



# Molecular simulation-based 3D structural construction of olfactory receptor with agonist binding

Takumi Hirao<sup>1</sup> · Yusuke Ihara<sup>2</sup> · Chiori Ijichi<sup>2</sup> · Genki Kudo<sup>3</sup> · Ryunosuke Yoshino<sup>4,5</sup> · Takatsugu Hirokawa<sup>4,5</sup>

Received: 2 July 2025 / Accepted: 27 October 2025  
© The Author(s) 2025

## Abstract

Olfactory receptors (ORs) form the largest subfamily of class A G protein-coupled receptors (GPCRs); however, only a few 3D structures of ORs have been determined. Structure-based virtual screening and improved structural insights are required to effectively identify novel odor molecules and elucidate their binding modes along with mechanisms of activation and inactivation. Herein, we propose a protocol to provide an active state model for the target OR (OR9Q2) with agonist molecules using AlphaFold2, molecular simulations, and virtual screening. Furthermore, we extracted ligand-stable bound sections using the ligand-stable duration (LSD) protocol defined in this study to analyze conformational ensembles of complex structures. We constructed promising complex structures and demonstrated their reliability by calculating the area under the receiver operating characteristic (ROC) curve in virtual screening tests using experimentally validated active and inactive compounds. This study offers a reliable structure-based screening protocol for olfactory receptors, which can subsequently aid novel odorant discovery and advancing fragrance, flavour, and biosensor industries.

**Keywords** Olfactory receptor · Olfactory receptor 9Q2 · Molecular dynamics simulation · Docking simulation · Active state · AlphaFold2

## Abbreviations

|                    |                                                        |
|--------------------|--------------------------------------------------------|
| AF2                | AlphaFold2                                             |
| AF2 <sup>ebi</sup> | AlphaFold2 ebi model                                   |
| AF2 <sup>tpl</sup> | AlphaFold2 with template model                         |
| AUC                | area under the curve                                   |
| AUROC              | area under the receiver operating characteristic curve |
| FMS                | pharmacophore matching similarity                      |
| GPCR               | G protein-coupled receptors                            |
| hRLuc              | luminescence of Renilla luciferase                     |
| LSD                | ligand-stable duration                                 |
| Luc                | luminescence of firefly luciferase                     |
| OR                 | olfactory receptor                                     |
| OR9Q2              | olfactory receptor family 9 subfamily Q member 2       |
| PDB                | protein data bank                                      |
| PLIF               | protein-ligand interaction fingerprint                 |
| POPC               | pure 1-palmitoyl-2-oleyl-sn-glycero-3-phosphocholine   |
| RESP               | restrained electrostatic potential procedure           |
| RMSD               | root mean square deviation                             |
| ROC                | receiver operating characteristic                      |

✉ Takatsugu Hirokawa  
t-hirokawa@md.tsukuba.ac.jp

Takumi Hirao  
s2430407@u.tsukuba.ac.jp

- <sup>1</sup> Doctoral Program in Medical Sciences, Degree Programs in Comprehensive Human Sciences, Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba 305-8575, Ibaraki, Japan
- <sup>2</sup> Institute of Food Sciences and Technologies, Food Products Division, Ajinomoto Co., Inc, Tokyo, Japan
- <sup>3</sup> Physics Department, Graduate School of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8571, Ibaraki, Japan
- <sup>4</sup> Division of Biomedical Science, Institute of Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8575, Ibaraki, Japan
- <sup>5</sup> Transborder Medical Research Center, University of Tsukuba, 1-1-1 Tennodai, 14 Ibaraki, Tsukuba 305-8575, Japan

TM transmembrane

## Introduction

G protein-coupled receptors (GPCRs) are membrane proteins comprising seven transmembrane (TM) helices. GPCRs are vital in cell signaling processes and respond to various stimuli, such as hormones, neurotransmitters, light, and odors [1, 2]. GPCRs can detect diverse ligands and proteins that control cell signaling [3]. Human GPCRs constitute a family of more than 800 members categorized into several classes according to their function and structure: class A (rhodopsin-like receptors), class B (secretin-like receptors), class C (metabotropic glutamate receptors), and class T (taste receptors) [4–6]. In particular, class A GPCRs comprise approximately 80% of this family, and the number of non-class A GPCRs is fewer than 100.0 [7].

GPCRs undergo substantial conformational changes between their inactive and active states. In particular, activation involves an outward movement of TM6 (toward an open conformation) and rearrangements within the ligand-binding pocket [8–10]. A conformational change occurs when an agonist binds to a GPCR, forming a complex structure with the heterotrimeric G-protein complex [11, 12]. GPCRs are the most intensively studied targets for drug discovery. Furthermore, 34% of the small molecule drugs act on GPCRs, and structure-based drug design methods are commonly used [3, 13, 14].

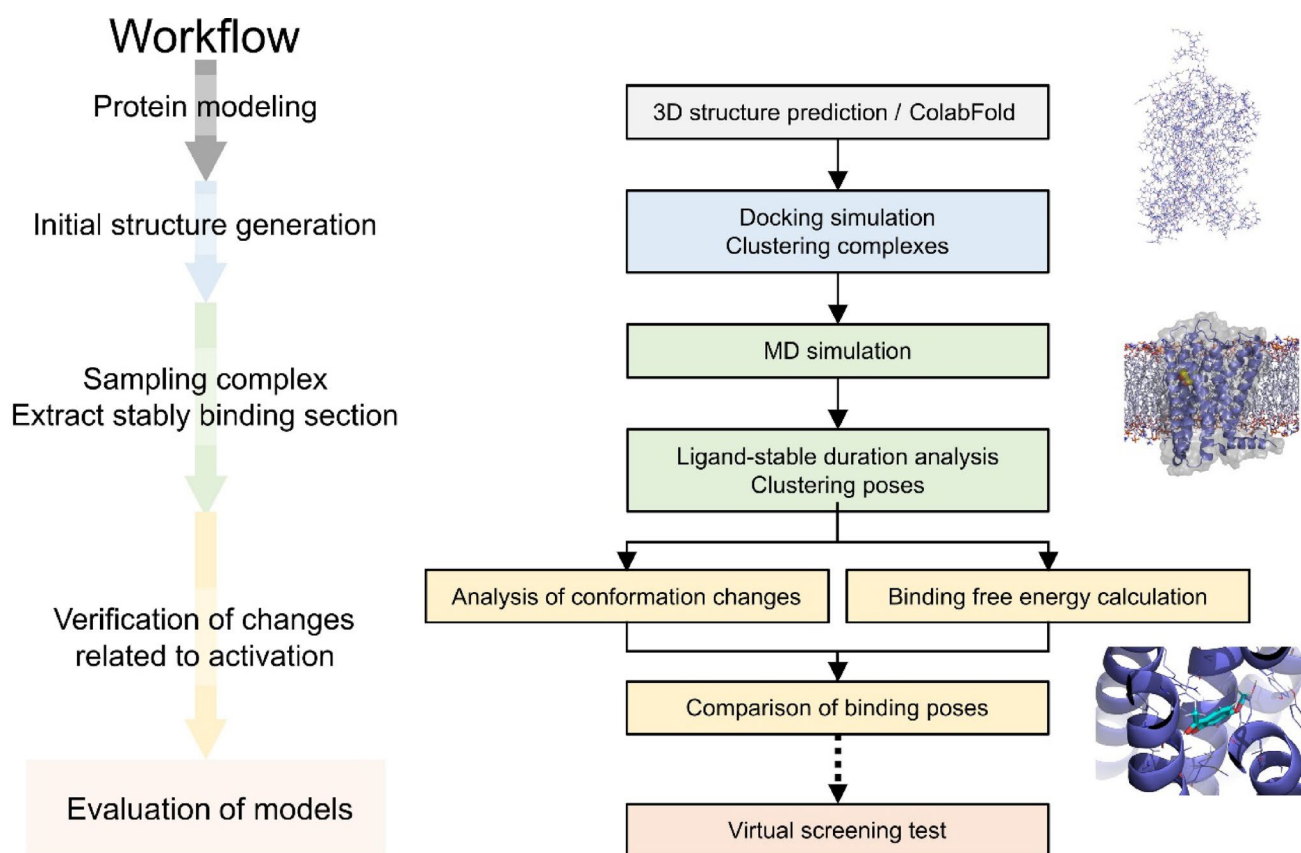
Olfactory receptors (ORs), the largest subfamily of class A GPCRs, are involved in odor molecule perception. Approximately 400 genes encode human ORs [15, 16]. Recently, the 3D structure of OR51E2, the first OR structure, was determined by cryo-electron microscopy by Billesbølle et al. [17], and further studies are expected to accelerate the molecular-level analysis of ORs. Humans recognize odor molecules as combined signals from multiple ORs that accept odor molecules of diverse chemotypes or modes of action [18–21]. Although several experimental studies have assigned odor molecules to ORs, few agonists and antagonists are known. Therefore, in addition to drug discovery research on non-OR GPCRs, *in silico* methods can be used to identify novel odor molecules.

