

# OlfactionDB: A Database of Olfactory Receptors and Their Ligands

D. Modena<sup>1</sup>, M. Trentini<sup>1,2</sup>, M. Corsini<sup>1</sup>, A. Bombaci<sup>1</sup>, A. Giorgetti<sup>1,2,\*</sup>

<sup>1</sup>Department of Biotechnology, University of Verona, strada Le Grazie 15, 37134, Verona, Italy

<sup>2</sup>German Research School for Simulation Science, FZ-Juelich and RWTH Aachen, Germany

**Abstract** Odorants are volatile molecules that efficiently carry chemical information, providing one of the main ways for communicating with the environment in all kingdoms of life. In the other hand, mammalian genomes codify for hundreds of olfactory receptors (ORs), e.g. about 400 in human and more than 1000 in mouse, underlying the crucial role of the sense of smell during evolution. Therefore, the olfactory system is capable to discriminate between ~10,000 different odors. The possibility of collecting and compiling information about odorants and their receptors is thus fundamental for a functional characterization of the signaling firing event. OlfactionDB, a manually curated database providing comprehensive information for nearly 400 odorant-receptor interactions at the current state, has been developed for managing information about odorants and their receptors. OlfactionDB is a free publicly database available online from: <http://molsim.sci.univr.it/OlfactionDB>.

**Keywords** Olfactory Receptors, Ligand-Receptor Interactions, Odorant Molecules

## 1. Introduction

The G protein-coupled receptors (GPCRs) family is the largest membrane-bound receptor family expressed by mammals (encompassing more than 1% of the genome). They are involved in an enormous variety of intra- and extracellular signaling pathways, including detection of light, odors and taste; neurotransmission; inflammation; cardiac and smooth muscle contractility[1]. Ligand (or photon) binding to GPCRs fires a cascade of events, producing an electrical signal as output. More than half of the GPCRs codified in mammalian genomes are olfactory receptors (ORs)[2,3], underlying the crucial role of the sense of smell during evolution. With such an impressive number of different ORs, the olfactory system is capable to discriminate between ~10,000 different odors: one odorant can activate numerous types of ORs, while a single OR can be activated by several different odorants. Thus, it is important to rationally collect and compile information about interaction affinities of odorants and their receptors. In the last few years, several human/mouse OR-odorants interaction affinities have been published[4-13]. Here we present a free, publicly available, database, OlfactionDB, that contains affinity data of human/mouse olfactory receptors (OR) and ligands, manually compiled and extracted from the literature,

containing experimental values for almost 400 OR-odorant interacting pairs.

## 2. Methods

The whole families of mouse and human ORs were retrieved from the Uniprot database[14]. The ORs for which experimental data exist[4-13] were then aligned using the program PROMALS[15] and manually checked in order to eliminate redundancies. We used in-home-written Python scripts to manipulate the data, whereas annotations regarding the interaction between G-proteins and ligands, the affinity of a particular interaction and the corresponding references were appended manually in a spreadsheet. The data has been organized on the basis of a relational model (Figure. 1) and stored in a PostgreSQL database system. The user has supervisory access through our Apache web server interferential software, which was developed in Java for database manipulation. This software tends to settle any web server's query.

The OR's sequence similarity searches can be carried out in the database using a server-side version of the program Ssearch[16] and the ligand similarity searches can be performed by using a server-side version of OpenBabel (<http://openbabel.sf.net>).

The three dimensional structures of the odorant molecules can be visualized with Jmol[17] and the multiple sequence alignment of the ORs included in the database can be analyzed by a Jalview[18] applet. A tree as the one showed in Figure. 2 can be easily obtained by using the

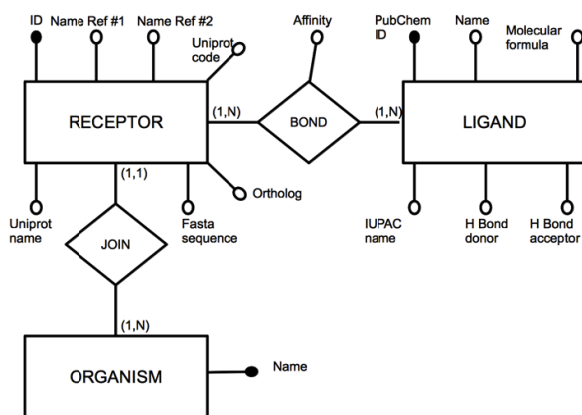
\* Corresponding author:

