Biophysical Chemistry 255 (2019) 106264

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: www.elsevier.com/locate/biophyschem

Investigation of the binding between olfactory receptors and odorant molecules in *C. elegans* organism

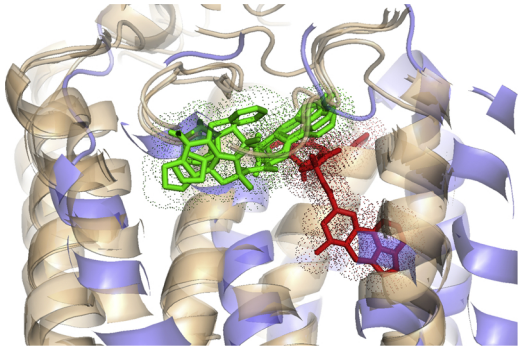
Edoardo Milanetti[a,b,*], Giorgio Gosti[a], Luca De Flaviis[b], Pier Paolo Olimpieri[b], Silvia Schwartz[a], Davide Caprini[a], Giancarlo Ruocco[a,b], Viola Folli[a]

[a] Center for Life Nanoscience, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161 Rome, Italy
[b] Department of Physics, Piazzale Aldo Moro 5, 00185 Rome, Italy

HIGHLIGHTS

- Investigating the bimolecular mechanism to underly the odor-receptor binding with a predictive method based on structural biology approach.
- A new strategy is proposed for G Protein-Coupled Receptors expressed on a *C. elegans* specific olfactory neuron AWC.
- The method is able to propose a set of possible candidate ligands for binding each GPCRs of AWC neurons, providing reliable results which are consistent with the experimental data.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Olfactory receptors
C.elegans
Volatile molecules
Chemosensory
G-protein-coupled receptor
Protein structure prediction

ABSTRACT

The molecular mechanisms regulating the complex sensory system that underlies olfaction are still not completely understood. The compounds formed from the interaction of Olfactory Receptors (ORs) with volatile molecules play a crucial role in producing the sense of olfaction. Therefore, it is necessary to investigate the binding mechanisms between these receptors and small ligands. In this work, we focus our attention on *C.elegans*, this is a particularly suitable model organism because it is characterized by a nervous system composed of only 302 neurons. To study olfaction in *C.elegans*, we select 21 ORs from its olfactory neurons, and present a pipeline, consisting of several computational methods, with the aim of proposing a set of possible candidates for binding the selected *C.elegans* ORs. This pipeline introduces an approach based on the selection of templates, and threading, that takes advantage of the structural redundancy among membrane receptors. This procedure is widely replicable because it is based on algorithms that are publicly available and are freely hosted on institutional servers.

1. Introduction

Chemoperception, and more specifically the sense of olfaction, is a versatile mechanism for the detection of volatile odorants, which allows for a highly specific discrimination of similar molecules [1]. The olfactory sense is fundamental for the survival of a wide range of living

* Corresponding author at: Department of Physics, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy.
E-mail addresses: edoardo.milanetti@uniroma1.it (E. Milanetti), giorgio.gosti@iit.it (G. Gosti), silvia.schwartz@iit.it (S. Schwartz), davide.caprini@iit.it (D. Caprini), Giancarlo.Ruocco@roma1.infn.it (G. Ruocco), viola.folli@iit.it (V. Folli).

https://doi.org/10.1016/j.bpc.2019.106264
Received 28 June 2019; Received in revised form 10 September 2019; Accepted 10 September 2019
Available online 11 September 2019

Table 1
Table of ligands in complex with GPCR templates. In the first column the ligand names are indicated for each template (fourth column). In the second and third columns are indicated the average mass (Da) and boiling point (°C at 760 mmHg) respectively. In the fourth column, pdb code for each protein template as provide by GPCR I-Tasser method. In the fifth column there are the set of *C.elegans* receptors for each template.

Ligand	Average mass (Da)	Boiling point (°C at 760 mm Hg)	PDB template	*C.elegans* gene
(*S*)-Carazolol	298.379	531.2 ± 40.0	2RH1	Srsx-5
ZM241385	337.336	–	3EML	Srj-21
-Quinuclidinyl benzilate	337.412	439.0 ± 24.0	3UON	Srsx-5
Glycine	75.067	240.9 ± 23.0	4BUO	Srj-22, Str-2
JDTic	465.628	701.9 ± 60.0	4DJH	Srd-5, Sri-14, Srj-21, Str-199, Odr-10
Neurotensin	1673	1616.4 ± 75.0	4GRV	Sra-13, Srab-16, Srd-17, Sri-14
Dihydroergotamine	583.677	899.5 ± 65.0	4IAQ	Srt-28, Srx-1
Ergotamine	581.661	914.5 ± 65.0	4IB4	Sra-13, Sre-4, Srsx-3, Srsx-5, Srsx-37, Srt-7, Srt-28, Srt-29, Srt-45, Srt-47, Str-199
ONO9780307	521.601	725.5 ± 60.0	4Z34	Sre-4, Srsx-5
AZ8838	234.269	440.3 ± 45.0	5NDD	Sra-13
AM6538	481.417	–	5TGZ	Srd-5, Srd-17, Sri-14, Srj-21, Srj-22, Srsx-7, Str-2, Str-130, Str-199
Taranabant	515.955	634.2 ± 55.0	5U09	Sra-13, Srab-16, Sre-4, Sri-14, Srsx-3, Srsx-5, Srsx-37, Srt-7, Srt-28, Srt-29, Srt-45, Srt-47, Srx-1, Odr-10
BMS-193885	590.710	707.5 ± 60.0	5ZBH	Sra-13, Srab-16, Srd-5, Srd-17, Sri-14, Srj-22, Srsx-3, Srsx-5, Srsx-37, Srt-7, Srt-47, Srx-1, Str-2, Str-199, Odr-10

organism. Nevertheless, it is still unclear how olfactory receptors recognise volatile molecules, and how the brain decodes the receptor responses in order to generate appropriate behaviours [1–3]. Currently, it is known that odorant molecules are perceived by organisms through specific membrane proteins, Odorant Receptors (ORs), which bind to volatile ligands. These proteins represent the largest subfamily within the G Protein-Coupled Receptors (GPCRs) family, and are present in all multicellular organisms [4–6].

Undestanding the ORs molecular mechanisms is crucial for many biological applications. Indeed, ORs are expressed in many cell types. In particular, ORs are present also in tumoral tissues and in these cells their level of expression is remarkably different compared to healthy tissues. Specifically, in prostate cancer cells, it has been shown that the activation of specific over-expressed ORs inhibits the proliferation of these cells. Although the complete understanding of their biological role in cancer is still elusive, it is clear that these receptors might constitute new targets for diagnosis and therapeutics [7].