Recently, AlphaFold2 (AF2) [22] was used to predict protein structures for which high precision using conventional homology modeling was challenging. Critical Assessment of Structure Prediction 14 (CASP14) and other previous studies have shown that AF2 predicts protein structure remarkably [23–25]. Although AF2 is expected to be useful for structural prediction-based discovery of novel odor molecules, the AF2 model is unsuitable for direct use as a structural model for virtual screening with docking

simulations [26–28]. In particular, AF2-predicted structures are less accurate than experimentally obtained structures for ligand-binding sites, because AF2 cannot predict the structure of the protein-ligand complex. Therefore, predicting the appropriate state in AF2 and optimizing the structure is important to obtain a structure suitable for virtual screening. On the other hand, AlphaFold3, as reported by Josh et al., enables the prediction of protein–ligand complexes with reasonable confidence and has been successfully applied to a variety of ligands [29]. In a previous study, the binding sites of OR structures were identified using AF2 and conventional homology modeling [21]. The authors used induced-fit docking for both models to refine the binding sites. Thus, the accuracy of virtual screening could be improved by using the model predicted with AF2, which refines the binding pocket toward a ligand-bound conformation and accounts for conformational perturbations introduced by MD simulations or template-based bias. Previous studies have generated the desired state by restricting AF2 training data to a certain state [30]. In this study, we propose a virtual screening model protocol for ORs. We refined the binding site of the target OR through molecular simulations using experimentally selected active compounds. We chose olfactory receptor family 9 subfamily Q member 2 (OR9Q2) as the target for validation and predicted its 3D structure using ColabFold [31]. Conformational sampling of the ligand-protein complex was performed using docking and molecular dynamics (MD) simulations. Ligand-stable duration (LSD) was defined and used to extract the section in which the ligand was stably bound based on the MD trajectory. In the extracted sections, we identified binding positions that could cause conformational changes upon activation. Virtual screening tests using representative structures showed that these models could discriminate between active and inactive (decoy) compounds to some extent. This protocol is an exploratory method for sampling protein-ligand complexes and virtual screening models for GPCR with small molecules.

## Materials and methods

Our protocol was divided into four steps, and the final models were validated using structure-activity relationships and virtual screening tests (Fig. 1). First, we modeled the 3D structure of OR9Q2. Second, a known agonist was docked into OR9Q2 as the initial complex structure. Third, MD simulations were performed using these complex structures. LSD analysis was conducted to extract the ligand-stable bound sections. Finally, the ligand-binding free energy and GPCR-specific conformational changes induced by the active odor molecules of representative LSDs were



**Fig. 1** Overview of the protocols. The figure shows the workflow of protocols. The right side of the chart shows the specific processes

analyzed. Subsequently, a virtual screening test using the active odor molecule and decoy dataset was performed on several binding-pose candidates to evaluate the accuracy of the binding models and their potential for virtual screening.

### Protein structure modeling

In this study, we modeled the 3D structure of the active state-like OR9Q2, which has no reported experimental-determined structure, evaluated its bioactivity, and studied the structure-activity relationship. The amino acid sequences of OR9Q2 (UniProt ID: Q8NGE9) and the guanine nucleotide-binding protein G(olf) subunit (UniProt ID: P38405) were obtained from the UniProt database [32]. The 3D structure was predicted to have a complex structure with the G(olf) subunit using AF2 as implemented in ColabFold (Ver. 1.3.0) [31, 33]. Multiple sequence alignment based on mmseqs2 was used to predict multimer. We performed a PDB sequence search using the OR9Q2 and G(olf) sequences and selected the melanocortin 4 receptor (MC4R) (PDB ID: 7F53 [34]) as a high-scoring template structure that forms a complex with an agonist-binding G-protein trimer. At the outset of this study, no experimentally determined olfactory receptor structures were available. In ColabFold, the active

MC4R was used as a template to predict active-like structures. The sequence identity between OR9Q2 and MC4R in the transmembrane region was 20.1%. Subsequently, the 3D structure of OR9Q2 was generated by removing G(olf) from the predicted complex using the highest pTM score model. Protonated state of the proteins was determined using the H++ server (ver. 4.0) [35–37] at pH 7.4.

