Machine Learning-based Modeling of Olfactory Receptors: Human OR51E2 as a Case Study

Mercedes Alfonso-Prieto[*,†] and Riccardo Capelli[*,‡]

†*Computational Biomedicine, Institute for Advanced Simulation IAS-5/Institute for Neuroscience and Medicine INM-9, Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, D-52428 Jülich, Germany*

‡*Dipartimento di Bioscienze, Università degli Studi di Milano, Via Celoria 26, I-20133 Milan, Italy*

E-mail: m.alfonso-prieto@fz-juelich.de; riccardo.capelli@unimi.it

Abstract

Atomistic-level investigation of olfactory receptors (ORs) is a challenging task due to the experimental/computational difficulties in the structural determination/prediction for members of this family of G-protein coupled receptors. Here we have developed a protocol that performs a series of molecular dynamics simulations from a set of structures predicted *de novo* by recent machine learning algorithms and apply it to a well-studied receptor, the human OR51E2. Our study demonstrates the need for extensive simulations to refine and validate such models. Furthermore, we demonstrate the need for the sodium ion at a binding site near $D^{2.50}$ and $E^{3.39}$ to stabilize the inactive state of the receptor. Considering the conservation of these two acidic residues across human ORs, we surmise this requirement also applies to the other ∼400 members of this family.

Olfactory receptors (ORs) are a family of G protein-coupled receptors (GPCRs) that plays a crucial role in the sense of smell.[1] The human genome encodes for approximately

800 GPCRs, out of which 50% are ORs.[2] Although initially identified in the nose, ORs are expressed in different parts of the body.[3] The investigation of the physiological roles of these extranasal ORs, as well as their possible involvement in pathological conditions, is attracting a growing interest.[4] Moreover, given that GPCRs are the target of ∼35% of FDA-approved drugs[5,6] and the wide range of odorants and other biologically active molecules binding to ORs,[7,8] these receptors are being explored as potential novel drug targets.[9,10] However, the lack of high-resolution structures for ORs has hindered the understanding of their functional mechanisms and the development of OR-targeting drugs.

Recently, the field of computational biology has made significant strides in protein structure prediction, following the development of AlphaFold2,[11] a deep learning(DL)-based algorithm that can predict the 3D structures of proteins from their amino acid sequences with high accuracy. The success of AlphaFold2 and other machine learning(ML)-based algorithms has provided a powerful tool to study protein structure and function.[12–14] Nonetheless, structural prediction of GPCRs, including ORs, still presents challenges. In particular, the algorithm predicts a single structure, despite multiple conformational states are possible for GPCRs,[15,16] and higher average confidence scores are obtained for proteins with close homologs in the training PDB set,[17] which is not the case for ORs.

To verify the reliability of an out-of-the-box *in silico* approach to predict OR structures and dynamics, we tested a set of models generated with six different predictors, followed by sub-microsecond molecular dynamics (MD) simulations. We chose to focus on the human olfactory receptor 51E2 (hOR51E2), associated to prostate cancer, because it has been widely studied, both experimentally and computationally (using a homology model).[18] Based on our test case, we propose a protocol to build reliable models of inactive, sodium-bound OR structures.

Structure prediction. We started our protocol by generating a set of six structural models of hOR51E2 via homology modeling and ML-based prediction algorithms. For ho-

mology modeling, we relied on the SwissModel (SM) web server,[19] while for ML-based prediction, we considered AlphaFold (AF),[11,20] RoseTTAFold (RF),[21] OmegaFold (OF),[22] and ESMFold (EF).[23] As a last candidate, we considered a model of the receptor in its inactive state (AF_{in}), generated with AlphaFold-MultiState.[15,24] For all the predictors considered, we tried as much as possible to use the models already available to the public (*i.e.,* without directly using the ML algorithm or modifying the default parameters – see details in the Methods section).

