fourier analysis. *J. Am. Chem. Soc.* **72**, 949–953 (1950).
- Junk, G. & Svec, H. The mass spectra of the α -amino acids. *J. Am. Chem. Soc.* **85**, 839–845 (1963).
- Locke, M. J. & McIver, R. T. Effect of solvation on the acid/base properties of glycine. *J. Am. Chem. Soc.* **105**, 4226–4232 (1983).
- Chapo, L. J., Paul, J. B., Provencal, R. A., Roth, K. & Saykally, R. J. Is arginine zwitterionic or neutral in the gas phase? Results from IR Cavity Ringdown Spectroscopy. *J. Am. Chem. Soc.* **120**, 12956–12957 (1998).
- Martinez, S. J. III, Alfano, J. C. & Levy, D. H. The electronic spectroscopy of the amino acids tyrosine and phenylalanine in a supersonic jet. *J. Mol. Spectrosc.* **156**, 421–430 (1992).
- Snoek, L. C., Robertson, E. G., Kroemer, R. T. & Simons, J. P. Conformational landscapes in amino acids: infrared and ultraviolet ion-dip spectroscopy of phenylalanine in the gas phase. *Chem. Phys. Lett.* **321**, 49–56 (2000).
- Bakker, J. M., Aleese, L. M., Meijer, G. & von Helden, G. Fingerprint IR spectroscopy to probe amino acid conformations in the gas phase. *Phys. Rev. Lett.* **91**, 203003 (2003).
- Oomens, J., Steill, J. D. & Redlich, B. Gas-phase IR spectroscopy of deprotonated amino acids. *J. Am. Chem. Soc.* **131**, 4310–4319 (2003).
- Hong, A. et al. Electronic circular dichroism spectroscopy of jet-cooled phenylalanine and its hydrated clusters. *J. Phys. Chem. Lett.* **7**, 4385–4390 (2016).
- Suenram, R. D. & Lovas, F. J. Millimeter wave spectrum of glycine. *J. Mol. Spectrosc.* **72**, 372–382 (1978).
- Brown, R. D., Godfrey, P. D., Storey, J. W. V. & Bassez, M. -P. Microwave spectrum and conformation of glycine. *J. Chem. Soc. Chem. Commun.* **13**, 547–548 (1978).
- Godfrey, P. D., Firth, S., Hatherley, L. D., Brown, R. D. & Pierlot, A. P. Millimeter-wave spectroscopy of biomolecules: alanine. *J. Am. Chem. Soc.* **115**, 9687–9691 (1993).
- Blanco, S., Lesarri, A., Lopez, J. C. & Alonso, J. L. The gas-phase structure of alanine. *J. Am. Chem. Soc.* **126**, 11675–11683 (2004).
- Tia, M. et al. VUV photodynamics and chiral asymmetry in the photoionization of gas phase alanine enantiomers. *J. Phys. Chem. A* **118**, 2765–2779 (2014).
- Hadidi, R. et al. Conformer-dependent VUV photodynamics and chiral asymmetries in pure enantiomers of gas phase proline. *Commun. Chem.* **4**, 72 (2021).
- Müller, T., Wiberg, K. B. & Vaccaro, P. H. Cavity ring-down polarimetry (CRDP): a new scheme for probing circular birefringence and circular dichroism in the gas phase. *J. Phys. Chem. A* **104**, 5959–5968 (2000).
- Lahiri, P., Wiberg, K. B. & Vaccaro, P. H. Intrinsic optical activity and conformational flexibility: the role of size-dependent ring morphology in model cycloketones. *J. Phys. Chem. A* **117**, 12382–12400 (2013).
- Prümper, G. et al. Upper limits for stereoselective photodissociation of free amino acids in the vacuum ultraviolet region and at the C1s edge. *Phys. Rev. A* **69**, 062717 (2004).
- Meinert, C. et al. Anisotropy spectra of amino acids. *Angew. Chem. Int. Ed.* **51**, 4484–4487 (2012).
- Snyder, P. A., Vipond, P. M. & Johnson, W. C. Jr Circular dichroism of the alkyl amino acids in the vacuum ultraviolet. *Biopolymers* **12**, 975–992 (1973).
- Bockova, J., Jones, N. C., Meierhenrich, U. J., Hoffmann, S. V. & Meinert, C. Chiroptical activity of hydroxycarboxylic acids: implications for the origin of biological homochirality. *Commun. Chem.* **4**, 86 (2021).
- Meierhenrich, U. J. et al. Circular dichroism of amino acids in the vacuum-ultraviolet region. *Angew. Chem. Int. Ed.* **49**, 7799–7802 (2010).
- Fukuyama, T., Matsuo, K. & Gekko, K. Vacuum-ultraviolet electronic circular dichroism of L-alanine in aqueous solution investigated by time dependent density functional theory. *J. Phys. Chem. A* **109**, 6928–6933 (2005).
- Cao, M., Newton, S. Q., Pranata, J. & Schaefer, L. Ab initio conformational analysis of alanine. *J. Mol. Struct.* **332**, 251–267 (1995).
- Csaszar, A. Conformers of gaseous α -alanine. *J. Phys. Chem.* **100**, 3541–3551 (1996).
- Pescitelli, G. & Bruhn, T. Good computational practice in the assignment of absolute configurations by TDDFT calculations of ECD spectra. *Chirality* **28**, 466–474 (2016).
- Iijima, K. & Nakano, M. Reinvestigation of molecular structure and conformation of gaseous L-alanine by joint analysis using electron diffraction data and rotational constants. *J. Mol. Struct.* **486**, 255–260 (1999).
- Balabin, R. M. The identification of the two missing conformers of gas-phase alanine: a jet-cooled Raman spectroscopy study. *Phys. Chem. Chem. Phys.* **12**, 11675–11683 (2010).
- Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function* (W. H. Freeman, 1980).
- Jones, N. C., Nielsen, S. B. & Hoffmann, S. V. On the delocalization length in RNA single strands of cytosine: how many bases see the light? *PCCP* **22**, 2188–2192 (2020).
- Burin, A. L., Armbruster, M. E., Hariharan, M. & Lewis, F. D. Sum rules and determination of exciton coupling using absorption and circular dichroism spectra of biological polymers. *Proc. Natl Acad. Sci. USA* **106**, 989–994 (2009).
- Osted, A., Kongsted, J. & Christiansen, O. Theoretical study of the electronic gas-phase spectrum of glycine, alanine, and related amines and carboxylic acids. *J. Phys. Chem. A* **109**, 1430–1440 (2005).
- Meierhenrich, U. J. *Amino Acids and the Asymmetry of Life* (Springer, 2008).
- Cronin, J. R. & Pizzarello, S. Enantiomeric excesses in meteoritic amino acids. *Science* **275**, 951–955 (1997).
- Glavin, D. P. & Dworkin, J. P. Enrichment of the amino acid L-isovaline by aqueous alteration on CI and CM meteorite parent bodies. *Proc. Natl Acad. Sci. USA* **106**, 5487–5492 (2009).
- Muñoz Caro, G. M. et al. Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* **416**, 403–406 (2002).
- De Marcellus, P. et al. Aldehydes and sugars from evolved precometary ice analogs: importance of ices in astrochemical and prebiotic evolution. *Proc. Natl Acad. Sci. USA* **112**, 965–970 (2015).
- Meinert, C. et al. Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science* **352**, 208–212 (2016).

41. Goesmann, F. et al. Organic compounds on comet 67P/Churyumov-Gerasimenko revealed by COSAC mass spectrometry. *Science* **439**, aab0689 (2015).
42. Wright, I. P. et al. CHO-bearing organic compounds at the surface of 67P/Churyumov-Gerasimenko revealed by Ptolemy. *Science* **439**, aab0673 (2015).
43. Greenberg, J. M. *Cosmic Rays, Supernovae and the Interstellar Medium* 57–68 (Springer, 1991).
44. Altwegg, K. et al. Prebiotic chemicals—amino acid and phosphorus—in the coma of comet 67P/Churyumov-Gerasimenko. *Sci. Adv.* **2**, e1600285 (2016).
45. Bailey, J. et al. Circular polarization in star-formation regions: implications for biomolecular homochirality. *Science* **281**, 672–674 (1998).
46. Kwon, J. et al. Near-infrared polarimetry of the outflow source AFGL 6366S: Detection of circular polarization. *Astrophys. J.* **156**, 1 (2018).
47. Forbes, J. C., Alves, J. & Lin, D. N. C. A solar system formation analogue in the Ophiuchus star-forming complex. *Nat. Astron.* **5**, 1009–1016 (2021).
48. De Marcellus, P. et al. Non-racemic amino acid production by UV irradiation of achiral interstellar ice analogs with circularly polarized light. *Astrophys. J. Lett.* **727**, L27 (2011).
49. Modica, P. et al. Enantiomeric excesses induced in amino acids by ultraviolet circularly polarized light irradiation of extraterrestrial ice analogs: a possible source of asymmetry for prebiotic chemistry. *Astrophys. J.* **788**, 79 (2014).
50. Meinert, C. et al. Photonenergy-controlled symmetry breaking with circularly polarized light. *Angew. Chem. Int. Ed.* **53**, 210–214 (2014).
51. Frank, F. C. On spontaneous asymmetric synthesis. *Biochim. Biophys. Acta* **11**, 459–463 (1953).
52. Laurent, G., Lacoste, D. & Gaspard, P. Emergence of homochirality in large molecular systems. *Proc. Natl Acad. Sci. USA* **118**, e2012741118 (2021).
53. Bianchi, L., Conti, A. & Shiao, B. The ultraviolet sky: an overview from the GALEX surveys. *Adv. Space Res.* **53**, 900–912 (2014).