[alejandro.giorgetti@univr.it](mailto:alejandro.giorgetti@univr.it) (A. Giorgetti)

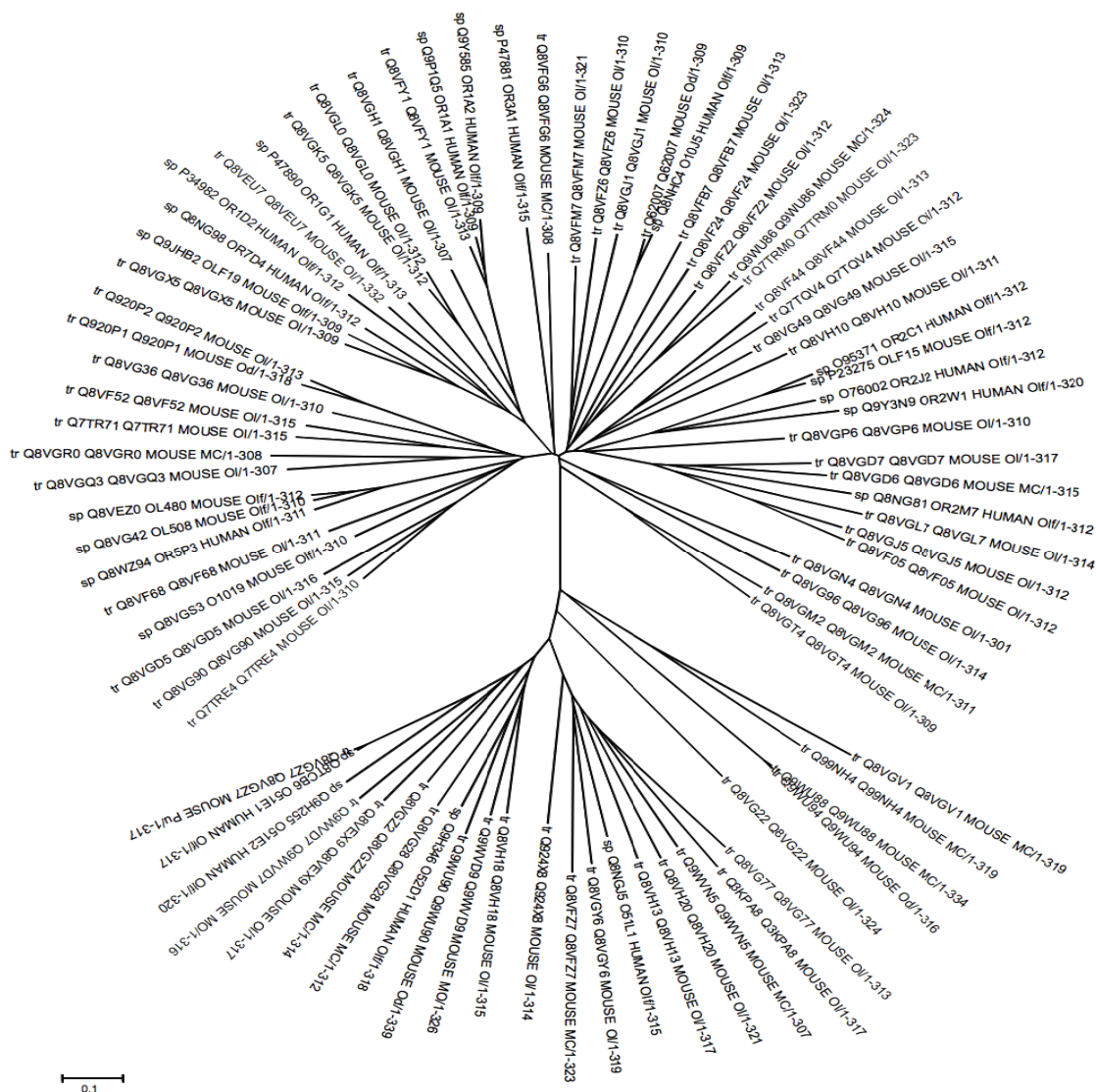
Published online at <http://journal.sapub.org/als>

Copyright © 2011 Scientific & Academic Publishing. All Rights Reserved

Jalview program.