### Docking simulation

We selected 2-hydroxy-4-methoxyacetophenone (compound 1) as the representative agonist (Table 1). The 3D structure of compound 1 was constructed and minimized using MOE (ver. 2022, Chemical Computing Group, LLC) [38]. The protonated state and the force field parameter for minimization were set to pH 7.0 and MMFF94x, respectively. Compound 1 was docked to the orthosteric binding site of the OR9Q2 model using DOCK6.9 [39]. The ligand-binding site was identified with reference to the typical orthosteric binding position observed in Class A GPCRs [40, 41]. The orthosteric binding site was also estimated using FTMap [42], and the centroid of the binding site was set as the grid center. The results of FTMap are shown in

**Table 1** OR9Q2 agonists and EC<sub>50</sub> values<sup>a</sup>

| Compound | Name                            | Structure | CAS       | EC <sub>50</sub> (μM) <sup>b</sup> | E <sub>max</sub> |
|----------|---------------------------------|-----------|-----------|------------------------------------|------------------|
| 1        | 2-Hydroxy-4-methoxyacetophenone |           | 552-41-0  | 6.08 ± 0.04                        | 6.42             |
| 2        | 4-Ethylphenol                   |           | 123-07-9  | 6.11 ± 0.03                        | 8.28             |
| 3        | 4-Chlorophenol                  |           | 106-48-9  | 6.40 ± 0.03                        | 8.33             |
| 4        | p-Cresol                        |           | 106-44-5  | 6.34 ± 0.03                        | 8.47             |
| 5        | 2,6-Dimethylphenol              |           | 576-26-1  | 3.56 ± 0.03                        | 7.79             |
| 6        | 3,4-Dimethylphenol              |           | 95-65-8   | 5.28 ± 0.03                        | 7.56             |
| 7        | 4-Methoxyphenol                 |           | 150-76-5  | 5.27 ± 0.04                        | 7.68             |
| 8        | 4-Isopropylphenol               |           | 99-89-8   | 5.22 ± 0.05                        | 5.75             |
| 9        | 4-Ethyl-1,3-benzenediol         |           | 2896-60-8 | 5.39 ± 0.04                        | 8.20             |
| 10       | 4-Chloro-3-methylphenol         |           | 59-50-7   | 5.78 ± 0.03                        | 7.65             |
| 11       | 3-Ethylphenol                   |           | 620-17-7  | 4.90 ± 0.03                        | 7.17             |
| 12       | 4-Ethenylphenol                 |           | 2628-17-3 | 4.40 ± 0.04                        | 7.61             |
| 13       | Phenol                          |           | 108-95-2  | 4.61 ± 0.04                        | 8.09             |

Figure S1. The binding site was evaluated using SiteMap (Schrödinger Release 2024-3: SiteMap, Schrödinger, LLC, New York, NY, 2024.) [43, 44] (Figure S2). Spheres within 5 Å around the grid center were selected, and the extra margin was set to 10 Å with grid spacing set to 0.3 Å. The force field parameters, ff14SB [45] and AM1-BCC [46], were

assigned as the charge parameters for OR9Q2 and compound **1** using Chimera [47], respectively.

The docked poses were hierarchically clustered using Ward's method based on a distance matrix calculated from the root mean square deviations (RMSDs) of heavy atoms. The poses were classified into five clusters, with a threshold value of 18.0 applied to avoid contamination by visually

flipped poses. The representative pose with the lowest grid score in each cluster was selected.

## MD simulation

MD simulations of five representative poses were performed for post-docking processing using GROMACS 2021.5 [48]. The system setup for representative poses was performed using Membrane Builder in CHARMM-GUI [49]. OR9Q2 was embedded in a pure 1-palmitoyl-2-oleyl-sn-glycero-3-phosphocholine (POPC) bilayer of approximately  $100 \times 100 \text{ \AA}^2$ . TIP3P water molecules and 0.15 M  $\text{Na}^+/\text{Cl}^-$  ions were added to each membrane face with a thickness of approximately 25  $\text{\AA}$ .

The CHARMM36m [50] force field was used for proteins and lipids, while CGenFF was used for compound **1**. The generated system was subjected to energy minimization, NVT equilibration, and NPT equilibration, following the six-step equilibration process, which gradually decreased the constraint energy, using the method proposed by Membrane Builder [51]. Different random seeds were generated during the initial equilibration stage, and no resampling of initial velocities performed in subsequent calculations. Details of the calculation conditions are provided in the Data and Software Availability section. We performed five independent 500 ns MD simulations using the NPT ensemble at 1 bar and 303.15 K. The time step was set to 2 fs and snapshots were extracted every 100 ps.

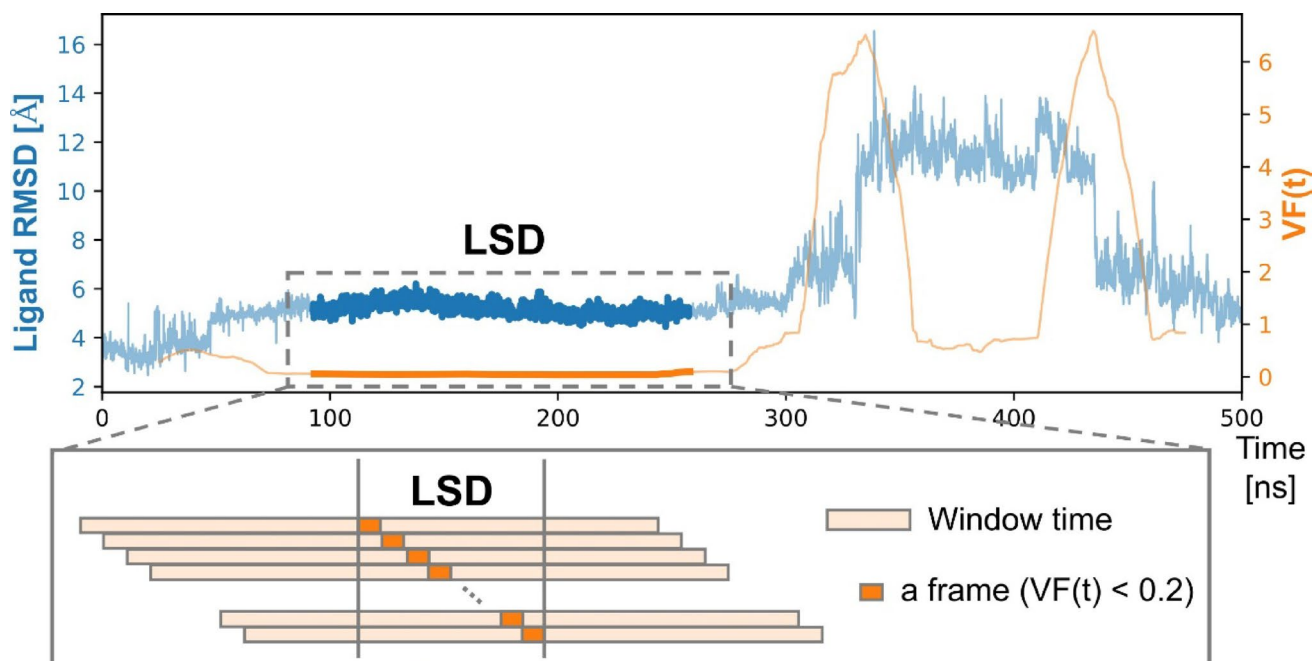
## Ligand stability sections and binding energy calculation

We proposed a new method for measuring the LSD to extract the sustainable binding state of the ligand. Specifically, the RMSD values of heavy atoms in the ligand were calculated by superimposing them onto the appropriate region of the protein. The initial structure of the production run was used as the reference structure for RMSD calculation. In this study, the Ca atoms in the TM region were used for superimposition. The RMSD variance of a frame (VF) at time  $t$  was calculated using snapshots with a sliding window of pre- and post-25 ns (a total of 501 frames).