Acknowledgements

This work was funded by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement 804144, C.M.). Further funding was provided by the French government through the UCA^{IED} Investments in the Future project managed by the National Research Agency (ANR), with reference number ANR-15-IDEX-01 (C.M.), and by the ANR under grant number ANR-18-CE29-0004-01 (U.J.M.), as well as by the project CALIPSOplus, under Grant Agreement 730872 (S.V.H., N.C.J., and C.M.) from the EU Framework Program for Research and Innovation HORIZON 2020. Financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project number 328961117—SFB 1319 ELCH (M.D. and R.B.) is acknowledged. A.D.G. is grateful for a PhD scholarship from the French Ministry of Science and Education. Computation for the work described in this paper was supported by the Université Côte d'Azur's Center for High-Performance Computing. We thank Dr. Jana Bocková for supporting the alanine sublimation experiments as well as Dr. Thomas Michel for the liquid chromatographic analyses.

Author contributions

U.J.M., C.M., L.N. and S.V.H. conceived and designed the experiments. C.M., A.D.G., J.T., N.C.J. and S.V.H. performed the spectroscopic investigations, J.T., M.D. and R.B. the calculations. C.M. assembled and wrote the Supplementary Information with contributions from J.T., N.C.J. and S.V.H. A.D.G. and J.T. contributed equally to the experiments and are listed alphabetically. C.M. wrote the manuscript with the contribution of all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-28184-0>.

Correspondence and requests for materials should be addressed to Cornelia Meinert or Uwe J. Meierhenrich.

Peer review information *Nature Communications* thanks Francesco Aquilante, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

SPACE SCIENCES

Chiroptical activity of gas-phase propylene oxide predicting the handedness of interstellar circular polarization in the presolar nebula

Adrien D. Garcia^{1†}, Jérémie Topin^{1†}, Jana Bocková¹, Nykola C. Jones², Søren V. Hoffmann², Cornelia Meinert^{1*}

Propylene oxide, the first chiral molecule recently detected in the interstellar medium, has once again raised the question whether biomolecular chirality might have cosmic origins. However, accurate chiroptical properties of propylene oxide in the ultraviolet spectral range necessary to suggest possible asymmetric synthetic routes in the gas phase are scarce. Here, we report on the first experimental measurements of the anisotropy spectra of gas-phase propylene oxide in the vacuum ultraviolet spectral range. Our experimental results provide novel insights into the handedness of interstellar circular polarization at the dawn of molecular evolution of our star- and planet-forming region. Besides the fundamental importance of this new investigation for understanding the origin and evolution of homochirality on Earth, our high-resolution experimental electronic circular dichroism data will inspire new efforts in quantum computational spectroscopy.

INTRODUCTION

Propylene oxide, also known as 2-methyl oxirane, has recently regained considerable attention being the first chiral molecule detected outside of our Solar System (1). Chiral molecules are extremely important in chemistry, physics, biology, and drug discovery (2–8). Despite their variable stereoselectivity in various chemical reactions, the homochiral nature of all chiral biomolecules is a decisive prerequisite for an enormous range of cellular functions (9). While homochirality is evolutionarily advantageous, it is still unknown how the molecular single-handedness of the building blocks of the complex trinity—DNA, RNA, and proteins—came about (10). Several asymmetric processes have been experimentally tested to induce chirality in molecular systems (11–20), but those focusing on circularly polarized light (CPL) appear to us to be the most encouraging, especially given the results reported on CPL-induced molecular chirality in amino acids (21–27). Chirality transfer from photons to matter and vice versa (28–30) is not only of fundamental importance to understanding one of the most puzzling questions of the early origins of life (31, 32) but also of practical importance as a preparative method for synthesis of chiral compounds (33–36) or next-generation photonic devices (37–39).

Experiments that use CPL illumination to induce a chiral bias require prior knowledge of the chiroptical response of the targeted matter to tune desired enantiomeric excesses (*ee*). Circular dichroism (CD) spectroscopy, based on the difference in absorption of left and right CPL (*l*- and *r*-CPL) of chiral molecules, remains the preferred technique to study chiral discrimination, because of its straightforward operation, high accuracy, and efficiency (40). The intrinsic enantioselectivity toward CPL of a given handedness and wavelength can be then quantified by the so-called anisotropy or

dissymmetry factor $g = \Delta\epsilon/\epsilon$, the ratio between the differential extinction coefficient $\Delta\epsilon$, and the extinction coefficient ϵ (41–43).

Propylene oxide (Fig. 1A), a relatively simple chiral molecule, has recently been detected in the interstellar medium toward the high-mass star-forming region Sagittarius B2(N) (1). At the time, the observations did not allow the determination of whether there is a difference between the relative abundance of the two propylene oxide enantiomers. However, high-precision polarization measurements could provide a definite answer in the future.

The spectroscopic properties of propylene oxide and its derivatives have been extensively studied in the past (44–48). Because of the minimal conformational freedom of propylene oxide, it is, moreover, an ideal model compound for theoretical method development for chiroptical spectroscopy (49–51). However, the only low-resolution experimental CD spectra of propylene oxide previously reported in the ultraviolet (UV) (46) and vacuum UV (VUV) wavelength range (44, 45) show only very weak and inconsistent features for the higher-lying electronic states with no mirror symmetry confirmation of relative intensities, signs, and band positions of the two enantiomers. These experimental data are thus inadequate for validation of theoretical approaches. We have recently refreshed the field of gas-phase CD and anisotropy spectroscopy with a new experimental setup that combines an entirely new gas cell with synchrotron radiation from the AU-CD beam line at ASTRID2, Aarhus (Denmark) (52).

Here, we report on the first experimental measurements of the anisotropy spectra of *R*-(+)- and *S*-(-)-propylene oxide along with synchrotron CD and absorption spectra in the VUV spectral region. The concerted experimental and in vacuo theoretical experiments ultimately provide structural information on the electronic and vibronic transitions as well as the oscillator and optical rotatory strengths. Quantitative models based on these newly measured anisotropy spectra revealed correlations between the *R*-(+)-enantiomer-enriched propylene oxide derivative detected in the Murchison meteorite (53) and the helicity of CPL in our presolar

¹Institut de Chimie de Nice, Université Côte d'Azur, UMR 7272 CNRS, Nice 06108, France. ²ISA, Department of Physics and Astronomy, Aarhus University, 8000 Aarhus, Denmark.

†These authors contributed equally to this work.

*Corresponding author. Email: cornelia.meinert@univ-cotedazur.fr

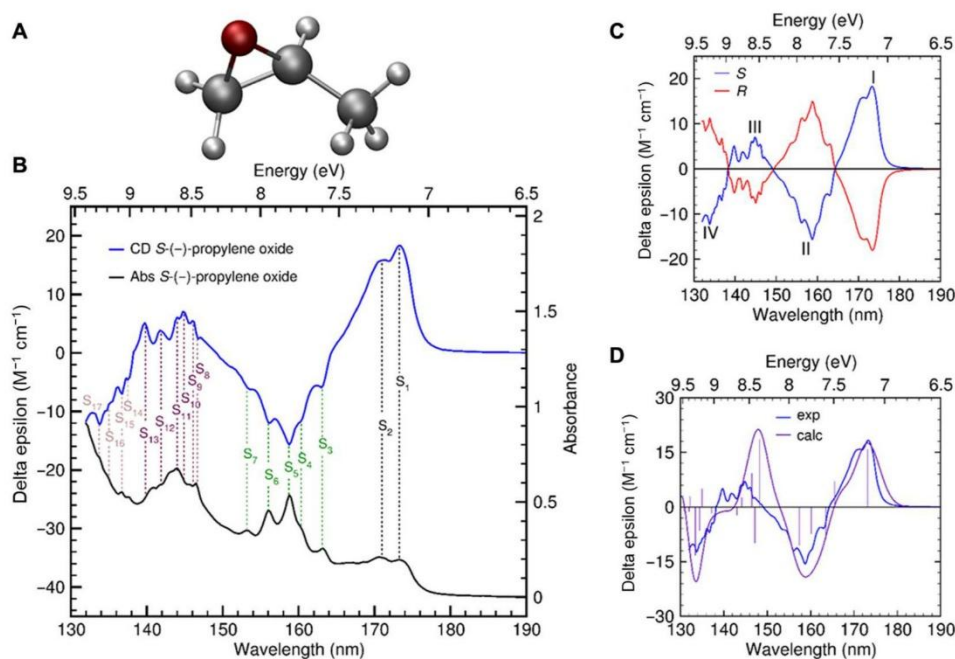


Fig. 1. Gas-phase chiroptical properties of propylene oxide. (A) Optimized structure of *S*-(–)-propylene oxide (CAM-B3LYP/aug-cc-pVQZ). (B) Experimental CD (blue) and absorbance (black) spectra of *S*-(–)-propylene oxide. (C) Experimental CD spectra of *R*-(+)-enantiomer (red) and *S*-(–)-enantiomer (blue) of propylene oxide exhibiting mirror symmetry. (D) Experimental (blue line) and theoretical CD spectrum (purple line) of *S*-(–)-propylene oxide based on vertical excitations (TD-DFT/CAM-B3LYP and aug-cc-pVQZ); FWHM, 0.12 eV; rotatory strengths are indicated as stick bars.

cloud with far-reaching consequences in understanding the origins of biomolecular homochirality.