**Figure 1.** Relational scheme. Each type of entry, i.e. receptor odorant, is represented by different fields, depending on the nature of the molecule. Both types of entries are linked together by their affinity values. While receptor type fields are: names, sequence, Uniprot accession code, orthologs and taxonomy; ligand type fields are: H-bonds, names, formula, and PubChem accession code.



**Figure 2.** Evolutionary relationships between olfactory receptors deposited in OlfactionDB. The evolutionary history was inferred using the Neighbour-Joining method. The optimal tree with the sum of branch length = 25.29363918 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 83 amino acid sequences. All positions containing gaps and missing data were eliminated.

### 3. Results

The current version of OlfactionDB contains comprehensive information for a) 85 odorants, b) 83 odorant receptors, totalizing information for about 400 ligand-receptors pairs. All the data were extracted and annotated from articles published in the last ten years[4-13]. The greatest amount of data was indeed extracted from the seminal work of Saito and co-workers[8].

One of the principal problems encountered during the preparation of the database, was the high heterogeneity of nomenclatures found in the different databases explored and on the reference articles. Thus, in order to simplify the database navigation and data retrieval, OlfactionDB offers, through its main page, several search tools: a) full database navigation, b) text and name search, including nomenclatures from at least three different independent sources and c) Uniprot accession code search. Furthermore, similarity searches can be performed using the programs Ssearch and OpenBabel, for receptors and ligands, respectively. An example of a similarity search using Ssearch is shown in Figure 3.

**STATISTICALLY SIGNIFICANT SSEARCH SCORES:**

UniProt Code	Description	Length	E-value
Q9Y585	OR1A2_HUMAN	309	2.2e-85
Q9P1Q5	OR1A1_HUMAN	309	2e-71
Q8VFPY1	Q8VFPY1_MOUSE	313	3e-65
Q8VGH1	Q8VGH1_MOUSE	307	4.9e-45
734982	OR1D2_HUMAN	312	2.6e-40
Q8VEU7	Q8VEU7_MOUSE	332	8.3e-40
Q8VGS5	Q8VGS5_MOUSE	312	4e-39
747890	OR1G1_HUMAN	313	9e-39
Q8VGL0	Q8VGL0_MOUSE	312	6.1e-38
Q8VGS3	Q1019_MOUSE	310	1.2e-34

**Figure 3.** Search output screen. Similarity searches through the internal database are carried out using the program Ssearch. A link to the corresponding entry is offered, together with the length of the subject and the e-value.

Summarizing, while a typical entry for an odorant (Figure 4) offers a chemical characterization of the compound, including information regarding the binding affinities, the number of putative H-bond donors and acceptors and the possibility of visualizing its three-dimensional structure; a typical entry for a receptor includes three different nomenclatures, Uniprot accession code, name, FASTA sequence and affinity data for its cognate ligands in terms of the EC50 reported in literature. Moreover, in each of the entry pages, direct links to the corresponding Uniprot and PubChem databases entries are given for the receptor and the ligand respectively. Finally, all the receptor sequences included in the databases were aligned and the resultant multiple sequence alignment can always be reached from the Jalview[18] link in the main menu, allowing further analysis through the applications, i.e. new sequences can be uploaded and realigned, phylogenetic trees can be created immediately and several color and edition possibilities are offered. Regarding this tool, it is important to analyse the

phylogenetic tree obtained from the receptors included in the database. As can be observed in Figure. 2, the phylogenetic tree vividly shows the evolutionary relationships between the different members of the family for which experimental information is available. In particular, it can be appreciated the way in which the putative orthologs, i.e. from human and mouse, group together. This observation, although being one of the principal achievements of the present work might be of fundamental importance for a future extensive family functional assessment and in the process of deorphanization of other members of the family.