Therefore, the aim of this work is to investigate how a generic odorant binds to a specific sub region. This information is fundamental in order to be able to design specific receptors for any given odorant, or, given an arbitrary receptor, to be able to predict the odorants activating it [8]. To address these issues, we select the model organism *C.elegans*, a roundworm able to detect a large class of volatile odorants. This nematode is easy to manipulate genetically and its nervous system, composed only of 302 neurons, is well known and completely mapped [9]. In *C.elegans*, olfaction is the primary mechanism for responding to environmental changes. The genome of *C.elegans* encodes an extremely large class of genes for GPCRs, approximately 1200 genes, over half of which have been suggested as chemosensory receptors. Since *C.elegans* has only 32 chemosensory neurons, each worms olfactory neuron co-expresses dozens of chemosensory receptors, and it is likely to detect a broad range of molecules. This redundancy and combinatorial complexity in the chemosensory receptor expression makes extremely difficult to associate *C.elegans* GPCRs to the sensations of specific cues and to their biological role. Specifically, *C.elegans* recognizes odorants mainly with three pairs of olfactory neurons AWA, AWC and AWB [1]. The first two mediate responses to attractive volatile odorants, while AWB neurons normally drive avoidance responses. In all three types of neurons, the chemoreception is mediated by members of the seven-transmembrane G-protein-coupled receptor class (7TM GPCRs). It has been recently shown that *C.elegans* is able to discriminate urine of cancer patients from healthy subject [10]. In fact, *C.elegans* shows an impressive attraction toward cancer metabolites. Attraction to volatile compounds is mediated mainly by a pair of wing neurons, AWCs, which express 20 GPCRs known as olfactory receptors. Even if we do not know the molecular picture of the activation of G protein coupled receptors, understanding how GPCRs bind volatile compounds can guide the determination of the dynamical coefficients in the biophysical neuron-scale models used for the characterization of the chemosensory circuit from receptors through the neurons to the effectors [11–14]. The development of a not computationally expensive strategy able to select a panel of putative ligands for these GPCRs, could guide the identification of the possible cancer metabolites activating the attractive behavior of the nematode. Despite knowledge of GPCR structures would provides crucial information in the drug design field, experimental determination of 3D structures of GPCR proteins has proved to be difficult. More specifically, for ORs no experimental structural data is available to date [6].

Here we present a computational protocol that aims to propose new possible volatile compounds as putative activators of *C.elegans*'s membrane receptors based on the idea that similar compounds bind similar receptors [15]. Therefore, for each AWC's GPCR, a set of similar receptors of known structure bounded with small compounds were selected. These will be the templates used to predict the 3-dimensional structure. Through a sequence alignment between the binding region of the templates and worm's GPCR, we give a confidence score for any proposed ligand. Furthermore, for each ligand, other molecules with similar chemical-physical properties have been proposed as additional candidates. Finally, a docking approach between any ligand and the corresponding predicted GPCR structure tests the reliability of the proposed procedure.