RESULTS

Gas-phase chiroptical spectra

The anisotropy factor *g* as a function of wavelength was studied by recording simultaneously the CD $\Delta\epsilon$ and absorbance ϵ spectra of each enantiomer of propylene oxide using a gas of analytes in a dedicated sample cell with a length of 500 mm at the AU-CD beam line on ASTRID2 at Aarhus University (52). Figure 1B shows the CD and associated absorption spectra of *S*-(–)-propylene oxide along with the most dominant resolved experimental electronic and vibronic transitions highlighting the high resolution of gas-phase synchrotron CD spectroscopy. To provide accurate anisotropy data of both enantiomers, the CD (fig. S1) and absorbance spectra (fig. S2 and table S1) were measured at various gas flow rates. Individually recorded CD spectra could be scaled directly with the cell pressure during measurements to yield spectra of similar shape and magnitude. The absolute CD signal, corrected for gas density and path length, is given in units of molar CD (per molar per centimeter) often referred to as Delta epsilon units as used in Fig. 1.

CD spectroscopy of both enantiomers of propylene oxide yielded two almost perfectly mirrored spectral curves (Fig. 1C), revealing the best resolved vibrational fine structure to date (fig. S1 and table S2). *S*-(–)-propylene oxide displays the most pronounced maxima at around $\lambda = 173$ (CD band I) and 145 nm (III) and

minima at around 159 (II) and 134 nm (IV), while *R*-propylene oxide shows mirror symmetry in CD signs and relative intensities.

To compute the CD spectrum of *S*-(–)-propylene oxide, the first 16 time-dependent density functional theory (TD-DFT) vertical excitations were convoluted with Gaussian broadening functions with full width at half maximum (FWHM) of 0.12 eV. Vertical excitation energies and rotatory strengths *R* computed with different functionals and basis sets are summarized in table S4 and compared with former theoretical results (table S5). Best agreement with our experimental data was found using the CAM-B3LYP functional, which was tailored explicitly to treat Rydberg-type and diffuse excitations (54). In combination with the aug-cc-pVQZ basis set, the convergence for the low-lying computed excitation energies and rotatory strengths is reasonably good, while larger deviation occurs for CD bands III and IV. According to previous literature (44, 45, 50, 51, 55), the CD transitions were assigned as the Rydberg transitions from either *n*(O), the lone pair orbital on the oxygen atom, or the σ -type orbital localized along the oxirane ring bonds (table S5, A and B). The often mixed *s*- and *p*-characters of involved Rydberg-type orbitals in combination with the fine structure of our experimental gas-phase CD spectra attributable to vibrational modes—not considered in our theoretical model—make definite electronic assignments of the experimental excitation energies rather challenging.

Apart from a spectral shift of CD band III, the sign and spectral position of the calculated CD bands match those of our measured spectrum (Fig. 1D), further supporting the consistency of our data. The discrepancies in the energy range of 8 to 9 eV obviously

highlight the challenges in predicting CD spectra of oxiranes, particularly where the overall signal is a result of cancellation of vibronic transitions of opposite signs with distinct vibrational structures (50, 51, 56). Our first coherent high-resolution experimental absorbance and CD data of both propylene oxide enantiomers in the VUV spectral region, therefore, open a plethora of avenues for profound optimization of computational methods that consider vibrational modes to accurately simulate chiroptical properties.

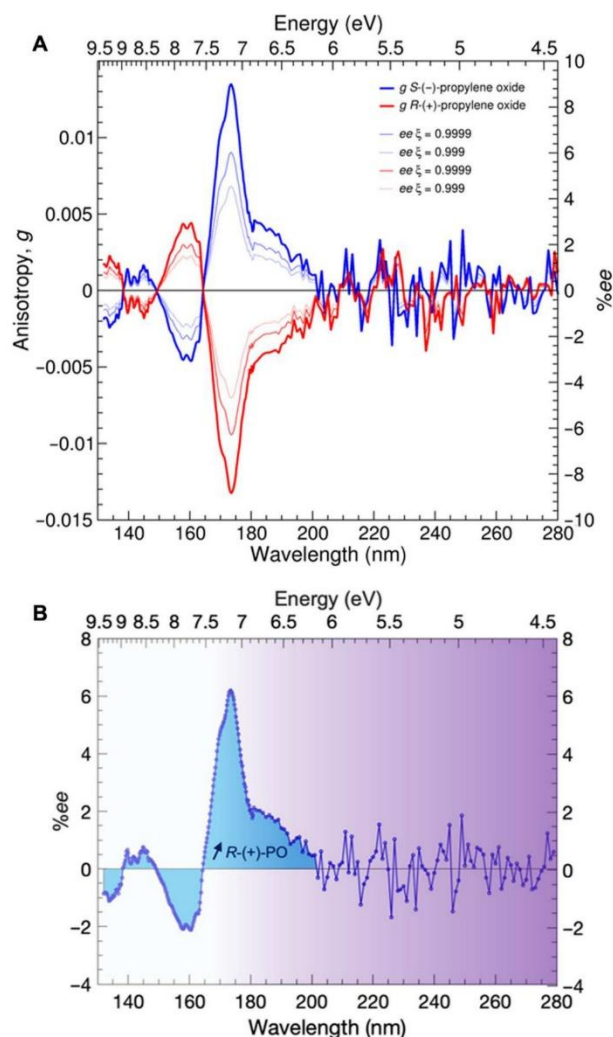


Fig. 2. Gas-phase anisotropy spectra of propylene oxide. (A) Anisotropy spectra of $R(+)$ -propylene oxide (red) and $S(-)$ -propylene oxide (blue). Thin lines indicate the lower limit of the inducible enantiomeric excess (ee) by either I -CPL or r -CPL as a function of wavelength and extent of reaction ξ calculated on the basis of the relation reported in (43). (B) Integrated area (blue) under the $\%ee$ curve obtained for $S(-)$ -propylene oxide with an extent of reaction $\xi = 0.9999$. Purple color and brightness indicate the decline in UV photon flux of low-mass main sequence stars (60). Stellar broadband I -CPL would induce an excess in R -propylene oxide (PO).

The intensity of the anisotropy bands follows the intensity pattern of the CD spectra. Band I is the most intense feature in both CD and anisotropy spectra, followed by bands II, IV, and III (Fig. 2A). Consequently, the highest $\%ee$ can be induced in the wavelength range of 165 to 200 nm with values predicted to be about 6% for a reaction rate of 0.9999 and $g_{174nm} = 1.3 \times 10^{-2}$ based on the relation reported in (43) and recapped in the Supplementary Text. The predicted ee curves over the entire anisotropy wavelength range are shown as thin lines in Fig. 2A for two selected photolysis rates ξ of 0.999 (99.90%) and 0.9999 (99.99%), respectively. Further anisotropy values and their calculated minimum inducible $\%ee$ by asymmetric photolysis at the extent of reaction 0.9999 at energies corresponding to the extrema of anisotropy bands I to IV are given in table S3.

Asymmetric photochemistry at the dawn of molecular cloud evolution

The enantiomeric enrichment of naturally occurring L-amino acids (57, 58) and D-sugar derivatives (59) found in carbonaceous meteorites offers the sole example of asymmetry outside of the biosphere and has raised the hypothesis of a possible link between astrochemical evolution and the origin of homochiral biomolecules. However, both the origin of this extraterrestrial chiral bias and the emergence of the first chiral molecule itself are still unanswered. Asymmetric photochemistry by stellar CPL is arguably the most compelling explanation behind the chiral bias found in meteorites. This idea gained support when CP radiation in the near-infrared was found in the active sites of star-forming regions of the Orion Nebula (60–62) and in NGC 6334-V (63). Considering that a primordial chiral bias originated in the interstellar medium and was at least partially incorporated into planetesimals and planets, extraterrestrial chiral molecules could have thus contributed to the chiral choice of homochiral life on Earth. Interstellar ice simulation experiments revealed a formation mechanism of racemic propylene oxide via galactic cosmic ray-driven nonequilibrium chemistry involving suprathermal oxygen at very low temperatures and its subsequent sublimation during temperature-programmed desorption at 100 K (64). While the predominant photodecomposition of gas-phase propylene oxide upon UV radiation is known to proceed via a ring opening followed by a 1,2-hydrogen shift to give propanal and to a minor extent acetone (65), our anisotropy data predict that chiral discrimination via asymmetric photolysis of racemic propylene oxide in the gas phase could lead to an enantiomeric enrichment when the photolysis is incomplete. No chiroptical data are now available for propylene oxide in the solid state, simulating the surface of interstellar dust grains. A derivative of propylene oxide has recently been detected in the ethanol extracts of the Murchison meteorite (53). The enantioselective analysis revealed a large ee_R of about 10%, i.e., an excess of the $R(+)$ -enantiomer.