Models without Sodium ion in its Binding Site. For the first set of MD runs, we submitted the aforementioned starting configurations (embedded in a POPC lipid bilayer and surrounded by a water solution with 150 mM NaCl) as set up by the CHARMM-GUI[25] webserver (see the Methods section and the Supporting Information). During the equilibration protocol, while the receptor and the membrane configurations were maintained as expected in the presence of restraints, when the system was left unconstrained in all cases we observed at least a partial rearrangement of the transmembrane helices and their interfaces. Interestingly, even before removal of the restraints on the protein structure, the interior of the receptor is flooded with water molecules passing from the intracellular to the extracellular part (Figure 1). During the last 500 ns of unconstrained simulation, the amount of flowing water increases, destabilizing the interaction network that keeps TM6 and TM7 close together, thus increasing the spacing (from $7-9$ Å to $13-15$ Å) and finally breaking the helical bundle fold.

One notable exception was represented by the SM structure. After ∼200 ns of unrestrained simulation, a sodium ion bound to the receptor, occupying the known ion binding pocket close to $D69^{2.50}$; $E110^{3.39}$ also coordinates the Na^+ ion. After this event, the structure appeared much more stable (despite the already broken fold). This suggests that a sodium-bound inactive structure might be more stable than an ion-devoid configuration and thus make us consider the possibility of positioning such an ion in the Na^+ binding pocket from the beginning of the MD protocol.

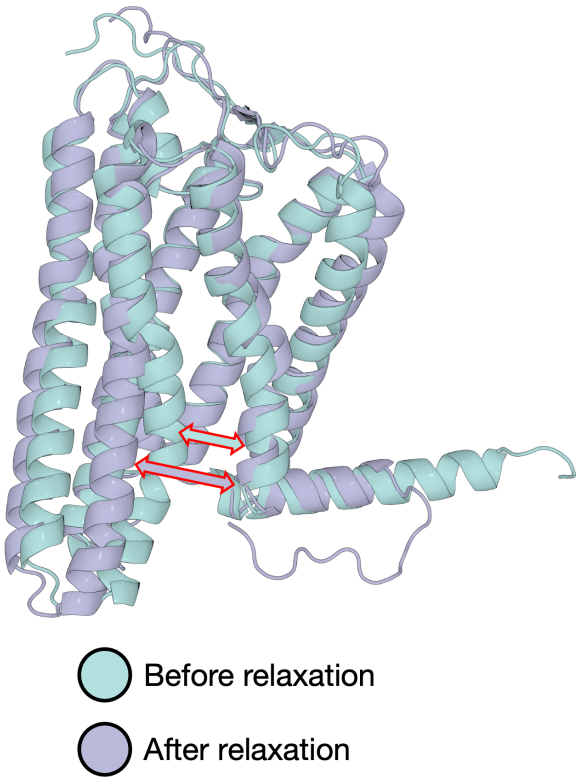


Figure 1: Opening of the TM6-TM7 interface in absence of sodium. In all the simulations without a sodium ion bound to the receptor, the interface between TM6 and TM7 is disrupted (red-contoured arrows) and thus the receptor changes between a closed conformation (cyan) at the end of the restrained equilibration to a completely open conformation (lilac) after unrestrained MD.

Models with Sodium in its Binding Site. In the second set of runs, we followed the same protocol but positioning a sodium ion close to the ion binding pocket in vicinity of D69$^{2.50}$. In all the six simulations, we observed conservation of the initial fold, with an RMSD of the heavy atoms around 5 Å (see Figure S2 in the SI), and the inactive conformation is preserved, as shown by the A^{100} descriptor[26] (see Figure S1 in the SI). Despite this qualitative change in the stability of the fold compared to the simulations without bound Na^+, the sodium-bound simulations started from AF, EF, RF, and SM configurations still showed a passage of water from the intracellular to the extracellular part through the receptor, that disrupts the interface between the transmembrane helices, mainly stabilized by hydrophobic interactions.

Instead, considering the AF_{in} and the OF simulations, water did not pass from the intracellular part to the transmembrane part of the receptor, maintaining the initially predicted fold and the TM6-TM7 distance through the whole 500 ns simulations.