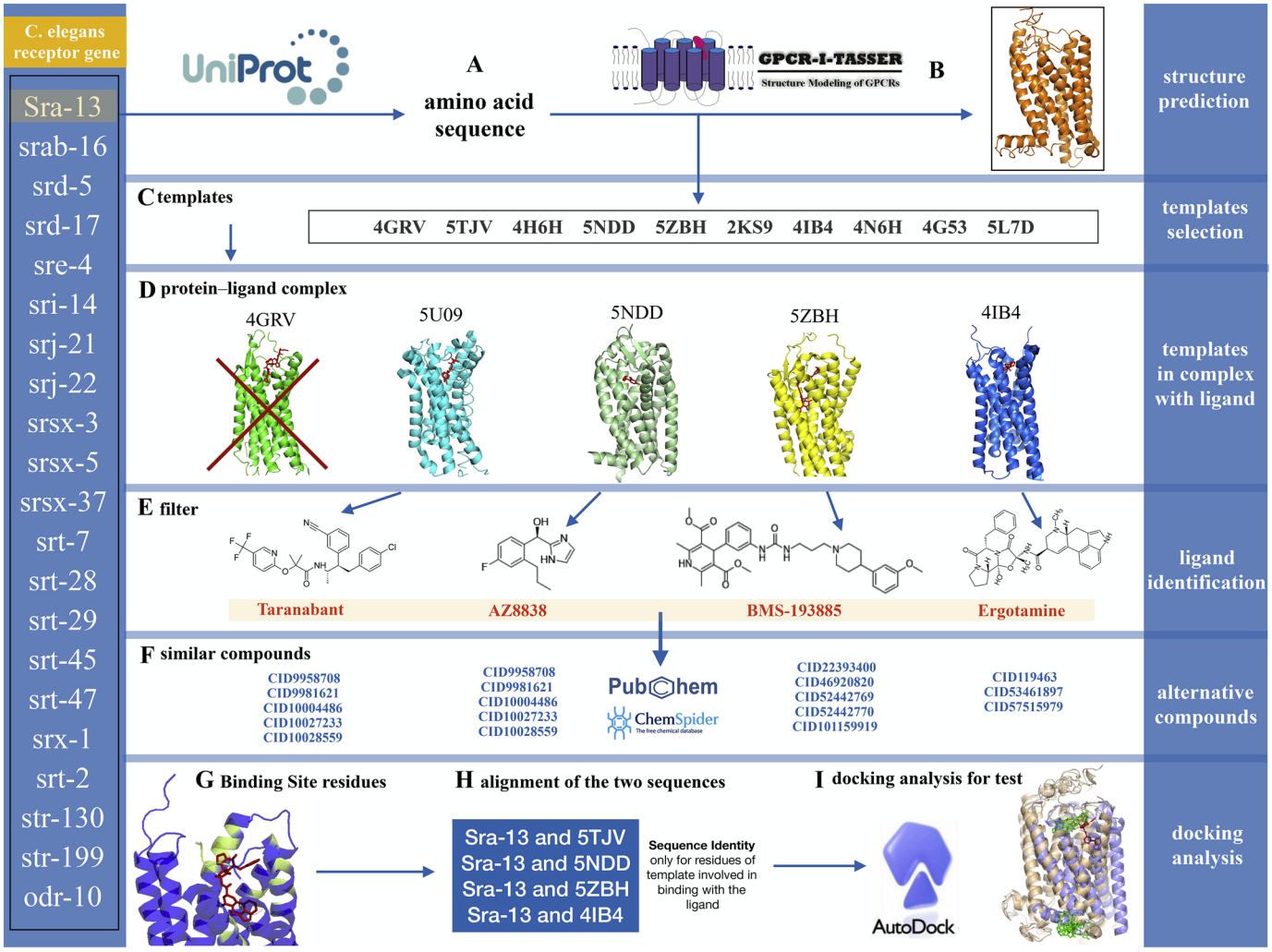


Fig. 1. As an example, the pipeline for sra-13 protein is shown. On the left column (in blue), there is the list of the target GPCRs belonging to the dataset of this work. On the right column (in blue), there are the main steps of our method: (A) starting from the nucleotide sequence, the amino acid sequence is determined by using UniProt server; (B) through the use of GPCR I-Tasser server, the three-dimensional structure of each protein is predicted; (C) list of the 10 best templates identified by GPCR I-Tasser; (D) among the 10 templates only those resolved in complex with a ligand are selected; (E) peptides are removed from the ligand candidates; (F) for each ligand, a set of chemically and physically analogous compounds are selected from PubChem and ChemSpider databases; (G) for each template, the binding site residues are identified (in yellow); (H) the binding site sequence identity between each template and the target protein is calculated through their alignment; (I) docking analysis is performed to test the method. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2. Methods

2.1. Datasets

We select all the 20 genes encoding GPCRs expressed on AWC neurons: sra-13, srab-16, srd-5, srd-17, sre-4, sri-14, srj-21, srj-22, srsx-3, srsx-5, srsx-37, srt-7, srt-28, srt-29, srt-45, srt-47, srx-1, str-2, str-130, str-199 [16]. In this discussion we include also odr-10, a gene expressed on AWA neuron, encoding one of the few GPCR directly linked to a specific odorant, namely diacetyl [17,18], also sensed by SRI-14. A summary of the dataset with information related to the biological activity of each gene is shown in Table 1. All information was obtained by consulting UniProt sever (http:// www.uniprot.org/) [19,20].

2.2. Structure prediction

Starting from 21 genes of *C.elegans*, codifying for different GPCRs of unknown structure, we obtained the corresponding amino acid sequence using Uniprot server (Fig. 1A). Their tridimensional structure was predicted through the GPCR I-TASSER computational method [21], an hybrid protocol able to construct GPCR structure model considering both experimental mutagenesis data and ab initio simulations of transmembrane helix assembly. The GPCR I-TASSER protocol can be divided in two parts: the first part individuates templates, and the second part predicts the structure. The first part finds the putative templates with Local Meta-Threading Server (LOMETS), a meta-threading method for template-based protein structure prediction [22]. In the second part, GPCR I-TASSER only uses the 10 best templates with the highest significance in terms of the threading alignments in order to predict the structure (Fig. 1B).

2.3. Selection of candidate compounds

For each target GPCR, there is a group of 10 GPCR templates (Fig. 1C). For each group, we have only selected the subset of GPCR proteins which have been resolved in complex with a ligand (Fig. 1D). We exclude the case in which the ligand was a peptide (Fig. 1E).

Table 1 reports the physical-chemical features of each found ligand. Furthermore, we associate at each ligand a list of alternative compounds with similar chemical and structural properties which forms a

Table 2
Pair alignment between the templates and the *C.elegans* GPCRs. The first column shows the pdb code for each protein template as provided by GPCR I-Tasser method. In the second column there is the corresponding *C.elegans* receptor target. The third and fourth columns give the GlobalAlig and BSAlig scores. The fifth column lists the percentage of residues in common between the template protein and the target one in the binding region. In the last column, the number of residues involved in the GPCR template-ligand binding is reported.

PDB	*C.elegans* gene	Globalalign	BSalign	BS res	N. res
2RH1	srsx-5	0.275	0.2	0.018	5
3EML	srj-21	0.285	0	0.003	1
3UON	srsx-5	0.28	0.111	0.032	9
4BUO	srj-22	0.244	0.188	0.05	16
4BUO	str-2	0.311	0.25	0.05	16
4DJH	srj-21	0.295	0.2	0.035	10
4DJH	srd-5	0.285	0.1	0.035	10
4DJH	sri-14	0.31	0.2	0.035	10
4DJH	str-199	0.303	0.4	0.035	10
4DJH	odr-10	0.303	0.2	0.035	10
4GRV	sri-14	0.297	0.25	0.04	12
4GRV	sra-13	0.296	0.167	0.04	12
4GRV	srab-16	0.312	0.25	0.04	12
4GRV	srd-17	0.327	0.417	0.04	12
4GRV	str-130	0.299	0.167	0.04	12
4IAQ	srt-28	0.289	0.182	0.04	11
4IAQ	srx-1	0.285	0.182	0.04	11
4IB4	srsx-5	0.328	0.353	0.06	17
4IB4	str-199	0.297	0.294	0.06	17
4IB4	sra-13	0.311	0.176	0.06	17
4IB4	srab-16	0.329	0.412	0.06	17
4IB4	srt-28	0.337	0.412	0.06	17
4IB4	sre-4	0.28	0.353	0.06	17
4IB4	srsx-3	0.356	0.294	0.06	17
4IB4	srsx-37	0.297	0.176	0.06	17
4IB4	srt-7	0.367	0.235	0.06	17
4IB4	srt-29	0.321	0.471	0.06	17
4IB4	srt-45	0.335	0.412	0.06	17
4IB4	srt-47	0.347	0.176	0.06	17
4Z34	srsx-5	0.33	0.286	0.024	7
4Z34	sre-4	0.299	0.143	0.024	7
5NDD	sra-13	0.331	0.125	0.027	8
5TGZ	srj-21	0.335	0.444	0.031	9
5TGZ	srj-22	0.351	0.222	0.031	9
5TGZ	str-2	0.315	0.333	0.031	9
5PDB	*C.elegans* gene	GlobalAlign	BSAlign	BS res	N. res
5TGZ	srd-5	0.271	0.111	0.031	9
5TGZ	sri-14	0.315	0.333	0.031	9
5TGZ	str-199	0.279	0.333	0.031	9
5TGZ	srd-17	0.323	0.222	0.031	9
5TGZ	str-130	0.371	0.222	0.031	9
5TGZ	srsx-37	0.367	0.444	0.031	9
5U09	srsx-5	0.335	0.083	0.042	12
5U09	sri-14	0.354	0.25	0.042	12
5U09	odr-10	0.336	0.417	0.042	12
5U09	sra-13	0.337	0.167	0.042	12
5U09	srab-16	0.29	0.333	0.042	12
5U09	srt-28	0.337	0.083	0.042	12
5U09	srx-1	0.314	0.333	0.042	12
5U09	sre-4	0.304	0.333	0.042	12
5U09	srsx-3	0.29	0.167	0.042	12
5U09	srsx-37	0.372	0.25	0.042	12
5U09	srt-7	0.317	0.167	0.042	12
5U09	srt-29	0.376	0.417	0.042	12
5U09	srt-45	0.306	0.167	0.042	12
5U09	srt-47	0.372	0.333	0.042	12
5ZBH	srsx-5	0.284	0.231	0.043	13
5ZBH	srj-22	0.285	0.231	0.043	13
5ZBH	str-2	0.272	0.308	0.043	13
5ZBH	srd-5	0.25	0	0.043	13
5ZBH	sri-14	0.282	0.154	0.043	13
5ZBH	str-199	0.291	0.231	0.043	13
5ZBH	odr-10	0.284	0.308	0.043	13
5ZBH	sra-13	0.266	0.231	0.043	13
5ZBH	srab-16	0.294	0.231	0.043	13
5ZBH	srd-17	0.31	0.154	0.043	13
5ZBH	srx-1	0.258	0.154	0.043	13
5ZBH	srsx-3	0.276	0.231	0.043	13
5ZBH	srsx-37	0.345	0.385	0.043	13
5ZBH	srt-7	0.359	0.308	0.043	13
5ZBH	srt-47	0.294	0.385	0.043	13