Under realistic astrophysical conditions, natural sources of circularly polarized radiation are spectrally broadband in the pertinent UV spectral region (60–62), and photochemical activation of multiple CD/anisotropy bands of opposite signs is expected. To investigate the net effect of asymmetric photolysis of non-monochromatic CPL emission, integration of each wavelength interval corresponding to an anisotropy band and integration over the entire investigated wavelength range have been performed (Table 1 and fig. S3). The results show that broadband UV irradiation by stellar I -CPL with a constant flux in the measured

Table 1. Inducible ee values for propylene oxide. Integrated anisotropy g bands and inducible minimum enantiomeric excess ee_R through asymmetric photolysis of racemic propylene oxide (PO) at the extent of reaction $\xi = 0.9999$ and broadband UV I -CPL with uniform flux density in the UV range.

λ/nm	Average g ($\times 10^{-2}$)*	Average ee_R [†]	Average ee_R [‡]
164–202	0.51		2.4
149–164	−0.29		−1.4
138–149	0.08	0.21	0.4
132–138	−0.18		−0.8

*Average anisotropy $g(\lambda)$ of each active g band of S-PO. †Average anisotropy $g(\lambda)$ over the entire g spectra of S-PO. ‡Average inducible ee_R of each active g band. §Average inducible ee_R over the entire g spectra.

wavelength range would induce an ee in favor of the R -enantiomer of gas-phase propylene oxide of at least 1%. The calculated net ee_R over the range of 138 to 202 nm can be thought of as a lower limit considering the reduction of UV photon flux in the far UV range (60), resulting in a weaker impact of the g bands between 132 and 164 nm compared to the most intense g band spreading from 164 to 202 nm. The most pronounced anisotropy band I coinciding with the dominant UV emission wavelength range of most of the stars (60) would thus dictate the handedness of UV CPL inducible enantiomeric excess (Fig. 2B).

Although we cannot exclude the impact of asymmetric solid-state photolytic reactions on the surface of icy grains, which may differ due to the influence of a surrounding polar solvent on the chiroptical activity of propylene oxide, we assume that its overall impact on the total inducible ee is rather limited. The lowest energy anisotropy band attributed to the $n_O \rightarrow R(3s)$ transition is expected to be in the far UV even in the solid state where the ice matrix dominates the stellar CPL absorption with energies above 6.5 eV. Nevertheless, future studies of the influence of a surrounding water matrix on the excitation energy and the sign of CD band I recorded in the gas phase between 7 and 7.5 eV by a surrounding water matrix are envisaged.

DISCUSSION

If our Solar System was formed in a high-mass star-forming region, which is supported by the presence of short half-life radionuclides such as ^{60}Fe and ^{26}Al in primitive meteorites (66)—according to our anisotropy-based prediction—it may have been irradiated with left-handed CPL to result in an excess of R -(+)-propylene

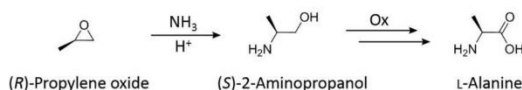


Fig. 3. Gas synthetic pathway from R -(+)-propylene oxide to L-alanine. First step: Epoxide opening in acidic condition. Second step: Over-oxidation of S -(+)-2-aminopropanol.

oxide and its derivatives in the Murchison meteorite. While our current prediction of the inducible ee_R in propylene oxide should be considered as the lower estimate due to the nonlinear decline of the stellar CPL in the VUV (60), desorption-condensation cycles connecting the solid interstellar ices with the surrounding gas may have been crucial for an ee amplification of interstellar propylene oxide starting from a few percent, as shown here in the gas phase, to about 10% ee_R as identified in the Murchison meteorite (53). Besides, it is a common view that not only meteoritic amino acids but also their chiral bias may have originated from their precursor structures via Strecker-type synthesis during parent body aqueous alteration (53). Note that R -(+)-propylene oxide can, in two steps, lead to the proteinogenic amino acid L-alanine. The nucleophilic attack of ammonia (NH_3) on the stereogenic carbon of the R -(+)-propylene oxide under acidic conditions leads to S -(+)-2-aminopropanol, the reduced form of the proteinogenic amino acid L-alanine (Fig. 3). Given the evidence of ammonia in the interstellar medium, this reaction could have occurred on interstellar icy dust particles, and the resulting amino alcohols could have been subsequently embedded in comets and asteroids. During aqueous alteration of the parent body, an oxidation process might have allowed the formation of L-alanine. Thereby, the discovery of an enantiomeric excess of the R -(+)-enantiomer of a propylene oxide derivative (53) could be related to the reported ee of the L-enantiomer of some meteoritic amino acids (67, 68).

Besides the substantial improvement of our CD and anisotropy spectra in terms of resolution and mirroring effect down to the far UV region, we believe that our data will stimulate future studies on establishing a unified view of the CPL scenario considering (i) the chiroptical properties of key biomolecules, (ii) the hypothesis that our Solar System encountered I -CPL radiation at the time of its formation, and (iii) the detection of chiral molecules in meteorites and samples returned from the asteroid sample-return missions Hayabusa2 and OSIRIS-REx.

MATERIALS AND METHODS

Chemicals

Enantiopure standards of R -(+)-propylene oxide ($\geq 98.5\%$) and S -(−)-propylene oxide (not specified) were purchased from Sigma-Aldrich (Germany). Both standards were of high optical purity (99% ee) and were used without any further purification. R -(+)-camphor (98%) used for calibration purposes of the gas cell before recording CD spectra was purchased from Sigma-Aldrich (Germany).

Experimental design

Gas-phase absorption spectra, electronic CD, and anisotropy spectra of individual enantiomers were recorded at the AU-CD beam line on ASTRID2 at Aarhus University (Denmark) (69, 70). A temperature- and pressure-controlled gas cell has been built for this purpose and is described in detail elsewhere (52). The cell is made of stainless steel with an optical path length of 500 mm and an inner diameter of 18 mm. The vapor pressure of the samples was monitored by a type 631D Baratron heated manometer (MKS Instruments). CaF_2 windows used at the extremities of the gas cell connected the apparatus with the photoelastic modulator to produce the CPL at one end and a photomultiplier tube detector at the other. Heating tapes and cartridge heaters allow heating of

the entire gas cell up to 200°C, and we can monitor and locally control the temperature at nine different points of the gas cell assembly in case of low vapor pressure from a sample. This was needed for earlier measurements on amino acids (2), but the vapor pressure of propylene oxide was high enough to allow operation of the gas cell at room temperature. However, differential heating of the windows to higher temperatures compared to the main body of the gas cell avoided local condensation of analytes. Therefore, throughout all measurements, the temperature of the CaF₂ windows was set to 30°C. Two all-metal valves are mounted on the main gas cell tube: one connects, via a flexible hose, to a Varian V70 turbo pump, while the other valve connects to the sample reservoir, a 25-ml single-neck round-bottom flask, connected at the center of the gas cell. This allows the pumping speed and sample inlet rate to the gas cell to be adjusted. Measured pressures increased rapidly above 0.8 mbar without sample flow, while it was kept in between 0.03 and 0.16 mbar for *R*-(+)- and *S*-(-)-propylene oxide during absorption and CD measurements.

Gas-phase synchrotron radiation CD and anisotropy spectroscopy

The operation of the experimental set up was confirmed through measurement of *R*-(+)-camphor, for which the CD spectrum in the gas phase with a characteristic transition at about 300 nm is known (71). The absorption ϵ and CD spectra $\Delta\epsilon$ of each propylene oxide enantiomer were recorded simultaneously in the gas phase. The gas cell was operated under flow conditions, where the valves to the sample and the turbo pump were partially open, to ensure a continuous renewal of the gas in the cell. This was done to avoid a build-up of any photodecomposition products. We were not able to maintain constant pressure under continuous flow conditions of each enantiomer but observed a steady increase of the pressure throughout the sample scans. Comparing different sample scans with each other, however, confirmed that the increase in pressure was due to the sample itself and not from other impurities such as water or decomposition products. The sample pressure was recorded for each data point in a spectrum, and all spectra are pressure-corrected.