To highlight differences and similarities in the fold suggested by different structure predictors, we performed a cluster analysis of the simulations. In particular, we concatenated all the MD trajectories and calculated the reciprocal RMSD of all the frames (Figure 2), considering the heavy atoms of the transmembrane helices only and ignoring the extra- and intracellular loops, which are less stable and usually predicted with a smaller confidence.[27–29] Then, clustering was performed with the gromos method[30] using an RMSD cutoff of 2 Å. From the cluster analysis we can make two observations: (i) the two most stable models

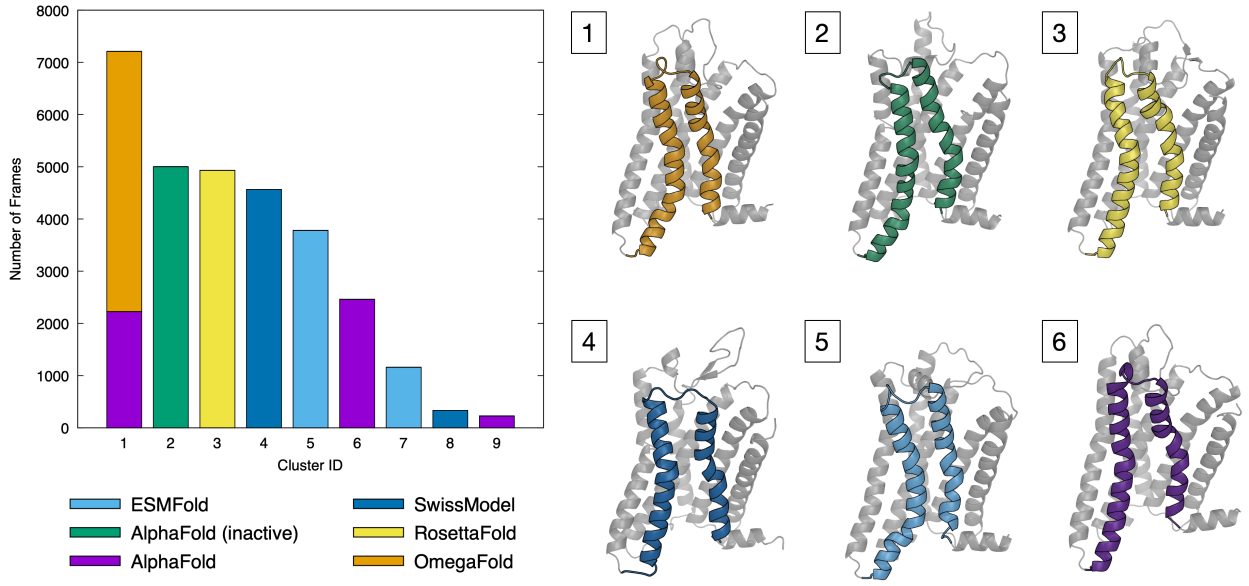


Figure 2: Cluster analysis population results and the most representative structures for the sodium bound simulations started from different initial receptor configurations.

–AF_{in} and OF– belong to different clusters (1 and 2 in Figure 2); (ii) the histogram shows no overlap between the different source structures (with the exception of cluster 1, where the initial part of AF and the OF trajectories are classified together). Therefore, the different structure prediction methods return significantly different conformations in the transmembrane part of the receptor, that should be the most stable (and thus less prone to errors in

the structure prediction). Interestingly, AF (which unfolds during the simulation) overlaps, at least in part, with the conformations sampled in the OF simulation (see cluster 1 in Figure 2). Visual inspection of the trajectory of the AF model reveals that, for the first 100 ns, the initial TM6-TM7 conformation is preserved (similarly to OF); afterwards, it widens up, losing its fold (passing from cluster 1 to clusters 6 and 9 in Figure 2).

The most evident change between the two best candidates, AF_{in} and OF, is the different structural alignment of the TM6-TM7 interface, as shown by the corresponding representative structures in Figure 2 (panels 1 and 2). Analysis of the contact map of the cluster centroids (via the webserver MAPIYA[31] (https://mapiya.lcbio.pl)) reveals a clear shift in the non-bonded (mainly hydrophobic) interactions that stabilize the TM6-TM7 interface in the two models (see Figure 3). The TM6 sequence is half helical turn behind in the AF_{in} model with respect to OF, whereas the TM7 helix is similar in both models. As a result, a mismatch between opposing amino acid pairs occurs at the TM6-TM7 interface of the OF model. In