VF values were calculated using Eq. (1):

$$VF(t) = \frac{1}{2 * window} \sum_{i=t-window}^{t+window} \left( RMSD_i - \overline{RMSD}(t) \right)^2 \quad (1)$$

$\overline{RMSD}(t)$  represents the local average of RMSD values within a symmetric time window centered at  $t$ . Based on VF [ $\text{\AA}^2$ ], we extracted the section with  $<0.2$  VF continuous for  $>100$  ns, and this section was defined as the LSD (Fig. 2). LSD indicated that compound **1** was stably bound to OR9Q2.



**Fig. 2** Diagram of ligand stability section Conception of ligand-stable duration (LSD) with a sample of LSD using real root mean square deviation (RMSD) plot and the RMSD variance of a frame (VF)

## Analysis of binding poses

In each LSD, the central structure of the ligand ensemble was selected based on the heavy-atom coordinates of compound **1**. All central structures were merged and classified into three clusters based on Ward's method for hierarchical clustering. For each LSD cluster, the mean snapshot derived from the longest LSD was selected as the representative LSD and structure. Protein-ligand interaction analysis was conducted using Maestro (Schrödinger Release 2023-1: Maestro, Schrödinger, LLC, New York, NY, 2023).

The binding free energy between OR9Q2 and **1** was calculated based on MM/PBSA [52] using gmx\_MMPBSA (ver. 1.5.7) [53] for representative LSDs. The statistical analysis was also evaluated with a 95% confidence interval (CI).

## Virtual screening

A virtual screening was performed to evaluate the performance of the active-inactive (decoy) discrimination of representative structures by docking simulation. Compounds **1–13** were used as active data, and decoy compounds (612 compounds) were generated using DUD-E “make decoy” command [54] with Compounds **1–13** as the query molecules (Table 1). The protonation states of all compounds were assigned at pH 7.0, and the force field parameter for minimization was set to MMFF94x. Docking simulations were performed using DOCK6.9, with the grid center defined by a sphere within 2.5 Å of compound **1** in the representative structure. Active and decoy compounds were docked into the receptor and evaluated based on grid and pharmacophore matching similarity (FMS) as primary and secondary scores, respectively. We used the FMS as an intra-compound pose constraint and the Grid score for inter-compound ranking. The step-by-step protocol is as follows:

1. Pose generation: For each ligand, multiple docking poses were generated.
2. Pharmacophore constraint (within-compound selection): Poses were evaluated by FMS; only poses satisfying the pharmacophore match (or the highest-FMS pose when multiple matched) were retained for that ligand.
3. Compound ranking (across compounds): Retained poses were scored by the Grid score, and compounds were ranked by their best (lowest) Grid score.
4. Penalty for no pharmacophore match: If a ligand yielded no pharmacophore-matched pose, a penalty was applied by adding a +100 kcal/mol repulsive term to its Grid score, effectively placing such ligands at the bottom of the ranked list.

The position of the benzene ring in compound **1** was used as a pharmacophore to restrict the aromatic and ring structures.

We used the area under the receiver operating characteristic (AUROC) curve [55] to evaluate the virtual screening model. ROC was plotted and AUROC was calculated using the Python Scikit-learn (v0.24.2) library. The statistical analysis also was evaluated with 95% CI.

We performed a protein-ligand interaction fingerprint (PLIF) analysis using MOE to compare the binding interactions of Compounds **1–13**.

## Measurement of OR activity

A luciferase reporter gene assay was performed using the Dual-Glo Luciferase Assay System (Promega) following a previously described protocol [56–59]. Cultured cells were transfected with 12.5 ng of the tagged OR9Q2 vector, 2.5 ng CRE/luc2P, 1.25 ng TK/RLuc, 2.5 ng RTP1S-pcDNA3.1 vector, 1.25 ng Ric8B-pcDNA3.1 vector, and 1.25 ng Golf-pcDNA3.1, using 0.0425 µL Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) per well. At 24 h after transfection, the medium was replaced with CD293 (Invitrogen) containing 1/10,000 v/v GlutaMAX (Thermo Fisher Scientific, Waltham, MA, USA), and the cells were stimulated with one of the listed odorants for four hours at 37 °C and 5% CO<sub>2</sub> (Table 1). Luminescence was detected using FDSS7000EX functional drug screening system (Hamamatsu Photonics, Hamamatsu, Japan). Fold increases were calculated as Luc/hRLuc (Luc [luminescence of firefly luciferase] divided by hRLuc [luminescence of *Renilla* luciferase]) stimulated with an odorant, divided by Luc/hRLuc upon no stimulation. The expression of OR9Q2 was confirmed by immunohistochemistry using the anti-Rhodopsin antibody (Rho 4D2, ab 9887, abcam).

## Screening OR9Q2 agonists

In the primary screening, we stimulated OR9Q2 with 2,630 odorants, most of which were applied at a concentration of 300 µM. We standardized the log-transformed fold increase values and ranked the odorants accordingly. For the secondary screening, we selected the top 13 odorants that exhibited standardized values exceeding the 6-sigma threshold (i.e., values greater than the mean plus six standard deviations across all standardized data).

For the 13 odorants, we generated concentration-response curves using seven concentrations, ranging from 1 µM to 1000 µM, 0.3 µM to 300 µM, or 0.1 µM to 1000 µM, depending on the activity in the primary screening. Each odorant-OR dose condition was tested twice, using

biological replicates consisting of one duplicate and one quadruplicate set. Measurements were obtained from independent wells seeded with cells derived from common parent plates, ensuring biological variability. A vector-only control was included for each odorant condition. The relationship between concentration  $x$  and response  $E(x)$  was fitted to a sigmoidal curve using R (version 4.2.2, R Foundation for Statistical Computing, Vienna, Austria) and the *drc* package [60]:

$$E(x) = \frac{top}{1 + 10^{(\log(EC_{50}) - \log(x))}} \quad (2)$$

We adopted previously established criteria for identifying a positive concentration-dependent response [57–59]. An odorant classified as agonist if all of the following conditions were met: (i) the 95% confidence interval of the fitted top parameter did not include zero; (ii) the standard error of the fitted  $\log EC_{50}$  was less than one log unit; and (iii) the extra sums-of-squares test confirmed that the odorant activated the receptor significantly more than the vector-only control.