The anisotropy spectra $g(\lambda) = \Delta\epsilon/\epsilon$ could be obtained as the absorbance of the gas, ϵ , was simultaneously measured with the CD. Absorbance is calculated using the regulated high voltage applied to photodetector and the gain value for detector, which has been determined through measurement on a series of samples on both the AU-CD beamline and a calibrated photospectrometer (Evolution 300, Thermo Scientific) (72). Approximately 10 ml of propylene oxide was added to a 25-ml pear-shaped glass container attached to the center of the gas cell by an all-metal valve. This amount of sample proved to be enough for multiple measurement runs. Before CD measurements, the sample was purged of dissolved gases by three freeze-pump-thaw cycles, and the valve was fully open to the cell to remove air or water vapor coming off the samples.

The gas-phase spectra were recorded from 4.4 to 9.5 eV (280 to 130 nm). Sample vapor pressures were chosen so that the optimal optical density near 0.8 providing the best signal-to-noise ratio was obtained. Multiple baseline scans of the empty gas cell were recorded before and after each propylene oxide enantiomer. Fortunately, the gas cell was designed to quickly be taken apart, cleaned, and reassembled as extensive white deposits within the cell and sample valve were observed. After extensive cleaning and remounting of

the sample inlet for the gas cell, we pumped down the cell for several hours before running new baselines and the opposite enantiomer.

The CD spectra of propylene oxide enantiomers could be scaled directly with the cell pressure during measurements, and the absolute CD signal (corrected for gas density and path length) could be calculated. Thus, the CD spectra shown are in units of molar CD (per molar per centimeter) often referred to as Delta epsilon units. The CD spectra were mildly smoothed with a seven-point Savitzky-Golay filter.

Computational details

The initial molecular structure of propylene oxide was built using GaussView (version 6). As the molecule exhibits no internal degrees of freedom, solely one conformer was considered. This structure was further optimized with Gaussian16, using a combination of different methods and different basis sets. We performed a single-point energy calculation with zero-point corrected energies including frequency calculations to verify the nature of true minima. Three methods previously applied for our amino acid gas-phase spectra were evaluated: CAM-B3LYP, ω B97X-D, and MO6-2X (52).

TD-DFT calculations were completed on the optimized structure using the three functionals combined with different basis sets. Two hundred excited states were calculated for the optimized conformer. Last, the combination of the range-separated functional CAM-B3LYP method with the quadruple- ζ with polarization and diffuse functions: aug-cc-pVQZ leads to the best fit between experiment and theory. To compare the theoretical data with the spectral shapes of the experiments, the theoretical electronic CD transitions have been convoluted by means of summing rotatory strength weighted Gaussian distribution functions with FWHM of 0.12 eV.

Supplementary Materials

This PDF file includes:

Supplementary Text
Figs. S1 to S3
Tables S1 to S5
References

REFERENCES AND NOTES

1. B. A. McGuire, P. B. Carroll, R. A. Loomis, I. A. Finneran, P. R. Jewell, A. J. Remijan, G. A. Blake, Discovery of the interstellar chiral molecule propylene oxide (CH₃CHCH₂O). *Science* **352**, 1449–1452 (2016).
2. J. R. Brandt, F. Salerno, M. J. Fuchter, The added value of small-molecule chirality in technological applications. *Nat. Rev. Chem.* **1**, 1–12 (2017).
3. F. Lancia, A. Ryabchun, N. Katsonis, Life-like motion driven by artificial molecular machines. *Nat. Rev. Chem.* **3**, 536–551 (2019).
4. T. Leigh, P. Fernandez-Trillo, Helical polymers for biological and medical applications. *Nat. Rev. Chem.* **4**, 291–310 (2020).
5. Z. Liu, H. Du, J. Li, L. Lu, Z.-Y. Li, N. X. Fang, Nano-kirigami with giant optical chirality. *Sci. Adv.* **4**, eaat4436 (2018).
6. G. Long, R. Sabatini, M. I. Saidaminov, G. Lakhwani, A. Rasmita, X. Liu, E. H. Sargent, W. Gao, Chiral-perovskite optoelectronics. *Nat. Rev. Mater.* **5**, 423–439 (2020).
7. R. Naaman, Y. Paltiel, D. H. Waldeck, Chiral molecules and the electron spin. *Nat. Rev. Chem.* **3**, 250–260 (2019).
8. G. Qu, A. Li, C. G. Acevedo-Rocha, Z. Sun, M. T. Reetz, The crucial role of methodology development in directed evolution of selective enzymes. *Angew. Chem. Int. Ed.* **59**, 13204–13231 (2020).
9. G. F. Joyce, G. M. Visser, C. A. A. van Boeckel, J. H. van Boom, L. E. Orgel, J. van Westrenen, Chiral selection in poly(C) directed synthesis of oligo(G). *Nature* **310**, 602–604 (1984).