The database here presented includes innovative and unique features such as the affinity values, not available in any other publicly accessible resource in the field, i.e. orDB and odordb[19], GPCRDB[20] and SuperScent[21]. The relational scheme as presented in Figure. 1, on which the presented database is organized, is especially designed to capture the coupling preferences of ORs and their ligands according to the reported data in the scientific literature. The database can be quickly updated upon new published ORs-odorants affinity values.

**Show complete list**

PubChem Id	Name	Iupac Name	Molecular Formula	H Bond Donor	H Bond Acceptor	Jmol
8174	1-DECANOL	decan-1-ol	C <sub>10</sub> H <sub>22</sub> O	1	1	

**KNOWN RECEPTORS**

UniProt Id	UniProt Gene Name	Ref #1 Name	Ref #2 Name	Organism	FASTA
Q8VGP6	Olf447	MOR261-1	MOR6.3.13	mus musculus	>tr Q8VGP6 Q8VGP6_MOUSE Olfactory r musculus GN=Olf447 PE=2 SV=1 MGGNQLTITQILLGFLSFRMMLLFAFLSFLFY FLSHLAIVDIAHACTVFPGLVNLMSRPFISFA RVAICRPSRYTAINRQVIVLISMLGVLLF LAVLKACSDTRINELMVIAGAVSVLWSPFSIVV SSHLCVVGLVYGTALVNYIGPQSGKNEQKYLII KRTLLKETS

**Figure 4.** Typical odorant entry. All the information provided in the webserver is shown: formula, names, amount of H-bonds able to form, structure and its cognate receptors.

### 4. Discussion

Olfactory receptors belong to the biggest superfamily of membrane proteins in mammals, i.e. GPCR, and constitute the subfamily with more members. Furthermore, during evolution ORs were shown to have an extremely important role for the survival of complex organisms. A deep insight into the molecular mechanisms underlying the olfactory signalling transduction is thus needed for a complete characterization of the way in which mammals communicate with the rest of the world. The database, i.e. OlfactionDB, here presented can be considered as an initial step into a more profound characterization of the relationships among receptors and their cognate ligands through the whole family of GPCRs. Moreover, detailed information regarding the odorant-receptor affinities may be the key for gaining insights into the structure and the molecular mechanisms underlying the function of the receptors. In particular, this may

help in the modelling of the binding cavities, which is the place into which inhibitors and or other kind of molecules may interact. Our group and collaborators have been involved, during the last decade, in the application of a combination of computational and experimental techniques aimed at the unravelling of the molecular mechanisms underlying the different steps of several signalling cascades, i.e. vision, olfaction and bitter taste, all of them including GPCRs as the initial part of the signalling firing event [22-28]. In fact, we have to consider that the main challenges for the near future will include the development and application of methods that permit the full description at the molecular/structural level of GPCR-ligand complexes, as they constitute one of the principal drug targets in the human organism. In this sense the availability of a freely-publicly database, accessible to the whole scientific community, able to offer details regarding ligand-receptor affinities may be of fundamental importance for the development of the field.

## ACKNOWLEDGEMENTS

Funding: Starting Grant from the department of Biotechnology, University of Verona.

The authors greatly acknowledge IllyCoffe Company (Trieste, Italy, <http://www.illy.com>), in particular Dr. Luciano Navarini for very illuminating discussions.