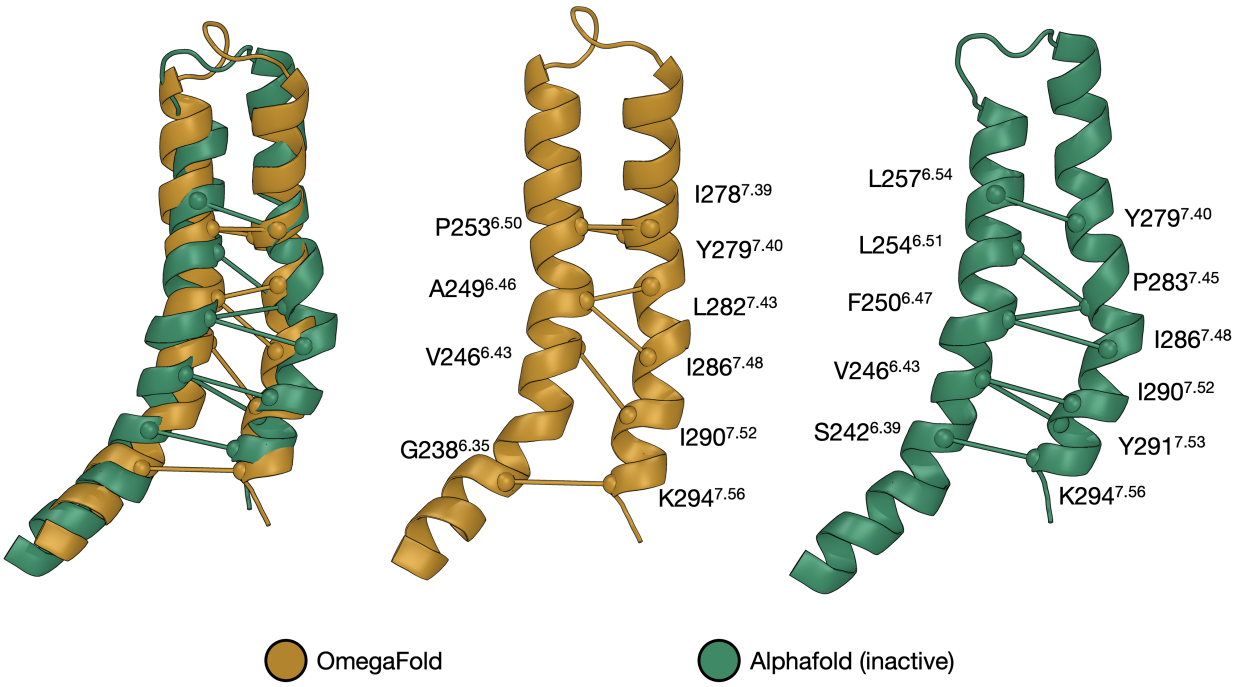


Figure 3: Comparison of the TM6-TM7 interface in AF_{in} and OF.

particular, the AF_{in} and OF structures have almost the same TM7 residues involved in the

interhelical contacts (Y279$^{7.40}$, I286$^{7.48}$, I290$^{7.52}$), while for TM6 the only residue identified as interacting in both models is V246$^{6.43}$. From a practical point of view, TM6 appears to be shifted by 1-2 residues in the structural alignment of the two models, similarly to what was observed in a recent work on another chemosensory receptor, TAS2R14,[32] when comparing two models built with AlphaFold and I-TASSER, respectively.[33]
This observation exemplifies that, although the global differences between structures generated with different predictors might seem minimal, small local differences can still result in significant changes and thus misleading predictions regarding structure-function relationships.

In conclusion, we set up a protocol to equilibrate and test models of olfactory receptors in their inactive state embedded in a POPC membrane. We can highlight four main observations from the protocol: (i) as already suggested in previous works,[34,35] the stability and reliability of structures should be tested extensively by MD simulations; (ii) *de novo* structural determination can lead to significantly different predictions in presence of a multi-state system (see AF vs. AF_{in}); (iii) the absence of conserved motifs between human ORs and other class A GPCRs, especially for TM6,[36] can lead to gross errors in structure reconstruction; and (iv) for hOR51E2 in its inactive state –but this is most probably valid for a large set of ORs– the presence of sodium in its binding pocket is crucial for the stabilization of its fold. Sodium binding to hOR51E2 can be attributed to the presence of two negatively charged residues, D69$^{2.50}$ and E110$^{3.39}$. The first one is a common site for ion binding in class A GPCRs, while the second one is usually occupied by Ser in non-olfactory class A GPCRs.[37] Instead, 93% of human ORs contain Asp/Glu at positions 2.50 and 3.39 (see Supporting Information Figure S3). Residue conservation in these sites can suggest a coevolutionary feature[38] supporting the structural stability role of this ion binding, as empirically observed by us. In line with this hypothesis, the presence of sodium in that position is also foreseen for hOR51E2 by the ML-based protein-ligand binding predictor, AlphaFill[39] (see