## Results

### Determination of $EC_{50}$ for each agonist

To identify OR9Q2 agonists, we conducted a high-throughput heterologous screening. In the primary screening, OR9Q2 was stimulated with 2630 odorants, resulting in a wide range of fold increase responses. After standardizing the log-transformed fold increase values, we identified 13 odorants that exceeded the 6-sigma threshold, indicating exceptionally strong activation of OR9Q2. In the secondary screening, concentration–response curves were successfully generated for all 13 selected odorants. Based on the activity observed in the primary screen, each odorant was tested across seven concentrations ranging from 0.1  $\mu\text{M}$  to 1000  $\mu\text{M}$ . Each odorant–OR dose condition was tested twice using biological replicates, consisting of one duplicate and one quadruplicate set. Measurements were obtained from independent wells seeded with cells derived from common parent plates. A vector-only control was included for each odorant condition to confirm specificity. Sigmoidal curve fitting using the *drc* package in R yielded reliable estimates of  $EC_{50}$  and top parameters for each odorant. Table 1 summarizes the  $EC_{50}$  and top values obtained from the quadruplicate experiments for all 13 agonists. Figure S3 shows the concentration–response curves for the 13 OR9Q2 agonists, illustrating the dose-dependent activation profiles and variability in potency. All 13 odorants met the predefined

criteria for agonist classification: (i) the 95% confidence interval of the fitted top parameter excluded zero, (ii) the standard error of the fitted  $\log EC_{50}$  was less than one log unit, and (iii) the extra sum-of-squares test indicated statistically significant activation compared to the vector-only control. The compounds with an  $EC_{50}$  less than 1  $\mu\text{M}$  were 4-Ethylphenol (2), 4-Chlorophenol (3), and p-Cresol (4). Notably, all 13 OR9Q2 agonists share a common structural feature—a phenol moiety—suggesting that OR9Q2 preferentially recognizes odorants with phenolic scaffolds. This structural convergence highlights a potential ligand recognition mechanism and provides a basis for further structure–activity relationship (SAR) analysis.

<sup>a</sup> Concentration–response curves are shown in Figure S3.

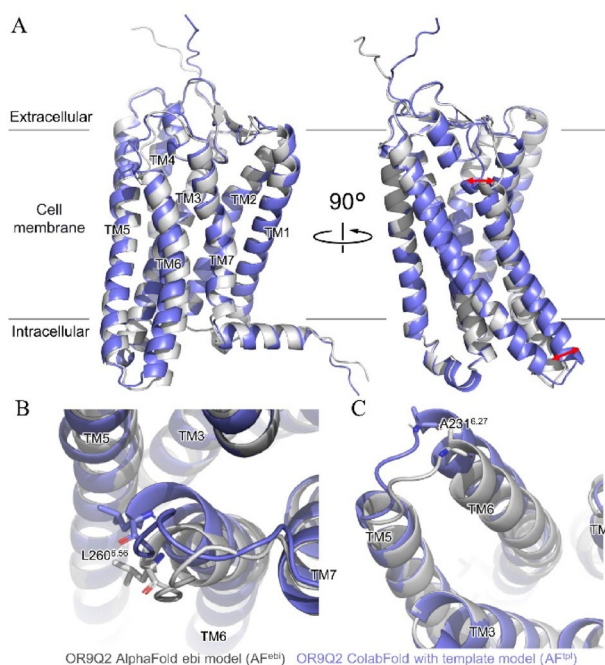
<sup>b</sup> The provided values are Mean  $\pm$  Standard Error (SE) from 2 biological replicas.

### Structural prediction for OR9Q2

The 3D structure of OR9Q2 was predicted using ColabFold with a template (AF2<sup>tpl</sup> model). Our predicted structure was compared with the structure predicted without a template from the AlphaFold2 ebi model (AF2<sup>ebi</sup> model). Figure 6A shows the comparison between the structures derived from the AF2 and AF2<sup>tpl</sup> models, and pLDDT scores are shown in Figure S4. The pLDDT scores for all regions except for TM6 were comparable. TM6, affected by the activation and inactivation range of pLDDT in the AF2<sup>tpl</sup> model, was scored lower than that in the AF2<sup>ebi</sup> model, indicating that it was affected by conformational changes due to complex formation with G(olf). RMSD with  $\text{Ca}$  atoms between two structures was 0.416 Å. We focused on the TM6 region, involved in the activation of class A GPCRs [61, 62]. The  $\text{Ca}$  atom of L260<sup>6,56</sup>, located on the extracellular side of TM6, differed by 3.9 Å between the two structures, and the AF2<sup>tpl</sup> model was oriented towards the inside of the TM bundle (Fig. 3B). In contrast, the  $\text{Ca}$  atom of A231<sup>6,27</sup>, located on the intracellular side of TM6, differed by 4.6 Å between the two structures, and the AF2<sup>tpl</sup> model opened towards the plasma membrane (Fig. 3C). Each terminal residue in TM6 was used for the measurements. These structural differences fall within the typical range observed for transitions between the active and inactive states of class A GPCRs [63].

### Docking and MD simulations

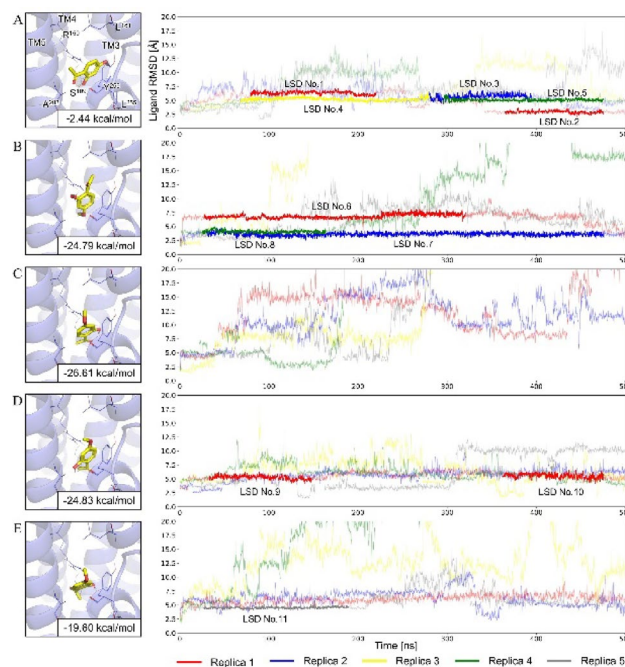
This docking and MD simulation analysis was conducted to identify potential binding modes of compound 1 within OR9Q2 and to evaluate the stability of the resulting complexes under dynamic conditions. Compound 1 was docked to the OR9Q2 (AF2<sup>tpl</sup>) model to generate initial complex structures. The generated poses were classified into five



**Fig. 3** Comparison of OR9Q2 structures between AlphaFold2 without template model and ColabFold with template model **A** Whole-model comparison between the AlphaFold2 ebi structure (gray) and ColabFold with the template structure (blue) from a parallel position of the cell membrane. **B** The viewpoint from outside the cell. The residue shown in stick representation is L260<sup>6,56</sup>, located on the extracellular side of TM6. The AF<sup>ebi</sup> structure (gray) opens outwards by approximately 3.9 Å from the TM bundle compared to the AF<sup>tpl</sup> structure (blue). **C** The viewpoint from inside the cell. The residue shown in stick representation is A231<sup>6,27</sup>, located on the intracellular side of TM6. The AF<sup>tpl</sup> structure (blue) opens outwards by approximately 4.6 Å from the TM bundle compared to the AF<sup>ebi</sup> structure (gray). All graphics of protein and ligands were generated using PyMOL (Ver. 2.5.2, Schrödinger & DeLano [2020])

clusters (Figure S5), and the pose with the lowest grid score in each cluster was selected as the initial complex structure. It should be noted that the binding free energy described here is derived from the docking score and should therefore be regarded as an estimated potential binding free energy, which is distinct from experimentally measured values obtained in binding assays. The five initial complex structures (A–E) and their grid scores are shown in Fig. 4A–E. A common interaction of cation- $\pi$  with R159<sup>4,60</sup> was found among all initial complex structures. The energetically lowest pose was the initial complex structure C, with a binding free energy of -26.61 kcal/mol. In contrast, the energetically worst pose was initial complex structure A, with a binding free energy of -2.44 kcal/mol.