---

## REFERENCES

- [1] Kroeze, W. K., Sheffler, D. J. & Roth, B. L. 2003. G-protein-coupled receptors at a glance. *J. Cell. Sci*116, 4867-4869
- [2] Buck, L. & Axel, R. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*65, 175-187
- [3] Takeda, S., Kadowaki, S., Haga, T., Takaesu, H. & Mitaku, S. 2002. Identification of G protein-coupled receptor genes from the human genome sequence. *FEBS Lett*520, 97-101
- [4] Oka, Y., Omura, M., Kataoka, H. & Touhara, K. 2004. Olfactory receptor antagonism between odorants. *EMBO J*23, 120-126.
- [5] Niimura, Y. & Nei, M. 2003. Evolution of olfactory receptor genes in the human genome. *Proc. Natl. A (Bozza T., Feinstein P., 2002) cad. Sci. U.S.A*100, 12235-12240
- [6] Kajiya, K. et al. 2001. Molecular bases of odor discrimination: Reconstitution of olfactory receptors that recognize overlapping sets of odorants. *J. Neurosci*21, 6018-6025
- [7] Pick, H. et al. 2009. Dual activities of odorants on olfactory and nuclear hormone receptors. *J. Biol. Chem*284, 30547-30555
- [8] Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J. D. 2009. Odor coding by a Mammalian receptor repertoire. *Sci Signal*2, ra9
- [9] Katada, S., Hirokawa, T., Oka, Y., Suwa, M. & Touhara, K., 2005. Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. *J. Neurosci*25, 1806-1815
- [10] Zhang, X. & Firestein, S. 2002. The olfactory receptor gene superfamily of the mouse. *Nat Neurosci*5, 124-133
- [11] Bozza, T., Feinstein, P., Zheng, C. & Mombaerts, P. 2002. Odorant receptor expression defines functional units in the mouse olfactory system. *J. Neurosci*22, 3033-3043
- [12] Malnic, B., Hirono, J., Sato, T. & Buck, L.B., 1999. Combinatorial receptor codes for odors. *Cell*96, 713-723
- [13] Shirokova, E. et al. Identification of specific ligands for orphan olfactory receptors. 2005. G protein-dependent agonism and antagonism of odorants. *J. Biol. Chem*280, 11807-11815
- [14] Wu, C.H. et al. 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res*34, D187-191
- [15] Pei, J., Kim, B., Tang, M. & Grishin, N.V. 2007. PROMALS web server for accurate multiple protein sequence alignments. *Nucleic Acids Res*35, W649-652
- [16] Ropelewski, A. J. Nicholas, H. B. Jr, Deerfield, D. W. 2004. Mathematically complete nucleotide and protein sequence searching using Ssearch. *Curr Protoc Bioinformatics*. Chapter 3, Unit3
- [17] Herráez, A. 2006. Biomolecules in the computer: Jmol to the rescue. *Biochem. Mol. Biol. Educ.*34, 255-261
- [18] Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*25, 1189-1191
- [19] Skoufos, E., Marenco, L., Nadkarni, P. M., Miller, P. L. & Shepherd, G. M. 2000. Olfactory Receptor Database: a sensory chemoreceptor resource. *Nucleic Acids Research* 28, 341-343
- [20] Vroling B, Sanders M, Baakman C, Borrmann A, Verhoeven S, Klomp J, Oliveira L, de Vlieg J, Vriend G. 2010. *Nucleic Acid Research*; 39(Database issue):D309-19.
- [21] Dunkel, M. et al. 2009. SuperScent--a database of flavors and scents. *Nucleic Acids Res* 37, D291-294
- [22] Nair, A. V., Mazzolini, M., Codega, P., Giorgetti, A. and Torre, V. 2006. Locking CNGA1 channels in the open and closed state. *Biophysical J.* 90: 3599-3607
- [23] Giorgetti A, Carloni P, Mistrik P, Torre V. A. Homology Model of the Pore Region of HCN channels. *Biophysical J.* 89(2):932-944
- [24] Giorgetti, A., A. V. Nair, P. Codega, V. Torre and P. Carloni. 2005. Structural Basis of Gating of CNG channels. *FEBS letters.* 579(9):1968-1972
- [25] Giorgetti, A., Carloni, P. 2003. Molecular Modeling of Ion Channels: Structural Predictions. *Current Opinion in Chemical Biology.* 7: 150-156
- [26] Mazzolini M, Marchesi A, Giorgetti A, Torre V. 2010. Gating in CNGA1 channels. *Pflugers Arch.* 459(4):547-55

- [27] Paola Lupieri, Chuong Ha HungNguyen, ZhalehGhaemi-Bafghi, Alejandro Giorgetti, and Paolo Carloni. 2009. Computationalmolecularbiologyapproaches to ligand-target interactions. *HFSP Journal*. 3(4):228-39
- [28] Biarnés X.; Marchiori A.; Giorgetti A.; Lanzara C.; Gasparini P.; Carloni P.; Born S.; Brockhoff A.; Behrens M.; Meyerhof W. Insightsinto the Binding of Phenyltiocarbamide (PTC) Agonist to Its Target Human TAS2R38 Bitter Receptor.2010 .*PLoS ONE* , vol. 5 , n. 8 - e12394 , pp. 1-6