https://alphafill.eu/model?id=Q9H255). As a further indirect validation, a recent experimental work[40] showed that mutation of $E^{3.39}$ dramatically enhances the *in vitro* expression of olfactory receptors, further supporting the structural and functional importance of this residue. As pointed out by a recent commentary,[41] *de novo* structure determination is dramatically limited by the "single answer problem": the algorithms return a single structure that is, following the training, the most probable candidate. From a general point of view, this can be correct only in the realm of single-state proteins, while here (and in the majority of the biologically-relevant cases) our target GPCR has a set of different conformational states. This problem can be solved (or attenuated) taking particular care of the structural knowledge that the algorithm employs to perform its prediction. In the Heo and Feig[15] or del Alamo *et al.*[16] approaches, this is accomplished by limiting the training set to a single state (here GPCR experimental structures annotated to be in the inactive state), to maximize the chances of a correct prediction. The majority of the *de novo* structure determination algorithms need a properly aligned multiple sequence alignment (MSA). In the case of GPCRs, and in particular ORs, special care needs to be taken when generating such MSA, especially in the absence of known conserved motifs. In particular, ORs lack the "rotamer toggle switch" involving $W^{6.48}$ sequence on helix TM6 present in non-olfactory class A GPCRs,[36] which results in different structural predictions. Some of these models seem to be not good enough, as evidenced by the stability (or lack thereof) of the predicted fold of the system. One possible way to overcome this problem can be represented by the use of manually-curated MSA based also on structural information and/or in the training of ML weights to target specific GPCRs subfamilies in their structure predictions.

Finally, during the writing of the manuscript, a preprint was published[42] presenting the first ever human OR structure determined experimentally, and coincidentally the structure corresponds to a construct based on the hOR51E2 sequence. Unfortunately, as of February 2023, the coordinates have not yet been deposited to the Protein Data Bank and thus we cannot make a direct comparison between our AF^{in} or OF inactive models and the agonist-

bound cryo-EM structure. In the future we plan to integrate this structural information to check the current findings and possibly build a better MSA for future structural predictions.

Methods

Initial structures generation

The hOR51E2 sequence was taken from UniProt[43] (accession code: Q9H255, amino acids 1-320). Such sequence was employed as input for all the structure prediction algorithms listed below.

AlphaFold (AF). We obtained the initial coordinates (accession code AF-Q9H255-F1, v3) from the AlphaFold Protein Structure Database[20] (https://alphafold.ebi.ac.uk); such model was generated with the vanilla version Alphafold.[11]

AlphaFold, inactive state ($\mathbf{AF_{in}}$). AlphaFold weights were trained on all experimental structures available at the PDB, regardless of their conformation in the case of multi-state proteins. Heo and Feig recently proposed[15] to perform structure predictions considering multiple sequence alignments built on state-annotated data (*i.e.,* using only active or inactive conformations). Since our intention was to model the apo receptor (*i.e.,* without ligands), we therefore chose the inactive conformation. In practice, such model was obtained from GPCRdb[24] (https://gpcrdb.org), which maintains an updated database of GPCR models generated with AlphaFold-MultiState.

RoseTTAFold (RF). The structure predicted by RoseTTAFold[21] was obtained via the web server of the Baker Lab, ROBETTA (https://robetta.bakerlab.org).

OmegaFold (OF). The OmegaFold[22] prediction was run on local resources using the public implementation available on GitHub (https://github.com/HeliXonProtein/OmegaFold), release v.1.1.0.

ESMFold (EF). The structure from ESMFold[23] was obtained from the python notebook available through the ColabFold interface (https://github.com/sokrypton/ColabFold).