We performed MD simulations for the initial complex structures. Figure 8 shows the RMSD of compound **1** heavy atoms during the MD simulations. In several trajectories with RMSD greater than 20 Å, compound **1** was completely dissociated from OR9Q2 (B: rep 4, C: rep 1–5, and E: rep



**Fig. 4** Root mean square deviation (RMSD) of the ligand in five replicas of molecular dynamics (MD) simulations for each initial cluster A–E. The initial complex structure A–E selected using docking simulation and pose clustering are shown in the left column. The ligand RMSDs of these initial complex structures are shown in the right plot. OR9Q2 is shown as a blue cartoon model and compound **1** as a yellow stick model. The MD replica is divided by color, and the bold line range is a section of ligand-stable duration (LSD)

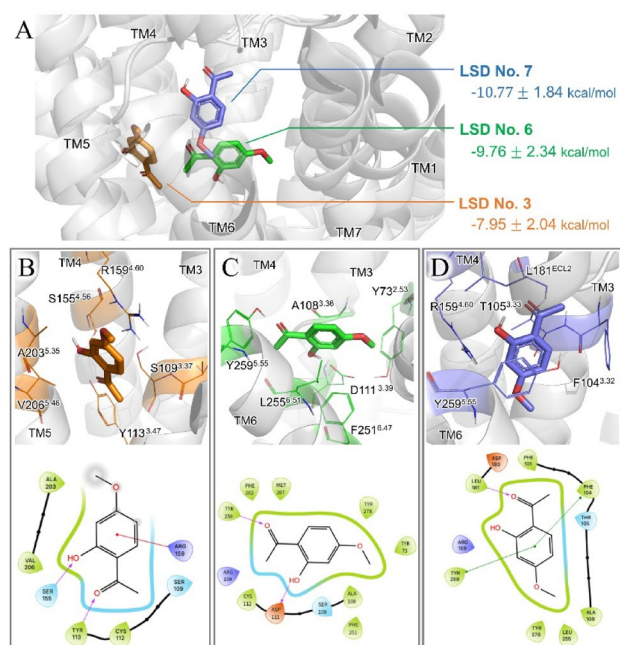
3,4). Neither the initial structure A nor D exhibited dissociation trajectories in any replicas.

Based on RMSD analysis, 11 LSDs were extracted from the MD trajectories. No LSDs were extracted from the trajectories using initial complex structure C. The longest LSD was LSD No. 7, with 445.1 ns from initial complex structure B, while the shortest was LSD No. 2, with 110.6 ns from initial complex structure A (Table S1).

### Comparison and evaluation of binding poses

Eleven LSDs were classified into three clusters based on the RMSD of compound **1** from the mean snapshot of each LSD (Figure S6). Representative structures for clusters 1–3 were selected from the mean snapshots of the longest LSD trajectories (LSD Nos. 6, 7, and 3), respectively. The binding free energies between compound **1** and OR9Q2 in the LSDs were calculated using the MM/PBSA method. The binding free energy of LSD No. 3, 6, and 7 were  $-7.95 \pm 2.04$ ,  $-9.76 \pm 2.34$ , and  $-10.77 \pm 1.84$  kcal/mol, respectively.

LSD 6 and 7 were located between TM3–TM6 (Fig. 5A). In particular, LSD No. 6 was embedded inside the TM bundle close to the intracellular side. This site is an important



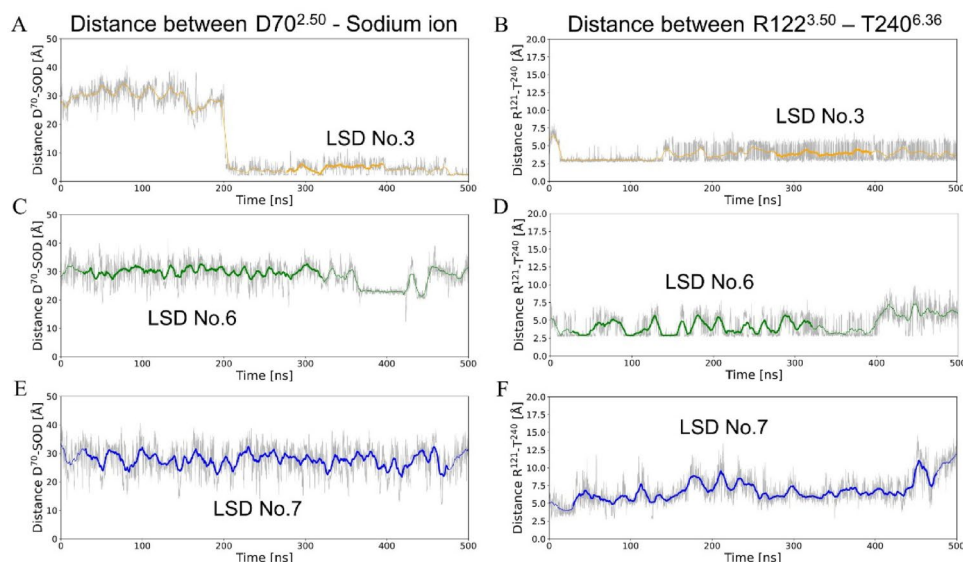
**Fig. 5** Representative structures for each cluster A. Representative structures for each cluster are shown (orange: ligand-stable duration [LSD] No. 3, green: LSD No. 6, blue: LSD No. 7). B–D. Binding pose (upper) and protein-ligand interaction diagram (lower) of LSD Nos. 3, 6, and 7 are shown, respectively

binding pocket in the OR [21]. In contrast, LSD No. 3 was located in a pocket around TM3–TM4–TM5.