set of possible candidates that could bind *C.elegans* receptors (Fig. 1F). All chemical and physical features were achieved through the use of ChemSpider database (www.chemspider.com) [23] and PubChem database [24].

2.4. Method validation

The method validation is composed of three tests:

- We select all the ligands proposed for odr-10 and srt-14 and then we compare their physical-chemical features with diacetyl which is experimentally proved to bind these receptors.
- Given a resolved GPCR in complex with a specific ligand, we verify if the same small molecule binds the predicted receptor; To do this, we initially perform a global alignment between the sequence of the template GPCR and the sequence of the unknown GPCR. Then we define a binding site region of the template mapping it into the unknown GPCR. Any template-ligand complex residue is considered to be part of the binding site, if at least one of its atoms is at a distance less than 5 Å from the ligan [25,26]. We perform the alignment between the two sequences by using the "pairwiseAlignment" function of R "Biostrings" package [27,28]. Using pid function of R [29] and setting the "PID2" alignment criteria, we calculate the percentage of sequence identity both for the entire sequences (named GlobalAlign) and for the residues of the Binding Site (named BSAlign). We use Autodock Vina to predict the bound conformations of the GPCR protein target with the corresponding template ligand [30,31]. This test has been performed with the known template 4IB4, a chimeric protein of 5-HT2B-BRIL in complex with ergotamine, and the predicted structure of the unknown SRT-29, the protein with the highest BSAlign.
- The test replicates the entire protocol for the 5 U09 complex. 5 U09 is a known complex with a resolved structure formed by the Human CB1 cannabinoid receptor and Taranabant. We submit the 5 U09 GPCR sequence to the I-Tasser procol [32–34], in order to understand which candidate ligands we find and where they bind, and compare them with the binding site of Taranabant. We set the upper sequence identity threshold to 0.25, maintaining similar conditions to what is imposed by the GPCRs belonging to the dataset analyzed in this work. Then, we perform molecular docking between any identified compound and the 5 U09 receptor. The top 3 docking poses for each test are taken into account. The poses are evaluated with the root-mean-square deviation (rmsd) values which are obatined by Autodock Vina considering "RMSD lower bound" option. Any compound structure is obtained by PubChem database [24].

3. Results and discussion

The aim of this work is to propose a strategy for investigating which volatile molecules interact with transmembrane receptors of *C.elegans*. In the nematode [35] three main pairs of neurons have been directly linked to the detection of volatile compounds (odorants). These olfactory neurons, AWAs, AWBs, AWCs, are largely studied in literature. Among them, AWCs are the main olfactory neurons mediating a behavioral attractive response toward a large battery of volatile molecules. While the set of volatile odorants sensed by AWC neurons is now

Table 3
For each ligand found to be in complex with a given GPCR (first column), the corresponding set of compounds with similar chemical and physical characteristics are listed (second column).

Ligand	Alternative compounds
(*S*)-Carazolol	CID162634, CID657621, CID744496, CID744498, CID10027233
ZM241385	CID557420, CID19097290, CID135955191
-Quinuclidinyl benzilate	CID129944, CID380245, CID425114, CID13773336, CID13773338
Glycine	CID161853, CID3826162, CID5744934, CID15175457, CID87848611
JDTic	CID62211446, CID119294836, CID119294836, CID134347871
Dihydroergotamine	CID53461897, CID57515979
Ergotamine	CID119463, CID53461897, CID57515979
ONO9780307	CID11351723, CID66861790, CID66862148, CID66862149, CID66862237
AZ8838	CID818136, CID818137, CID3061227, CID3407115, CID3782154
AM6538	CID59703686
Taranabant	CID9958708, CID9981621, CID10004486, CID10027233, CID10028559
BMS-193885	CID22393400, CID46920820, CID52442769, CID52442770, CID101159919

well known (e.g. chemotaxis to benzaldehyde, butanone, isoamylalcohol, 2,3pentanedione and 2,4,5 trimethylthiazole), little is understood on the receptor-odor pairings for the several GPCRs of the AWC neurons. So, we select all the 20 genes encoding GPCRs expressed on AWC neurons. We include also odr-10 which is a gene expressed on AWA neurons known to sense diacetyl [18]. For each of these GPCRs a set of possible volatile molecules have been proposed.

Most drug discovery approaches are based on the identification of the protein target in order to obtain a verified drug target. Therefore, in ordinary studies investigating the binding between protein targets and small molecules, we are required to determine experimentally the protein three-dimensional structure, which is usually obtained using either x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy [36]. However, structure determination is not yet a straightforward process: x-ray crystallography is constrained by the difficulty of getting some proteins to form crystals, especially for transmembrane proteins, and NMR is typically to protein molecules which are smaller than GPCR [37]. Furthermore, olfactory receptors belong to the GPCR hyperfamily for which little structural data are available [38]. Therefore, in order to face the study from a structural biology point of view, we need to resort to protein structural predictive computational methods.