10. W. A. Bonner, The origin and amplification of biomolecular chirality. *Orig. Life Evol. Biosph.* **21**, 59–111 (1991).
11. F. C. Frank, On spontaneous asymmetric synthesis. *Biochim. Biophys. Acta* **11**, 459–463 (1953).
12. M. Klussmann, T. Izumi, A. J. White, A. Armstrong, D. G. Blackmond, Emergence of solution-phase homochirality via crystal engineering of amino acids. *J. Am. Chem. Soc.* **129**, 7657–7660 (2007).
13. D. K. Kondepudi, R. J. Kaufman, N. Singh, Chiral symmetry breaking in sodium chlorate crystallization. *Science* **250**, 975–976 (1990).
14. M. Quack, How important is parity violation for molecular and biomolecular chirality? *Angew. Chem. Int. Ed.* **41**, 4618–4630 (2002).
15. J. B. Ribo, J. Crusats, F. Sagues, J. Claret, R. Rubires, Chiral sign induction by vortices during the formation of mesophases in stirred solutions. *Science* **292**, 2063–2066 (2001).
16. G. L. Rikken, E. Raupach, Enantioselective magnetochiral photochemistry. *Nature* **405**, 932–935 (2000).
17. R. A. Rosenberg, M. Abu Haila, P. J. Ryan, Chiral-selective chemistry induced by spin-polarized secondary electrons from a magnetic substrate. *Phys. Rev. Lett.* **101**, 178301 (2008).
18. K. Soai, S. Osanai, K. Kadowaki, S. Yonekubo, T. Shibata, I. Sato, *d*- and *l*-Quartz-promoted highly enantioselective synthesis of a chiral organic compound. *J. Am. Chem. Soc.* **121**, 11235–11236 (1999).
19. G. Tranter, Parity-violating energy differences of chiral minerals and the origin of biomolecular homochirality. *Nature* **318**, 172–173 (1985).
20. F. Vester, T. Ulbricht, H. Krauch, Optische Aktivität und die Paritätsverletzung im β -Zerfall. *Naturwissenschaften* **46**, 68–68 (1959).
21. P. De Marcellus, C. Meinert, M. Nuevo, J.-J. Filippi, G. Danger, D. Deboffe, L. Nahon, L. Le Sergeant d'Hendecourt, U. J. Meierhenrich, Non-racemic amino acid production by ultra-violet irradiation of achiral interstellar ice analogs with circularly polarized light. *Astrophys. J. Lett.* **727**, L27 (2011).
22. J. J. Flores, W. A. Bonner, G. A. Massey, Asymmetric photolysis of (RS)-leucine with circularly polarized ultraviolet light. *J. Am. Chem. Soc.* **99**, 3622–3625 (1977).
23. U. J. Meierhenrich, L. Nahon, C. Alcaraz, J. H. Bredehöft, S. V. Hoffmann, B. Barbier, A. Brack, Asymmetric vacuum UV photolysis of the amino acid leucine in the solid state. *Angew. Chem. Int. Ed.* **44**, 5630–5634 (2005).
24. C. Meinert, S. V. Hoffmann, P. Cassam-Chenai, A. C. Evans, C. Giri, L. Nahon, U. J. Meierhenrich, Photonenergy-controlled symmetry breaking with circularly polarized light. *Angew. Chem. Int. Ed.* **53**, 210–214 (2014).
25. H. Nishino, A. Kosaka, G. A. Hembury, F. Aoki, K. Miyauchi, H. Shitomi, H. Onuki, Y. Inoue, Absolute asymmetric photoreactions of aliphatic amino acids by circularly polarized synchrotron radiation: Critically pH dependent photo behavior. *J. Am. Chem. Soc.* **124**, 11618–11627 (2002).
26. W. L. Noorduin, A. A. C. Bode, M. van der Meijden, H. Meekes, A. F. van Etteger, W. J. P. van Enckevort, P. C. M. Christianen, B. Kaptein, R. M. Kellogg, T. Rasing, E. Vlieg, Complete chiral symmetry breaking of an amino acid derivative directed by circularly polarized light. *Nat. Chem.* **1**, 729–732 (2009).
27. B. Norden, Was photoreolution of amino acids the origin of optical activity in life? *Nature* **266**, 567–568 (1977).
28. J. A. L. Bel, Sur les relations qui existent entre les formules atomiques des corps organiques et le pouvoir rotatoire de leurs dissolutions. *Bull. Soc. Chim. Paris* **22**, 337–347 (1874).
29. N. P. Huck, W. F. Jager, B. De Lange, B. L. Feringa, Dynamic control and amplification of molecular chirality by circular polarized light. *Science* **273**, 1686–1688 (1996).
30. J.-Y. Kim, J. Yeom, H. Calcaterra, G. Zhao, P. Zhang, N. Kotov, Assembly of gold nanoparticles into chiral superstructures driven by circularly polarized light. *J. Am. Chem. Soc.* **141**, 11739–11744 (2019).
31. A. D. García, C. Meinert, H. Sugahara, N. C. Jones, S. V. Hoffmann, U. J. Meierhenrich, The astrophysical formation of asymmetric molecules and the emergence of a chiral bias. *Life* **9**, 29 (2019).
32. I. Myrberg, C. Meinert, Z. Martins, L. Le Sergeant d'Hendecourt, U. J. Meierhenrich, Molecular chirality in meteorites and interstellar ices, and the chirality experiment on board the ESA cometary Rosetta mission. *Angew. Chem. Int. Ed.* **54**, 1402–1412 (2015).
33. H. K. Bisoyi, Q. Li, Light-directed dynamic chirality inversion in functional self-organized helical superstructures. *Angew. Chem. Int. Ed.* **55**, 2994–3010 (2016).
34. C. He, G. Yang, Y. Kuai, S. Shan, L. Yang, J. Hu, D. Zhang, Q. Zhang, G. Zou, Dissymmetry enhancement in enantioselective synthesis of helical polydiacetylene by application of superchiral light. *Nat. Commun.* **9**, 5117 (2018).
35. H. Kagan, A. Moradpour, J. F. Nicoud, G. Balavoine, G. Tsoucaris, Photochemistry with circularly polarized light. Synthesis of optically active hexahelicene. *J. Am. Chem. Soc.* **93**, 2353–2354 (2002).
36. S. T. Wu, Z. W. Cai, Q. Y. Ye, C. H. Weng, X. H. Huang, X. L. Hu, C. C. Huang, N. F. Zhuang, Enantioselective synthesis of a chiral coordination polymer with circularly polarized visible laser. *Angew. Chem. Int. Ed.* **53**, 12860–12864 (2014).
37. J. Lu, Y. Xue, K. Bernardino, N. N. Zhang, W. R. Gomes, N. S. Ramesar, S. Liu, Z. Hu, T. Sun, A. F. de Moura, N. A. Kotov, K. Liu, Enhanced optical asymmetry in supramolecular chiro-plasmonic assemblies with long-range order. *Science* **371**, 1368–1374 (2021).
38. L. Wan, J. Wade, F. Salerno, O. Arteaga, B. Laidlaw, X. Wang, T. Penfold, M. J. Fuchter, A. J. Campbell, Inverting the handedness of circularly polarized luminescence from light-emitting polymers using film thickness. *ACS Nano* **13**, 8099–8105 (2019).
39. M. D. Ward, J. Wade, X. Shi, J. Nelson, A. J. Campbell, M. J. Fuchter, Highly selective high-speed circularly polarized photodiodes based on π -conjugated polymers. *Adv. Opt. Mater.* **10**, 2101044 (2021).
40. G. Pescitelli, L. Di Bari, N. Berova, Conformational aspects in the studies of organic compounds by electronic circular dichroism. *Chem. Soc. Rev.* **40**, 4603–4625 (2011).
41. W. Kuhn, E. Braun, Photochemische Erzeugung optisch aktiver Stoffe. *Naturwissenschaften* **17**, 227–228 (1929).
42. W. Kuhn, E. Knopf, Photochemische Erzeugung optisch aktiver Stoffe. *Naturwissenschaften* **18**, 183 (1930).
43. C. Meinert, J. H. Bredehöft, J. J. Filippi, Y. Baraud, L. Nahon, F. Wien, N. C. Jones, S. V. Hoffmann, U. J. Meierhenrich, Anisotropy spectra of amino acids. *Angew. Chem. Int. Ed.* **51**, 4484–4487 (2012).
44. A. Breest, P. Ochmann, F. Pulm, K. H. Gödderz, M. Carnell, J. Hormes, Experimental circular dichroism and VUV spectra of substituted oxiranes and thiranes. *Mol. Phys.* **82**, 539–551 (1994).
45. M. Carnell, Experimental and quantum theoretical investigation of the circular dichroism spectrum of R-methyloxirane. *Chem. Phys. Lett.* **180**, 477–481 (1991).
46. D. Cohen, M. Levi, H. Basch, A. Gedanken, Excited electronic states of optically active substituted ethylene oxides: (–)-(S)-2-methyloxirane and (–)-(S, S)-2, 3-dimethyloxirane. *J. Am. Chem. Soc.* **105**, 1738–1742 (1983).
47. J. Kukras, J. Kauczor, P. Norman, A. Rizzo, G. L. J. A. Rikken, S. Coriani, A complex-polarization-propagator protocol for magneto-chiral axial dichroism and birefringence dispersion. *Phys. Chem. Chem. Phys.* **18**, 13267–13279 (2016).
48. M. M. Rafiee Fanood, I. Powis, M. H. Janssen, Chiral asymmetry in the multiphoton ionization of methyloxirane using femtosecond electron-ion coincidence imaging. *J. Phys. Chem. A* **118**, 11541–11546 (2014).
49. L. Alagna, S. di Fonzo, T. Prosperi, S. Turchini, P. Lazzaretto, M. Malagoli, R. Zanasi, C. R. Natoli, P. J. Stephens, Random phase approximation calculations of K-edge rotational strengths of chiral molecules: Propylene oxide. *Chem. Phys. Lett.* **223**, 402–410 (1994).
50. M. Hodecker, M. Biczysko, A. Dreuw, V. Barone, Simulation of vacuum UV absorption and electronic circular dichroism spectra of methyl oxirane: The role of vibrational effects. *J. Chem. Theory Comput.* **12**, 2820–2833 (2016).
51. D. Kroner, Laser-driven electron dynamics for circular dichroism in mass spectrometry: From one-photon excitations to multiphoton ionization. *Phys. Chem. Chem. Phys.* **17**, 19643–19655 (2015).
52. C. Meinert, A. D. García, J. Topin, N. C. Jones, M. Diekmann, R. Berger, L. Nahon, S. V. Hoffmann, U. J. Meierhenrich, Amino acid gas phase circular dichroism and implications for the origin of biomolecular asymmetry. *Nat. Commun.* **13**, 502 (2022).
53. S. Pizzarello, C. T. Yarnes, Chiral molecules in space and their possible passage to planetary bodies recorded by meteorites. *Earth Planet. Sci. Lett.* **496**, 198–205 (2018).
54. T. Yanai, D. P. Tew, N. C. Handy, A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **393**, 51–57 (2004).
55. T. Miyahara, J.-y. Hasegawa, H. Nakatsuji, Circular dichroism and absorption spectroscopy for three-membered ring compounds using Symmetry-Adapted Cluster-Configuration Interaction (SAC-CI) method. *Bull. Chem. Soc. Jpn.* **82**, 1215–1226 (2009).
56. J. Neugebauer, E. J. Baerends, M. Nooijen, J. Autschbach, Importance of vibronic effects on the circular dichroism spectrum of dimethyloxirane. *J. Chem. Phys.* **122**, 234305 (2005).
57. J. R. Cronin, S. Pizzarello, Enantiomeric excesses in meteoritic amino acids. *Science* **275**, 951–955 (1997).
58. D. P. Glavin, J. P. Dworkin, Enrichment of the amino acid L-isovaline by aqueous alteration on CI and CM meteorite parent bodies. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5487–5492 (2009).
59. G. Cooper, A. C. Rios, Enantiomer excesses of rare and common sugar derivatives in carbonaceous meteorites. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3322–E3331 (2016).
60. J. Bailey, A. Chrysostomou, J. H. Hough, T. M. Gledhill, A. McCall, S. Clark, F. Ménard, M. Tamura, Circular polarization in star-formation regions: Implications for biomolecular homochirality. *Science* **281**, 672–674 (1998).
61. M. Buschermöhle, D. C. B. Whittet, A. Chrysostomou, J. H. Hough, P. W. Lucas, A. J. Adamson, B. A. Whitney, M. J. Wolff, An extended search for circularly polarized infrared radiation from the OMC-1 region of Orion. *Astrophys. J.* **624**, 821–826 (2005).