SwissModel (SM). A homology model was retrieved from the SwissModel repository.[19] The template used was the human neuropeptide Y receptor type 2 (PDB structure 7X9B, chain D), which has a sequence identity of 18.12% with hOR51E2. Unfortunately, hORs exhibit a low sequence identity ($< 20\%$) with other class A GPCRs and thus a better template cannot be found.[36] The SM model covers residues 22-307.

System preparation

Initially, we preprocessed all the structures obtained via *de novo* and homology modeling algorithms by using the Protein Preparation Wizard implemented in Schrödinger Maestro version 2022-3,[44] which automatically assigns the amino acid protonation states. Two exceptions were represented by D69$^{2.50}$ and E110$^{3.39}$, that were kept in charged state. All the structures prepared in this way were further processed via the online interface of CHARMM-GUI[25,45] (https://charmm-gui.org/). First, we built an disulfide bond between C96$^{3.25}$ and C178$^{45.50}$, then we defined a cubic simulation box with dimensions (100 Å) $\times$ (100 Å) $\times$ (120 Å), with the receptor in the center, embedded in a POPC lipid bilayer. The bilayer and the receptor were solvated in water with a NaCl concentration of 0.15 M, in line with standard experimental and physiological conditions for GPCRs. The protein, lipids, and ions were parameterized using the Charmm36m force field,[46] while water was describe with the TIP3P[47] model.

Molecular dynamics simulations

The simulations performed here, either in presence or absence of the sodium ion in close to D$^{2.50}$, were based on an extended version of the standard CHARMM-GUI workflow (see Supporting Information). The last production step was a 500 ns-long unrestrained MD simulation with a time step of 2 fs. Velocity rescale thermostat[48] was applied to keep the temperature to 310 K, and pressure was kept at 1 bar via the cell rescale barostat,[49] with semi-isotropic coupling.

All the simulations were performed using GROMACS[50] version 2021.2 patched with PLUMED.[51,52]

Data Availability

All the input data needed to reproduce the results shown in this paper (structures, topology, GROMACS and PLUMED input files, etc.) are available at Zenodo (https://doi.org/10.5281/zenodo.7657973).

Acknowledgement

RC thanks Carlo Camilloni, Federico Ballabio, and Amit Kumawat for useful discussions. MA-P acknowledges financial support in part from the DFG Research Unit FOR2518 "Functional Dynamics of Ion Channels and Transporters – DynIon" Project P6.

Supporting Information Available

Extensive details on the simulation protocol steps, description of A^{100} and its implementation in PLUMED, figures with the time evolution of A^{100} and RMSD and with the amino acid conservation for positions 2.50 and 3.39, as well as a table with the Ballesteros-Weinstein numbering for hOR51E2, as listed in the GPCRdb.[24]

References

(1) Buck, L.; Axel, R. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* **1991**, *65*, 175–187.

(2) Fredriksson, R.; Lagerström, M. C.; Lundin, L.-G.; Schiöth, H. B. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic

Analysis, Paralogon Groups, and Fingerprints. *Molecular Pharmacology* **2003**, *63*, 1256–1272.

(3) Maßberg, D.; Hatt, H. Human Olfactory Receptors: Novel Cellular Functions Outside of the Nose. *Physiological Reviews* **2018**, *98*, 1739–1763.

(4) Drew, L. Olfactory receptors are not unique to the nose. *Nature* **2022**, *606*, S14–S17.

(5) Hauser, A. S.; Attwood, M. M.; Rask-Andersen, M.; Schiöth, H. B.; Gloriam, D. E. Trends in GPCR drug discovery: new agents, targets and indications. *Nature Reviews Drug Discovery* **2017**, *16*, 829–842.

(6) Hauser, A. S.; Chavali, S.; Masuho, I.; Jahn, L. J.; Martemyanov, K. A.; Gloriam, D. E.; Babu, M. M. Pharmacogenomics of GPCR Drug Targets. *Cell* **2018**, *172*, 41–54.e19.

(7) Saito, H.; Chi, Q.; Zhuang, H.; Matsunami, H.; Mainland, J. D. Odor Coding by a Mammalian Receptor Repertoire. *Science Signaling* **2009**, *2*.

(8) Cong, X.; Ren, W.; Pacalon, J.; Xu, R.; Xu, L.; Li, X.; de March, C. A.; Matsunami, H.; Yu, H.; Yu, Y.; Golebiowski, J. Large-Scale G Protein-Coupled Olfactory Receptor–Ligand Pairing. *ACS Central Science* **2022**, *8*, 379–387.