We analyzed the differences in the interaction modes of compound 1 with OR9Q2 among the three representative binding poses obtained from MD simulation. In LSD No. 3, the hydroxyl and acetyl groups in compound 1 formed hydrogen bonds with S155<sup>4.56</sup> and Y113<sup>3.47</sup>, respectively. A cation- $\pi$  interaction was observed between R159<sup>4.60</sup> and benzene in compound 1 (Fig. 5B). In LSD No. 6, the hydroxyl group in compound 1 formed a hydrogen bond with D111<sup>3.39</sup> (Fig. 5C). Furthermore, the ketones formed a hydrogen bond Y259<sup>5.55</sup>, and the benzene in compound 1 formed hydrophobic interactions with A108<sup>3.36</sup>, F104<sup>3.32</sup>, and F251<sup>6.47</sup>. In LSD No.7, compound 1 formed hydrophobic interactions with F101<sup>3.29</sup>, F104<sup>3.32</sup>, A108<sup>3.36</sup>, L255<sup>6.51</sup>, and Y259<sup>5.55</sup> (Fig. 5D).  $\pi$ - $\pi$  stack interactions with benzene rings were formed at F104<sup>3.32</sup> and Y259<sup>5.55</sup>.

We focused on the binding of sodium ions to TM2 and the formation of an ionic lock to analyze the conformational change in OR9Q2 induced by the binding of compound 1.

Previous studies have suggested that sodium ions bind to Asp/Glu<sup>2.50</sup> and Asp/Glu<sup>3.39</sup> in the apo- and antagonist-bound states in class A GPCR [64–66]. First, the distance between D70<sup>2.50</sup>, which is suggested to bind sodium ions, and the nearest sodium ion was measured during MD simulation (Fig. 6A, C and E, and S7). In LSD No.3, all sections formed a salt bridge between D70<sup>2.50</sup> and sodium ions (Fig. 6E). In LSD No.6, the binding of sodium ion to D70<sup>2.50</sup> was



**Fig. 6** Changes in distances between the sodium ion pocket and ionic locks A, C, D. Distance between D70<sup>2.50</sup> and sodium ion, LSD No. 3, LSD No. 6, and LSD No. 7, respectively. The distance between D70<sup>2.50</sup> and the sodium ion closest to D70<sup>2.50</sup> on the trajectories was calculated. The OD1 or OD2 atoms in D70<sup>2.50</sup> were measured as the distance between D70<sup>2.50</sup> and the sodium ion, and the shorter one was selected. B, D, F. Distances between R122<sup>3.50</sup> and T240<sup>6.36</sup>, LSD No.

3, LSD No. 6, and LSD No. 7, respectively. The distance between TM3 and TM6 on the intracellular side was measured. The NH1 or NH2 atoms in R122<sup>3.50</sup> and OG1 atoms in T240<sup>6.36</sup> were used to measure the distance between TM3 and TM6, and the shorter one was selected. The color traces represent smoothed values with an averaging window of 8 ns; the grey traces represent unsmoothed values, and the bold line range is a section of LSD

not observed through the trajectories, as the distance was approximately 30 Å (Fig. 6A). LSD No.6 belongs to LSD cluster 1, and the hydroxyl group in compound **1** forms a hydrogen bond with the side chain of D111<sup>3,39</sup> (Fig. 5B). The interaction between D111<sup>3,39</sup>, which constitutes part of the sodium ion pocket, and compound **1** may prevent sodium ion binding. In LSD No.7, the sodium ion and D70<sup>2,50</sup> maintained a distance of approximately 30 Å with no binding through the trajectories (Fig. 6C). These results suggest that clusters 1 and 2 represent potential binding conformations for agonists.

Second, we analyzed the conformational changes in the ionic lock. An ionic lock refers to a stabilizing interaction between the intracellular sides of TM3 and TM6, involving a network of hydrogen bonds and salt bridges, such as the interaction between the conserved D(E)RY motif on TM3 and residues on TM6 [67]. The polar interactions formed between TM3 and TM6 on the intracellular side dissociate following GPCR activation. Herein, we calculated the distance between R122<sup>3,50</sup> and T240<sup>6,36</sup>. R122<sup>3,50</sup> is part of the DRY motif on TM3, and T240<sup>6,36</sup> is a highly conserved residue in the TM6 of OR. Figure 6B, D and F, and S8 show the distances between these residues in the trajectories containing LSD No. 3, 6, and 7. In the trajectory of LSD No. 6, the intracellular ends of TM3 and TM6 remained within 6 Å until 400 ns, indicating that these helices stayed in close proximity. After 400 ns, the distance exceeded 6 Å, indicating that the intracellular side of TM3-TM6 was beginning to open (Fig. 6D). In the trajectory of LSD No. 7, the distance was about 6 Å until 450 ns, indicating no interaction (Fig. 6E). After 450 ns, the intracellular side of TM3-TM6 opened as did the trajectory of LSD No. 6. In the trajectory of LSD No. 3, R122<sup>3,50</sup> - T240<sup>6,36</sup> distance fluctuated in the range 3–6 Å over the last 350 ns (Fig. 6F). No behavior could be considered a conformational change at that distance, and

the intracellular side of TM3–TM6 was inferred to have remained closed in the simulation.

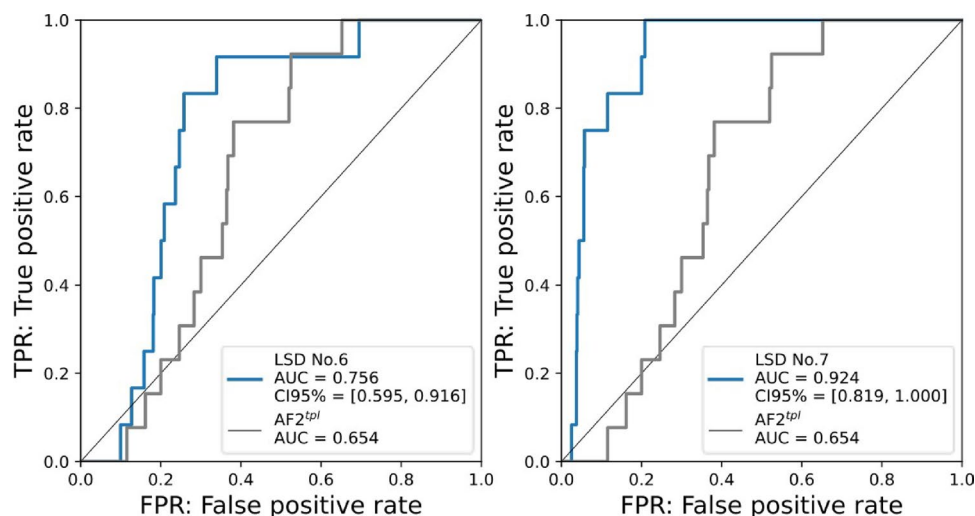
## Virtual screening

A virtual screening was conducted to evaluate the performance of the proposed docking-based virtual screening model. The grid score and FMS were adapted for re-ranking and pose selection, respectively. The discrimination between active and decoy compounds was estimated using AUROC.

For LSD No. 6, AUROC were 0.753, and for LSD No. 7, these were 0.924 (Fig. 7). The AUROC measured for the AF<sup>2pl</sup> model was 0.654. Both LSD-based models performed better than the random model and the AF<sup>2pl</sup> model in terms of discrimination indices.