There are two fundamental template-based approaches in protein structure prediction, homology modeling-based methods and threading-based methods. Typically, methods of structural prediction of proteins are based on homology modeling which is an approach that uses the information from homologues with an experimentally solved 3D structure, i.e. templates. Unfortunately, the ORs belong to the GPCR family and the pairwise sequence identity between proteins in this family is low (~ 0.25) [39]. For this reason, we can not adopt homology modeling-based approaches, but we have to use protein threading-based methods. Conveniently, in the GPCR family the overall structure and the key residues in each helix are highly conserved [40]. We expect that this improves the stability of the structure prediction. Nowadays, many methods and tools for GPCRs modeling are available [41–44]. Among these we used GPCR-I-TASSER, because it is one of the best performing algorithms and was designed specially for the predction of GPCR 3D structure. GPCR-I-TASSER uses an hybrid protocol [21]. In fact, its strategy is initially to search homologous templates and then to use mutagenesis data as constraints for structure assembly simulations approach [45]. When homologous templates are not available, GPCR-I-TASSER adopts an ab initio folding program to assemble artificial helices into a 7-TM-helix bundle.

The ORs sequence heterogeneity explains their ability to bind different classes of odorant molecules which are distinct in terms of shape (aliphatic, cyclic), length, hydrophobicity, and functional groups [7]. More specifically, the binding is not highly specific because the ligand-receptor complexes are mainly characterized by weak hydrophobic intramolecular interactions. This means that a single OR could bind odorant molecules with different chemical features. At the same time, a given odorant molecule may activate different ORs [46–48]. From the acknowledgement of this redundancy in the odorant-receptor pairing, we develop an algorithm that associates a diversified set of putative binding compounds for each *C.elegans* OR protein.

3.1. Computational protocol

For a given sequence, GPCR-I-TASSER method returns five predicted structures and then we select the model with the highest score. Then, the algorithm finds ten templates and we select only those experimentally resolved in complex with a ligand. The compounds which are associated to any *C.elegans* gene with this method are shown in Table 1. When the vapor pressure for a given compound is not avalilable, we describe its volatility by reporting the mass and the boiling point [49]. The next steps are based on the idea that two proteins with an high structural similarity have a great propensity to bind similar compounds, because it is more likely that also the chemical-physical properties may be preserved [26,50–52]. Therefore, our rationale is that the ligand coupled to a given template should be accommodated in similar binding regions of the corresponding *C.elegans* GPCR because the two proteins are similar in structure. In principle, the hypothetical binding site of the predicted protein may be in a different structural region of the template protein. Nevertheless, the probability that the candidate ligand binds the target protein of *C.elegans* increases as the global sequence identity (GlobalAlign) increases. In particular, the two proteins also share more likely the same binding site region if their local sequence identity (BSAlign) is higher, see Table 2.

The BSAlign will be used as a score about the goodness of the binding between the candidate ligand and the target protein. To expand the set of candidates, for each ligand we search all those compounds with the most similar chemical-physical characteristics (molecular weight, number of donors and acceptors, hydrophobicity, accessible surface area). For each ligand in complex with a given GPCR, the alternative chemical compounds are reported in Table 3. The results corresponding to each GPCR of *C.elegans* are shown in Fig. 2.

Specifically, Fig. 2 shows how promiscuity works in both directions. For a given GPCR the protocol proposes a set of putative ligands with different chemical physical properties associated to the corresponding templates. This is the first kind of promiscuity where different binding site regions bind different ligands. In the next step, for each of the proposed ligands we identify a broader range of similar compounds, binding the same receptive range of the receptor, as shown via docking test. This is the second kind of promiscuity.

3.2. Test results

In our first test we compare the chemical features of the proposed compounds for odr-10 and sri-14 with diacetyl which is experimentally