62. A. Chrysostomou, P. W. Lucas, J. H. Hough, Circular polarimetry reveals helical magnetic fields in the young stellar object HH 135–136. *Nature* **450**, 71–73 (2007).
63. J. Kwon, M. Tamura, P. W. Lucas, J. Hashimoto, N. Kusakabe, R. Kandori, Y. Nakajima, T. Nagayama, T. Nagata, J. H. Hough, Near-infrared circular polarization images of NGC 6334-V. *Astrophys. J.* **765**, L6 (2013).
64. A. Bergantini, M. J. Abplanalp, P. Pokhilko, A. I. Krylov, C. N. Shingledecker, E. Herbst, R. I. Kaiser, A combined experimental and theoretical study on the formation of interstellar propylene oxide ($\text{CH}_3\text{CHCH}_2\text{O}$)—A chiral molecule. *Astrophys. J.* **860**, 108 (2018).
65. D. R. Paulson, A. S. Murray, D. Bennett, E. Mills Jr., V. O. Terry, S. D. Lopez, Photochemistry of epoxides. 3. Direct irradiation of propylene oxide in the gas phase. *J. Org. Chem.* **42**, 1252–1254 (1977).
66. J. C. Forbes, J. Alves, D. N. C. Lin, A solar system formation analogue in the Ophiuchus star-forming complex. *Nat. Astron.* **5**, 1009–1016 (2021).
67. J. R. Cronin, S. Pizzarello, Amino acid enantiomer excesses in meteorites: Origin and significance. *Adv. Space Res.* **23**, 293–299 (1999).
68. S. Pizzarello, J. R. Cronin, Non-racemic amino acids in the Murray and Murchison meteorites. *Geochim. Cosmochim. Acta* **64**, 329–338 (2000).
69. A. J. Miles, S. V. Hoffmann, Y. Tao, R. W. Janes, B. A. Wallace, Synchrotron radiation circular dichroism (SRCD) spectroscopy: New beamlines and new applications in biology. *Spectroscopy* **21**, 245–255 (2007).
70. A. J. Miles, R. W. Janes, A. Brown, D. T. Clarke, J. C. Sutherland, Y. Tao, B. A. Wallace, S. V. Hoffmann, Light flux density threshold at which protein denaturation is induced by synchrotron radiation circular dichroism beamlines. *J. Synchrotron Radiat.* **15**, 420–422 (2008).
71. F. Pulm, J. Schramm, J. Hormes, S. Grimme, S. D. Peyerimhoff, Theoretical and experimental investigations of the electronic circular dichroism and absorption spectra of bicyclic ketones. *Chem. Phys.* **224**, 143–155 (1997).
72. A. C. Evans, C. Meinert, J. H. Bredehoeft, C. Giri, N. C. Jones, S. V. Hoffmann, U. J. Meierhenrich, Anisotropy spectra for enantiomeric differentiation of biomolecular building blocks. *Top. Curr. Chem.* **341**, 271–299 (2013).
73. J. Bloino, M. Biczysko, F. Santoro, V. Barone, General approach to compute vibrationally resolved one-photon electronic spectra. *J. Chem. Theory Comput.* **6**, 1256–1274 (2010).
74. A. Rizzo, O. Vahtras, *Ab initio* study of excited state electronic circular dichroism. Two prototype cases: Methyl oxirane and R-(+)-1,1'-bi(2-naphthol). *J. Chem. Phys.* **134**, 244109 (2011).

Acknowledgments: A.D.G. is grateful for a PhD scholarship from the French Ministry of Science and Education. Computation for the work described here was supported by the Azzura High-Performance Computing center of the Observatoire Pluridisciplinaire des Alpes-Maritimes (OPAL) research facility. We are grateful to U.J. Meierhenrich for discussions on gas phase spectroscopy. **Funding:** This work was supported by the EU Framework Program for Research and Innovation Horizon 2020 (grants ERC 804144 and CALIPSOplus 730872) and by the French National Research Agency (grant ANR-18-CE29-0004-01). **Author contributions:** Conceptualization: N.C.J., S.V.H., and C.M. Methodology: N.C.J., S.V.H., and C.M. Investigation: A.D.G., J.T., J.B., N.C.J., S.V.H., and C.M. Visualization: J.T. Supervision: C.M. Writing—original draft: C.M. Writing—review and editing: A.D.G., J.T., J.B., N.C.J., S.V.H., and C.M. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 13 June 2022
 Accepted 21 October 2022
 Published 18 November 2022
 10.1126/sciadv.add4614

Future direction and project

Project 1 : Deciphering odor recognition at the molecular level

One of the most fundamental and still unanswered questions in the field of olfaction is how odorant molecules are recognized by olfactory receptors and how this molecular recognition is translated into receptor activation or inhibition, ultimately giving rise to odor perception. Despite the central role of ORs in sensory biology, the molecular rules governing ligand selectivity, efficacy, and antagonism remain poorly understood. Addressing this gap represents a core objective of my long-term research program.

I have the ambition of establishing a predictive framework linking odorant chemical structure, receptor binding modes, and functional outcomes. To achieve this, I will follow an integrative research strategy that combines computational molecular modeling with targeted experimental validation.

The cornerstone of this strategy is a tight coupling between *in silico* and *in vitro* approaches. On the computational side, advanced numerical methods (including molecular docking, molecular dynamics simulations, and free-energy calculations) are systematically employed to characterize odorant-receptor complexes at atomic resolution. These structural analyses are not an end in themselves but are used to generate precise, testable hypotheses regarding the functional role of individual amino acids in ligand binding, receptor activation, and inhibition.

Experimentally, these hypotheses are evaluated through close collaboration with *in vitro* platforms, using site-directed mutagenesis combined with functional assays that quantify receptor responses upon odorant stimulation. This iterative loop between modeling and experimentation constitutes a central pillar of my long-term approach, allowing continuous refinement of molecular models based on experimental feedback.

Within this framework, my research progressively focuses on representative receptor systems, including ectopically expressed ORs and other receptor such as TAARs directly involved in the perception odors. These models provide complementary perspectives, enabling the identification of general principles of odor recognition while also addressing receptor-specific mechanisms linked to perceptual relevance.

By pursuing this long-term, integrative strategy, my objective is to move beyond descriptive studies of individual receptors and contribute to a unified molecular understanding of olfaction, ultimately enabling rational predictions of odorant–receptor interactions and their functional consequences.

Project 2: Functional study of Odorant Binding Proteins

Understanding the physiological role of OBPs has remained a major and unresolved challenge since their initial discovery and characterization. Although OBPs were originally described as key actors of vertebrate olfaction, accumulating experimental evidence has progressively challenged this view. In particular, heterologous expression studies have clearly demonstrated that ORs are fully functional in the absence of OBPs. This observation has profoundly questioned the canonical role of OBPs in odor perception and has left their true physiological function largely undefined.

More recent data suggest that OBPs may instead play a more prominent role in pheromone detection, particularly within the vomeronasal organ. However, even in this context, the molecular mechanisms by which OBPs contribute to chemosensory signaling remain poorly understood. A central unanswered

question, which constitutes a core axis of my long-term research program, is whether mammalian OBPs directly interact with olfactory or vomeronasal receptors and, if so, how such interactions modulate ligand recognition, selectivity, and signal transduction.

My long-term objective is to establish a comprehensive molecular and biophysical framework describing OBP function across physiological contexts. To achieve this goal, I am developing an integrated strategy that combines molecular modeling, structural biophysics, and functional assays, with a particular emphasis on bridging conditions that more closely reflect the native chemosensory environment. The affinity of OBPs for small VOCs is in the micromolar range and they have specificity for certain VOCs¹. Notably, OBPs can be expressed in bacterial systems at low cost and are easily purified. The currently available information on their structures and affinities for different molecules allows the design of specific mutants with targeted characteristics, representing a solid basis for tailoring OBPs to dedicated applications. One of the main challenges in using OBPs as sensors is rooted in the stability and activity of these proteins when in the dry phase.

A first strategic axis of this research focuses on characterizing ligand-induced conformational changes in OBPs from different mammalian species. Indeed, in order that the complex is recognized by the ORs, structure of OBP should be modified during the odorant binding process. By combining molecular dynamics (MD) simulations with synchrotron-based circular dichroism (CD) measurements, I aim to determine whether binding of volatile organic compounds (VOCs) induces conserved or species-specific structural rearrangements. This approach will provide key insights into whether OBPs behave as passive carriers or as dynamic molecular sensors capable of transmitting conformational information upon ligand binding.

A second major axis addresses a critical but largely unexplored aspect of OBP biology: their behavior in non-aqueous, or dry, environments. OBP-VOC interactions have been extensively characterized in the liquid phase, where ligand binding is predominantly driven by hydrophobic interactions within a well-defined orthosteric cavity. However, the relevance of these observations to physiological conditions remains uncertain, as the chemosensory interface is characterized by fluctuating hydration states. A fundamental question is whether OBPs retain their structural integrity, functional activity, and ligand specificity under dry or low-humidity conditions.

In this context, my research aims to determine whether VOC binding in the dry phase still occurs within the canonical binding pocket, or whether alternative, possibly allosteric, binding sites become favored, leading to altered specificity or promiscuous interactions. Addressing these questions will require a combination of advanced MD simulations explicitly modeling dehydration effects and experimental biophysical measurements capable of probing protein stability and ligand binding outside classical aqueous environments.

Beyond their biological significance, these investigations have important translational implications. By systematically evaluating *i)* the structural stability of OBPs in the dry phase, *ii)* their ability to remain functionally competent under such conditions, and *iii)* the preservation or loss of ligand specificity, this long-term research program will establish the foundations necessary to exploit OBPs as robust biosensing elements for the detection of volatile chemicals.