(9) Lee, S.-J.; Depoortere, I.; Hatt, H. Therapeutic potential of ectopic olfactory and taste receptors. *Nature Reviews Drug Discovery* **2018**, *18*, 116–138.

(10) Pizio, A. D.; Behrens, M.; Krautwurst, D. Beyond the Flavour: The Potential Druggability of Chemosensory G Protein-Coupled Receptors. *International Journal of Molecular Sciences* **2019**, *20*, 1402.

(11) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A., et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

(12) Akdel, M. et al. A structural biology community assessment of AlphaFold2 applications. *Nature Structural & Molecular Biology* **2022**, *29*, 1056–1067.

(13) Jumper, J.; Hassabis, D. Protein structure predictions to atomic accuracy with AlphaFold. *Nature Methods* **2022**, *19*, 11–12.

(14) Baek, M.; Baker, D. Deep learning and protein structure modeling. *Nature Methods* **2022**, *19*, 13–14.

(15) Heo, L.; Feig, M. Multi-State Modeling of G-protein Coupled Receptors at Experimental Accuracy. *Proteins: Structure, Function, and Bioinformatics* **2022**,

(16) Del Alamo, D.; Sala, D.; Mchaourab, H. S.; Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *Elife* **2022**, *11*, e75751.

(17) Thornton, J. M.; Laskowski, R. A.; Borkakoti, N. AlphaFold heralds a data-driven revolution in biology and medicine. *Nature Medicine* **2021**, *27*, 1666–1669.

(18) Wolf, S.; Jovancevic, N.; Gelis, L.; Pietsch, S.; Hatt, H.; Gerwert, K. Dynamical Binding Modes Determine Agonistic and Antagonistic Ligand Effects in the Prostate-Specific G-Protein Coupled Receptor (PSGR). *Scientific Reports* **2017**, *7*, 16007.

(19) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **2018**, *46*, W296–W303.

(20) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A., et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research* **2022**, *50*, D439–D444.

(21) Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876.

(22) Wu, R.; Ding, F.; Wang, R.; Shen, R.; Zhang, X.; Luo, S.; Su, C.; Wu, Z.; Xie, Q.; Berger, B., et al. High-resolution de novo structure prediction from primary sequence. *BioRxiv* **2022**,

(23) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y., et al. Evolutionary-scale prediction of atomic level protein structure with a language model. *bioRxiv* **2022**,

(24) Pándy-Szekeres, G.; Caroli, J.; Mamyrbekov, A.; Kermani, A. A.; Keserű, G. M.; Kooistra, A. J.; Gloriam, D. E. GPCRdb in 2023: state-specific structure models using AlphaFold2 and new ligand resources. *Nucleic Acids Research* **2022**,

(25) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry* **2008**, *29*, 1859–1865.

(26) Ibrahim, P.; Wifling, D.; Clark, T. Universal activation index for class A GPCRs. *Journal of Chemical Information and Modeling* **2019**, *59*, 3938–3945.

(27) Kufareva, I.; Katritch, V.; Stevens, R. C.; Abagyan, R. Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. *Structure* **2014**, *22*, 1120–1139.

(28) Lee, C.; Su, B.-H.; Tseng, Y. J. Comparative studies of AlphaFold, RoseTTAFold and Modeller: a case study involving the use of G-protein-coupled receptors. *Briefings in Bioinformatics* **2022**, *23*.

(29) Nicoli, A.; Haag, F.; Marcinek, P.; He, R.; Kreißl, J.; Stein, J.; Marchetto, A.; Dunkel, A.; Hofmann, T.; Krautwurst, D.; Pizio, A. D. Modeling the Orthosteric Bind-

ing Site of the G Protein-Coupled Odorant Receptor OR5K1. *Journal of Chemical Information and Modeling* **2023**,

(30) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Van Gunsteren, W. F.; Mark, A. E. Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition* **1999**, *38*, 236–240.

(31) Badaczewska-Dawid, A. E.; Nithin, C.; Wroblewski, K.; Kurcinski, M.; Kmiecik, S. MAPIYA contact map server for identification and visualization of molecular interactions in proteins and biological complexes. *Nucleic Acids Research* **2022**, *50*, W474–W482.