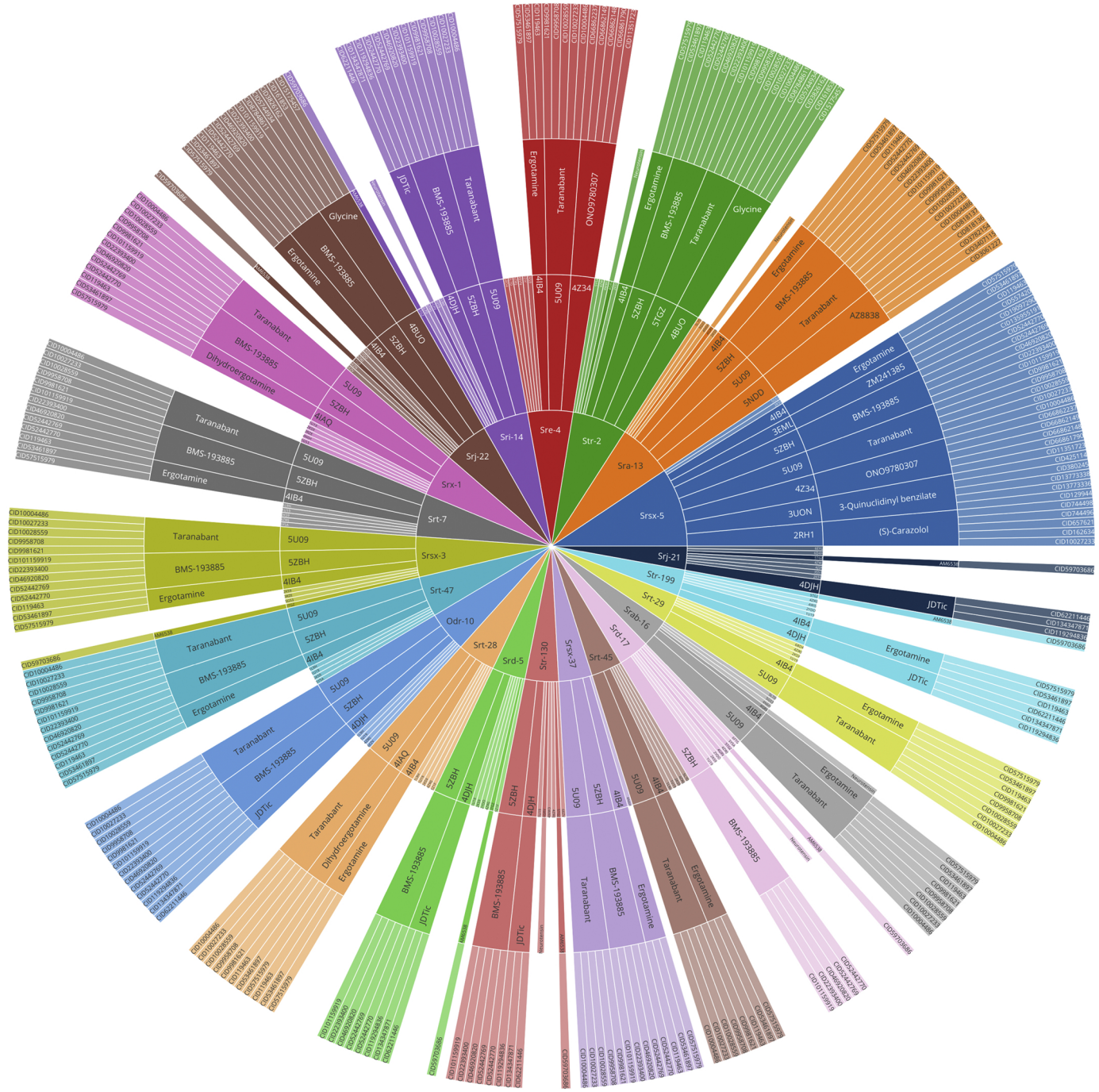


Fig. 2. For each GPCR of *C.elegans* the entire pipeline path is shown. Starting from the inner circle, the names of the 21 genes are indicated. Each of these is mapped on a larger circle where GPCR I-Tasser template proteins are indicated. Considering only GPCR template proteins in complex with a ligand, the third circle shows the names of the ligands associated with the template proteins. The outermost layer contains the names of the alternative ligands which are similar to the ligands selected in terms of physical-chemical characteristics.

proved to bind these receptors. Surprisingly, we found that these ligands share with diacetyl a central amide with two symmetric hydrophobic groups. In general, we stress that even if their dimensions are typically larger than those of ligands found to activate AWC neurons, like isoamyl alcohol, benzaldeide, butanone, 2,3 pentanedione, 2,4,5 trimethyl thiazole, etc., the candidates we identify for ODR-10 and SRI-14 share key chemical properties with those that interact with GPCRs expressed on AWCs.

In the second test, we first select the GPCR template with a binding site composed by the largest number of residues compared to all the other templates: 4IB4, a receptor binding the ergotamine ligand, having 17 residues involved in the binding site. We then investigate the binding between ergotamine ligand and a predicted *C.elegans* target protein which has 4IB4 as a template, namely SRT-29, since it has the largest binding site alignment score value among the overall target proteins considered (47.1%). To investigate if the binding between SRT-29 and ergotamine takes place in the same region of 4IB4 complex, we use a molecular docking approach (see Methods). We accept only the docking poses inside the receptor channel. Although GlobalAlig between 4IB4 and SRT-29 is only of 32.1%, the high BSAlign score (47.1%) gives a large number of shared residues between the two binding regions as shown in Fig. 3.

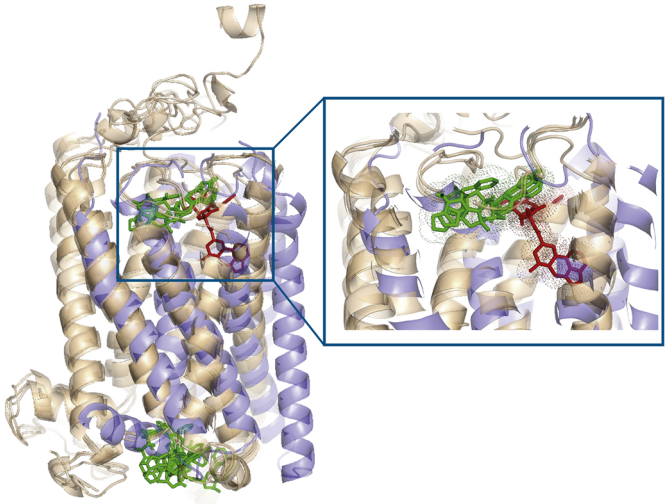


Fig. 3. Cartoon representation of molecular docking between receptors (in blue the 4IB4 template and in gold the corresponding target SRT-29) and ligand (ergotamine, in red). In green the ergotamine docking poses are shown for the predicted structure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In the third test to further validate our approach, we repeat all the protocol for 5 U09 receptor as a target protein, selecting its templates with a sequence identity less than 25%. All ligands in complex with the corresponding templates are used as chemical compound input for molecular docking with x-ray 5 U09 receptor structure. The three best docking poses have been evaluated in terms of root-mean-square deviation (rmsd). The rmsd average value among the 15 proposed docking poses is 2.66 ± 1.22. The obtained results shown in Fig. 4, are very encouraging and demonstrate that this approach proposes ligands with physical-chemical properties that maintain the binding site (Fig. 4A) and the binding mode of the complex resolved experimentally (Fig. 4B,C).

4. Conclusions

In this work, we propose a strategy to individuate a list of chemical compounds for GPCRs which are expressed on one of the olfactory neurons in *C.elegans*, namely AWC. The strategy of the protocol is based on the assumption that similar compounds are more likely to share a similar binding accommodation in the same receptor protein. Furthermore, similar target and template proteins may form binding regions for analogous ligands. The method proposed here provides reliable results which are consistent with the experimental data. Moreover, the tests in the paper, based on molecular docking approaches between GPCR structures and proposed candidate compounds, confirm the rationale and the effectiveness of our approach. Even if the dimensions of ligands proposed here are typically larger than those of ligands found to activate AWC neurons, like isoamyl alcohol, benzaldeide, butanone, 2,3 pentanedione, 2,4,5 trimethyl thiazole, etc., the candidates we find still share common fundamental features with those that interact with GPCRs expressed on AWCs. For example, as we have shown, diacetyl, known to form complexes with ODR-10 and SRI-14, shares chemically and physically analogous characteristics with the ligands that our method proposes. For future applications, this work lays the basis for guiding protein deletion experiments in *C.elegans* and allows us to discover the largely unknown odor-receptor pairings in olfactory neurons, more in general, it can be applied to other families of receptors.