Overall, this career-long effort aims to reposition OBPs within chemosensory biology, not as auxiliary carriers of odorants, but as active molecular players whose function depends on their structural dynamics, environmental context, and interaction networks. Through this integrative strategy, my goal is to resolve a long-standing question in olfactory research while opening new avenues at the interface of sensory biology and bio-inspired sensing technologies.

Project 3: Odor and emotion

Since antiquity, smells have been attributed the power to trigger physical and psychological reactions. Pharmacopoeias from ancient civilizations, including Egyptian, Chinese, and Indian traditions, describe the use of essential oils and aromatic substances for treating disorders of the body and mind. In a more contemporary context, aromatherapy associates essential oils with anti-stress, energizing, and antidepressant properties. However, despite widespread use, a robust scientific foundation for these claims is still lacking. Nevertheless, it is intuitive that odors play a role in well-being, as olfactory perception is intimately connected to brain systems involved in mood, memory, and desire. These longstanding observations foreshadow modern neuroscientific findings demonstrating close links between olfaction and emotion.

Neuroanatomical links between olfaction and emotion

The sense of smell is uniquely connected to emotional processing. From a neuroanatomical perspective, the olfactory system is interconnected through reciprocal projections with brain regions involved in emotion (amygdala and orbitofrontal cortex) and memory (hippocampus). Olfactory perception is therefore modulated by experience and physiological state. Conversely, odors themselves can influence both psychological and physiological states. The olfactory system also exhibits strong connections with the limbic system²⁻⁴, indicating that odor perception can directly affect emotions, cognition, and behavior⁵.

The impact of olfactory stimulation on physiological responses governed by the autonomic nervous system has been clearly demonstrated. In humans, such effects can be assessed by measuring physiological variables under autonomic control, including skin temperature, electrodermal activity, heart rate, respiratory rate, and muscle contraction. For example, essential oils of ylang-ylang and peppermint, as well as the molecule linalool (with lavender- and bergamot-like odor), have been reported to lower body temperature, suggesting relaxing properties. In another study, correlations were observed between the perceived hedonic value of odorants and autonomic nervous system responses: molecules judged as pleasant induced a decrease in electrodermal conductance, consistent with a relaxed state.

Together, these results demonstrate the relevance of monitoring autonomic nervous system activity to evaluate the emotional effects of olfactory stimulation.

Challenges in identifying emotions triggered by odors

Emotions are fundamental to social interaction and interpersonal communication in living organisms⁶. They are commonly defined as complex psychological states integrating subjective experience, physiological activation, and expressive behavior, and are typically intense yet transient in nature⁶. Motivation constitutes a goal-oriented process that is tightly interconnected with emotional functioning^{6, 7}. Both motivational and emotional processes recruit the orbitofrontal cortex, a brain region central to decision-making. Because emotional states shape decision-making, everyday behaviors are largely influenced by emotions⁸.

Emotional valence is defined here according to the bi-dimensional model of emotions⁹, in which emotion is characterized along two principal dimensions: valence (pleasant/unpleasant) and arousal (from very calm and relaxed to highly alert or excited). These dimensions can be quantified using questionnaires and human psychophysiological measurements.

France is a major international player in the fields of perfumery and flavor creation. Yet, within these domains, the identification of new compounds of interest still relies largely on serendipity, as relationships between molecular structure and sensory properties, such as odor, taste, and emotional valence, remain largely elusive.

Intuitively, expert perfumers appear to master such relationships, suggesting that they do exist and can be deciphered. Beyond their control of conscious olfactory perception, these experts design compositions with the objective of eliciting subtle emotional responses that are difficult to define and quantify. This difficulty arises from the extreme complexity of chemosensory biological systems and from cultural and genetic variability across individuals.

A variety of approaches have been developed for emotion recognition. Dimensional models based on subjective self-reports are widely used^{10, 11}, along with odor-specific instruments such as the Geneva Emotion and Odor Scale¹². Behavioral approaches include facial expression analysis¹³ and vocal parameters such as pitch¹⁴, most often centered on Ekman's six basic emotions¹⁵.

Physiological approaches, including electroencephalography-based emotion recognition, have also been explored. However, two major limitations have been identified: the lack of comprehensive datasets allowing all emotions to be recognized using common variables, and the absence of consensus for certain emotions, such as amusement¹⁶. These difficulties may arise from differences in induction modalities (visual, olfactory, auditory), failure to elicit the intended emotion, or overlapping physiological patterns across distinct emotions.

Although physiological-based emotion recognition is often considered more objective than subjective approaches¹⁷, it remains largely restricted to basic emotions, whereas a wide diversity of positive emotions still lacks objective characterization. Interestingly, the most effective strategy combines subjective, behavioral, and physiological components, yet such multimodal approaches remain rare due to their methodological complexity. Accordingly, the main objective of the project is to reliably elicit emotions of interest and to measure both corresponding physiological responses and subjective ratings. Visual stimuli are traditionally used for emotion induction¹⁶, but recent technological advances have promoted the use of virtual reality^{18, 19}.

The identification of emotions elicited by odors remains particularly challenging because of the difficulty in accurately labeling subjective emotional experiences. Consequently, a second objective is to characterize perfumes according to the emotions they evoke by combining physiological measurements with subjective evaluations. Finally, because odor perception is strongly shaped by individual preferences and personal experiences²⁰, a goal is to develop a methodology for individual-level emotion recognition integrating physiological signals and subjective reports.

Uncategorized References

1. Briand, L.; Nespoulous, C.; Perez, V.; Rémy, J. J.; Huet, J. C.; Pernollet, J. C., Ligand-binding properties and structural characterization of a novel rat odorant-binding protein variant. *European journal of biochemistry* **2000**, 267 (10), 3079-3089.
2. Brand, G., La latéralisation olfactive chez l'homme—Revue de la littérature—. *Neurophysiologie Clinique/Clinical Neurophysiology* **1999**, 29 (6), 495-506.
3. Millot, J.-L.; Brand, G.; Morand, N., Effects of ambient odors on reaction time in humans. *Neuroscience Letters* **2002**, 322 (2), 79-82.
4. Savic, I., Processing of odorous signals in humans. *Brain research bulletin* **2001**, 54 (3), 307-312.
5. Degel, J.; Köster, E. P., Odors: implicit memory and performance effects. *Chemical senses* **1999**, 24 (3), 317-325.
6. Hockenbury, D. H.; Hockenbury, S. E., *Discovering psychology*. Macmillan: 2010.
7. Raymond, J., Interactions of attention, emotion and motivation. *Progress in brain research* **2009**, 176, 293-308.
8. Shin, D.; Wang, Z., The experimentation of matrix for product emotion. *Procedia Manufacturing* **2015**, 3, 2295-2302.
9. Lang, P. J.; Greenwald, M. K.; Bradley, M. M.; Hamm, A. O., Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* **1993**, 30 (3), 261-273.

10. Plutchik, R., A psychoevolutionary theory of emotions. Sage Publications: 1982.
11. Russell, J. A., A circumplex model of affect. *Journal of personality and social psychology* **1980**, 39 (6), 1161.
12. Chrea, C.; Grandjean, D.; Delplanque, S.; Cayeux, I.; Le Calvé, B.; Aymard, L.; Velazco, M. I.; Sander, D.; Scherer, K. R., Mapping the semantic space for the subjective experience of emotional responses to odors. *Chemical senses* **2009**, 34 (1), 49-62.
13. Tracy, J. L.; Robins, R. W.; Schriber, R. A., Development of a FACS-verified set of basic and self-conscious emotion expressions. *Emotion* **2009**, 9 (4), 554.
14. Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W.; Taylor, J. G., Emotion recognition in human-computer interaction. *IEEE Signal processing magazine* **2001**, 18 (1), 32-80.
15. Ekman, P., An argument for basic emotions. *Cognition & emotion* **1992**, 6 (3-4), 169-200.
16. Kreibig, S. D., Autonomic nervous system activity in emotion: A review. *Biological psychology* **2010**, 84 (3), 394-421.
17. Wang, X.-W.; Nie, D.; Lu, B.-L., Emotional state classification from EEG data using machine learning approach. *Neurocomputing* **2014**, 129, 94-106.
18. Voigt-Antons, J. N.; Lehtonen, E.; Palacios, A. P.; Ali, D.; Kojic, T.; Möller, S. In *Comparing Emotional States Induced by 360° Videos Via Head-Mounted Display and Computer Screen*, 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 26-28 May 2020; 2020; pp 1-6.
19. Yan, X.; Yan, Y.; Cao, M.; Xie, W.; O'Connor, S.; Lee, J. J.; Ho, M.-H., Effectiveness of virtual reality distraction interventions to reduce dental anxiety in paediatric patients: a systematic review and meta-analysis. *Journal of dentistry* **2023**, 132, 104455.
20. Arshamian, A.; Gerkin, R. C.; Kruspe, N.; Wnuk, E.; Floyd, S.; O'Meara, C.; Rodriguez, G. G.; Lundström, J. N.; Mainland, J. D.; Majid, A., The perception of odor pleasantness is shared across cultures. *Current Biology* **2022**, 32 (9), 2061-2066. e3.