(32) Fierro, F.; Peri, L.; Hübner, H.; Tabor-Schkade, A.; Waterloo, L.; Löber, S.; Pfeiffer, T.; Weikert, D.; Dingjan, T.; Margulis, E., et al. Inhibiting a promiscuous GPCR: iterative discovery of bitter taste receptor ligands. *bioRxiv* **2022**, 2022–11.

(33) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nature methods* **2015**, *12*, 7–8.

(34) Raval, A.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* **2012**, *80*, 2071–2079.

(35) Heo, L.; Arbour, C. F.; Janson, G.; Feig, M. Improved Sampling Strategies for Protein Model Refinement Based on Molecular Dynamics Simulation. *Journal of Chemical Theory and Computation* **2021**, *17*, 1931–1943.

(36) Fierro, F.; Suku, E.; Alfonso-Prieto, M.; Giorgetti, A.; Cichon, S.; Carloni, P. Agonist binding to chemosensory receptors: a systematic bioinformatics analysis. *Frontiers in molecular biosciences* **2017**, *4*, 63.

(37) Zarzycka, B.; Zaidi, S. A.; Roth, B. L.; Katritch, V. Harnessing Ion-Binding Sites for GPCR Pharmacology. *Pharmacological Reviews* **2019**, *71*, 571–595.

(38) Baldessari, F.; Capelli, R.; Carloni, P.; Giorgetti, A. Coevolutionary data-based interaction networks approach highlighting key residues across protein families: The case of the G-protein coupled receptors. *Computational and Structural Biotechnology Journal* **2020**, *18*, 1153–1159.

(39) Hekkelman, M. L.; de Vries, I.; Joosten, R. P.; Perrakis, A. AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nature Methods* **2022**, *20*, 205–213.

(40) Fukutani, Y.; Nakamura, Y.; Muto, N.; Miyanaga, S.; Kanemaki, R.; Ikegami, K.; Noguchi, K.; Ohsawa, I.; Matsunami, H.; Yohda, M. Hot Spot Mutagenesis Improves the Functional Expression of Unique Mammalian Odorant Receptors. *International Journal of Molecular Sciences* **2021**, *23*, 277.

(41) Lane, T. J. Protein structure prediction has reached the single-structure frontier. *Nature Methods* **2023**, 1–4.

(42) Billesboelle, C. B.; de March, C. A.; van der Velden, W. J.; Ma, N.; Tewari, J.; Linas del Torrent, C.; Li, L.; Faust, B.; Vaidehi, N.; Matsunami, H., et al. Structural basis of odorant recognition by a human odorant receptor. *bioRxiv* **2022**, 2022–12.

(43) The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* **2020**, *49*, D480–D489.

(44) Schrödinger, LLC, Maestro 2022-4, 2022; New York, NY.

(45) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y., et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations us-

ing the CHARMM36 additive force field. *Journal of chemical theory and computation* **2016**, *12*, 405–413.

(46) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* **2017**, *14*, 71–73.

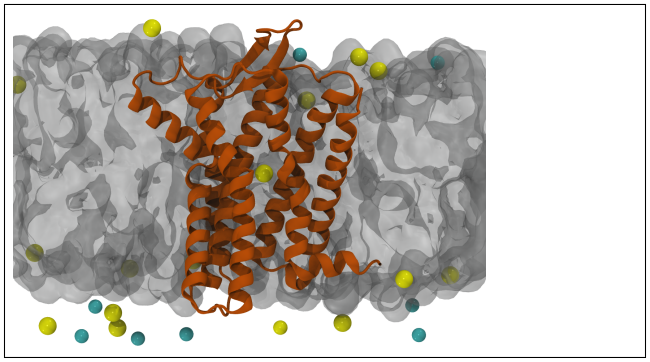
(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, *79*, 926–935.

(48) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **2007**, *126*, 014101.

(49) Bernetti, M.; Bussi, G. Pressure control using stochastic cell rescaling. *The Journal of Chemical Physics* **2020**, *153*, 114107.

(50) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(51) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer physics communications* **2014**, *185*, 604–613.

(52) Bonomi, M. et al. Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods* **2019**, *16*, 670–673.

TOC Graphic