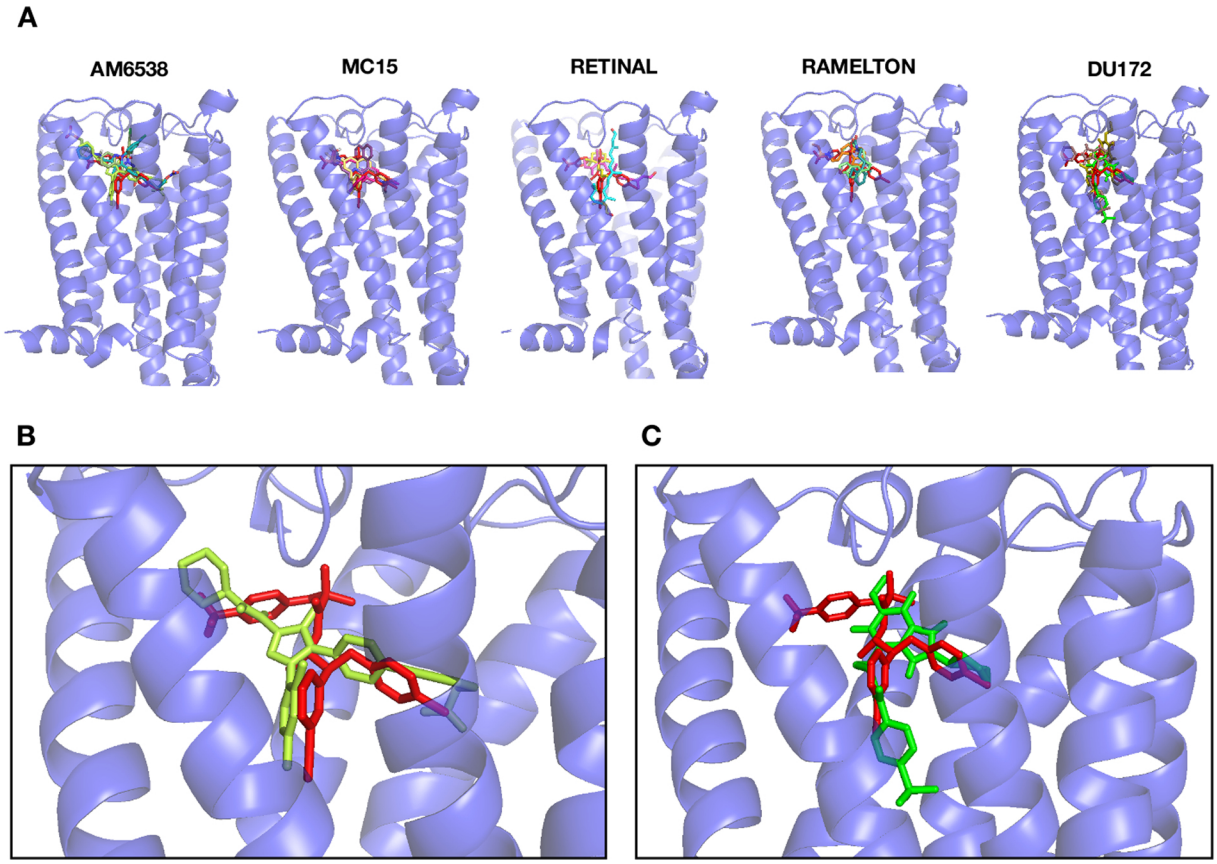


Fig. 4. Cartoon representation of molecular docking results between each identified ligand and 5 U09 receptor. The entire 5 U09 structure with the best 3 docking poses for each ligand (A). (B) and (C) show respectively a more detailed representation for the best AM6538 and DU172 poses.

Funding

This work was supported by the CrestOptics-IIT JointLab for Advanced Microscopy.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] C.I. Bargmann, E. Hartwieg, H.R. Horvitz, Odorant-selective genes and neurons mediate olfaction in c. elegans, Cell 74 (3) (1993) 515–527.
[2] Y. Jiang, N.N. Gong, X.S. Hu, M.J. Ni, R. Pasi, H. Matsunami, Molecular profiling of activated olfactory neurons identifies odorant receptors for odors in vivo, Nat. Neurosci. 18 (10) (2015) 1446.
[3] A. Dewan, A. Cichy, J. Zhang, K. Miguel, P. Feinstein, D. Rinberg, T. Bozza, Single olfactory receptors set odor detection thresholds, Nat. Commun. 9 (1) (2018) 2887.
[4] L. Buck, R. Axel, A novel multigene family may encode odorant receptors: a molecular basis for odor recognition, Cell 65 (1) (1991) 175–187.
[5] Y. Niimura, A. Matsui, K. Touhara, 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 (9) (2014) 1485–1496.
[6] C.A. de March, S.-K. Kim, S. Antonczak, W.A. Goddard III, J. Golebiowski, G protein-coupled odorant receptors: from sequence to structure, Protein Sci. 24 (9) (2015) 1543–1548.
[7] G. Launay, G. Sanz, E. Pajot-Augy, J.-F. Gibrat, Modeling of mammalian olfactory receptors and docking of odorants, Biophys. Rev. 4 (3) (2012) 255–269.
[8] T. Tesileanu, S. Cocco, R. Monasson, V. Balasubramanian, Adaptation of olfactory receptor abundances for efficient coding, Elife 8 (2019) e39279.
[9] J.G. White, E. Southgate, J.N. Thomson, S. Brenner, The structure of the nervous system of the nematode caenorhabditis elegans, Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 314 (1165) (1986) 1–340.
[10] T. Hirotsu, H. Sonoda, T. Uozumi, Y. Shinden, K. Mimori, Y. Maehara, N. Ueda, M. Hamakawa, A highly accurate inclusive cancer screening test using caenorhabditis elegans scent detection, PLoS One 10 (3) (2015) e0118699.
[11] M. Nicoletti, A. Loppini, L. Chiodo, V. Folli, G. Ruocco, S. Filippi, Biophysical modeling of c. elegans neurons: single ion currents and whole-cell dynamics of awcon and rmd, PLoS One 14 (7) (2019) e0218738.
[12] Q. Liu, P.B. Kidd, M. Dobosiewicz, C.I. Bargmann, C. elegans awa olfactory neurons fire calcium-mediated all-or-none action potentials, Cell 175 (1) (2018) 57–70.
[13] E. Mirzakhalili, B.I. Epureanu, E. Gourgou, A mathematical and computational model of the calcium dynamics in caenorhabditis elegans ash sensory neuron, PLoS One 13 (7) (2018) e0201302.
[14] M. Usuyama, C. Ushida, R. Shingai, A model of the intracellular response of an olfactory neuron in caenorhabditis elegans to odor stimulation, PLoS One 7 (8) (2012) e42907.
[15] H. Liu, J. Sun, J. Guan, J. Zheng, S. Zhou, Improving compound–protein interaction prediction by building up highly credible negative samples, Bioinformatics 31 (12) (2015) i221–i229.
[16] B. Vidal, U. Aghayeva, H. Sun, C. Wang, L. Glenwinkel, E.A. Bayer, O. Hobert, An atlas of caenorhabditis elegans chemoreceptor expression, PLoS Biol. 16 (1) (2018) e2004218.
[17] P. Sengupta, J.H. Chou, C.I. Bargmann, Odr-10 encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl, Cell 84 (6) (1996) 899–909.
[18] G. Taniguchi, T. Uozumi, K. Kiriyama, T. Kamizaki, T. Hirotsu, Screening of odorant-receptor pairs in caenorhabditis elegans reveals different receptors for high and low odor concentrations, Sci. Signal. 7 (323) (2014) ra39.
[19] Uniprot: the universal protein knowledgebase, Nucleic Acids Res. 45 (D1) (2016) D158–D169.
[20] R. Apweiler, A. Bairoch, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, et al., Uniprot: the universal protein knowledgebase, Nucleic Acids Res. 32 (suppl_1) (2004) D115–D119.
[21] J. Zhang, J. Yang, R. Jang, Y. Zhang, Gpcr-i-tasser: a hybrid approach to g protein-coupled receptor structure modeling and the application to the human genome, Structure 23 (8) (2015) 1538–1549.
[22] S. Wu, Y. Zhang, Lomets: a local meta-threading-server for protein structure prediction, Nucleic Acids Res. 35 (10) (2007) 3375–3382.
[23] H.E. Pence, A. Williams, Chemspider: An Online Chemical Information Resource, (2010).
[24] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P.A. Thiessen, B. Yu, et al., Pubchem 2019 update: improved access to chemical data, Nucleic Acids Res. 47 (D1) (2018) D1102–D1109.
[25] H. Tsujikawa, K. Sato, C. Wei, G. Saad, K. Sumikoshi, S. Nakamura, T. Terada, K. Shimizu, Development of a protein–ligand-binding site prediction method based on interaction energy and sequence conservation, J. Struct. Funct. Genom. 17 (2–3) (2016) 39–49.
[26] H. Hwang, F. Dey, D. Petrey, B. Honig, Structure-based prediction of ligand–protein interactions on a genome-wide scale, Proc. Natl. Acad. Sci. 114 (52) (2017) 13685–13690.
[27] K. Malde, The effect of sequence quality on sequence alignment, Bioinformatics 24 (7) (2008) 897–900.
[28] R. Durbin, S.R. Eddy, A. Krogh, G. Mitchison, Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids, Cambridge university press, 1998.
[29] G.P. Raghava, G.J. Barton, Quantification of the variation in percentage identity for protein sequence alignments, BMC Bioinform. 7 (1) (2006) 415.
[30] O. Trott, A.J. Olson, Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2) (2010) 455–461.
[31] A. Vina, Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading trott, oleg; olson, Arthur j, J. Comput. Chem. 31 (2) (2010) 455–461.
[32] Y. Zhang, I-tasser server for protein 3d structure prediction, BMC Bioinform. 9 (1) (2008) 40.
[33] A. Roy, A. Kucukural, Y. Zhang, I-tasser: a unified platform for automated protein structure and function prediction, Nat. Protoc. 5 (4) (2010) 725.
[34] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The i-tasser suite: protein structure and function prediction, Nat. Methods 12 (1) (2015) 7.
[35] C.I. Bargmann, Chemosensation in c. elegans, WormBook: The Online Review of *C. elegans* Biology [Internet], WormBook (2006).
[36] E. Afergan, Y. Najajreh, D. Gutman, H. Epstein, O. Elmalak, G. Golomb, 31p-Nmr and Differential Scanning Calorimetry Studies for Determining Vesicles Drug Physical State and Fraction in Alendronate Liposomes, (2010).
[37] D. Sherwood, J. Cooper, Crystals, X-Rays and Proteins: Comprehensive Protein Crystallography, OUP Oxford, 2010.
[38] O. Man, Y. Gilad, D. Lancet, Prediction of the odorant binding site of olfactory receptor proteins by human–mouse comparisons, Protein Sci. 13 (1) (2004) 240–254.
[39] J.C. Mobarec, R. Sanchez, M. Filizola, Modern homology modeling of g-protein coupled receptors: which structural template to use? J. Med. Chem. 52 (16) (2009) 5207–5216.
[40] T. Mirzadegan, G. Benkö, S. Filipek, K. Palczewski, Sequence analyses of g-protein-coupled receptors: similarities to rhodopsin, Biochemistry 42 (10) (2003) 2759–2767.
[41] G. Pándy-Szekeres, C. Munk, T.M. Tsonkov, S. Mordalski, K. Harpsøe, A.S. Hauser, A.J. Bojarski, D.E. Gloriam, Gpcrdb in 2018: adding gpcr structure models and ligands, Nucleic Acids Res. 46 (D1) (2017) D440–D446.
[42] M. Esguerra, A. Siretskiy, X. Bello, J. Sallander, H. Gutiérrez-de Terán, Gpcrmodsim: a comprehensive web based solution for modeling g-protein coupled receptors, Nucleic Acids Res. 44 (W1) (2016) W455–W462.
[43] P. Miszta, P. Pasznik, J. Jakowiecki, A. Sztyler, D. Latek, S. Filipek, Gpcrm: a homology modeling web service with triple membrane-fitted quality assessment of gpcr models, Nucleic Acids Res. 46 (W1) (2018) W387–W395.
[44] C.L. Worth, F. Kreuchwig, J.K. Tiemann, A. Kreuchwig, M. Ritschel, G. Kleinau, P.W. Hildebrand, G. Krause, Gpcr-ssfe 2.0a fragment-based molecular modeling web tool for class a g-protein coupled receptors, Nucleic Acids Res. 45 (W1) (2017) W408–W415.
[45] J. Zhang, Y. Zhang, Gpcrrd: G protein-coupled receptor spatial restraint database for 3d structure modeling and function annotation, Bioinformatics 26 (23) (2010) 3004–3005.
[46] B. Malnic, J. Hirono, T. Sato, L.B. Buck, Combinatorial receptor codes for odors, Cell 96 (5) (1999) 713–723.
[47] G. Sanz, C. Schlegel, J.-C. Pernollet, L. Briand, Comparison of odorant specificity of two human olfactory receptors from different phylogenetic classes and evidence for antagonism, Chem. Senses 30 (1) (2005) 69–80.
[48] D.G. Laing, P. Legha, A.L. Jinks, I. Hutchinson, Relationship between molecular structure, concentration and odor qualities of oxygenated aliphatic molecules, Chem. Senses 28 (1) (2003) 57–69.
[49] T. Salthammer, Very volatile organic compounds: an understudied class of indoor air pollutants, Indoor Air 26 (1) (2016) 25–38.
[50] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, C. Friedman, Drugdrug interaction through molecular structure similarity analysis, J. Am. Med. Inform. Assoc. 19 (6) (2012) 1066–1074.
[51] J. Hert, M.J. Keiser, J.J. Irwin, T.I. Oprea, B.K. Shoichet, Quantifying the relationships among drug classes, J. Chem. Inf. Model. 48 (4) (2008) 755–765.
[52] M.J. Keiser, B.L. Roth, B.N. Armbruster, P. Ernsberger, J.J. Irwin, B.K. Shoichet, Relating protein pharmacology by ligand chemistry, Nat. Biotechnol. 25 (2) (2007) 197.