We performed a PLIF analysis for the docking poses with these two models and compounds **1–13** (Figure S9). Compounds **1–13** were separated into high and middle activation groups. The high activation group contained the compounds whose  $-\log EC_{50}$  values were higher than 6. The middle activation group contained other compounds. In LSD No. 6, most compounds contacted with A108<sup>3,36</sup> and L255<sup>6,51</sup>. In particular, the high activation group formed an h-bond with D111<sup>3,39</sup> and contacted Y278<sup>7,41</sup>. The compounds in the middle activation group interacted either with D111<sup>3,39</sup> or Y278<sup>7,41</sup>. Therefore, the conditions that activated OR9Q2 included interaction with A108<sup>3,36</sup> and L255<sup>6,51</sup>. Furthermore, interaction with D111<sup>3,39</sup> and Y278<sup>7,41</sup> may have contributed to enhanced activation. The major interactions were with TM3 and TM6. In LSD No. 7, most compounds contacted with T105<sup>3,33</sup> and Y259<sup>5,55</sup>. In two of the compounds in the high activation group, arene interactions were observed with F104<sup>3,32</sup> and T105<sup>3,33</sup>. Most compounds of the middle activation group formed both or either of the T105<sup>3,33</sup> interactions with R159<sup>4,60</sup>. Compared to LSD No. 7, LSD No. 6 showed more specific interactions common

**Fig. 7** Virtual screening results A, B. The receiver operating characteristic (ROC) curve for virtual screening with representative complexes of ligand-stable duration (LSD) No.6 and 7 (blue line). The corresponding AUROC values were calculated for each model. For comparison, the ROC curve and AUROC obtained using the AF<sup>2pl</sup> model are shown in gray. The CI 95% for AF<sup>2pl</sup> was 0.492–0.820. In each panel, the diagonal line indicates the expected random trend



**Table 2** Comparison of profile binding poses

| No. of LSD                                 | LSD No.3        | LSD No.6    | LSD No.7     |
|--------------------------------------------|-----------------|-------------|--------------|
| Cluster                                    | 3               | 1           | 2            |
| MMPBSA±SD<br>[kcal/mol]                    | -7.95±2.04      | -9.76±2.34  | -10.77±1.84  |
| D70 <sup>2.50</sup> -Sodium ion            | Binding         | No binding  | No binding   |
| R122 <sup>3.50</sup> -T240 <sup>6.36</sup> | Not dissociated | Dissociated | Dissociated  |
| A <sup>100</sup> tool                      | -7.66±8.40      | 31.71±14.38 | -26.45±15.22 |

to the high activation group, and the correlation between interactions and activity values was clearer.

## Comprehensive evaluation of ligand binding poses

We suggest some binding positions for compound **1** in OR9Q2 using structural prediction, docking simulation, MD simulation, and extraction of the stable section. Table 2 summarizes the structural comparison among the agonist-binding conformation of the three representative LSDs. In LSD No.3, indications of activation were not confirmed, as D70<sup>2.50</sup> binds sodium ions in the sodium ion pocket while the ionic lock maintains the polar interaction. The binding free energy was the highest among the three LSDs. Furthermore, the A<sup>100</sup> tool, a classification method for the activation state of class A GPCRs by Ibrahim et al. [68], was used for structural evaluation. The A<sup>100</sup> activation model was trained using simulations and validated against X-ray structural data, enabling quantitative structure-based scoring of activation states in class A GPCRs. A higher value for the A<sup>100</sup> tool indicated that the state was similar to that in the active conformation. The mean value for LSD No.3 was  $-7.66 \pm 8.40$ , indicating a low probability of these structures being active states. In LSD No.6, activation behavior was observed in two structural analyses (Fig. 6). The binding free energy of LSD No. 6 was lower than those of LSD No.

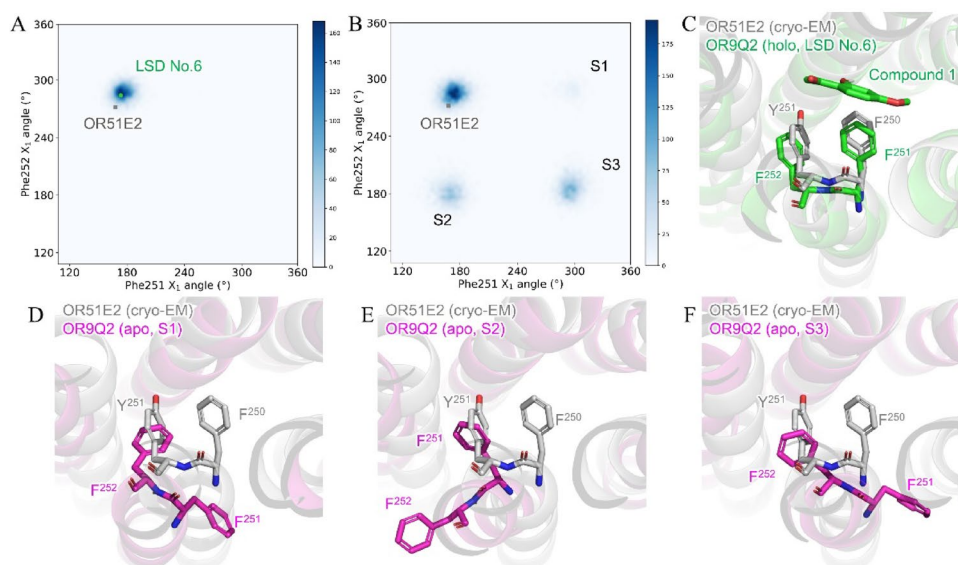
3 and LSD No.6, showing the largest stable section (Table S1). The mean value obtained from the A<sup>100</sup> tool was  $31.71 \pm 14.38$ , indicating that these structures were in an intermediate state, closer to an active state than LSD No. 3. The activation behavior was observed for LSD No.7, indicating it was stable and had a low binding free energy. However, the mean value of the A<sup>100</sup> tool was  $-26.45 \pm 15.22$ , indicating that these structures were far from being in an active state. Based on these results, LSD No.6 is proposed as a promising binding pose for compound **1** to OR9Q2 for activation.

## Discussion

### Comparison with a known 3D structure of an OR

OR51E2 is the only OR for which the 3D structure has been determined [17]. The sequence similarity between OR9Q2 and OR51E2 28.5%. The RMSD between the AF<sup>tp1</sup> model and OR51E2 was 1.447 Å for the full-length alignment, and 1.021 Å when restricted to the ECL2 region, which is implicated in ligand binding [69]. Therefore, we believe that the structure of the AF<sup>tp1</sup> model is reasonable. Billesbølle et al. suggested that the orientation of the side chain in Phe<sup>6.47</sup> and Tyr<sup>6.48</sup> relate to the OR51E2 activation mechanism. These aromatic rings are conserved in sequences F251<sup>6.47</sup> and F252<sup>6.48</sup> in OR9Q2 and others [40]. To analyze the orientation of these aromatic rings in our agonist-binding model, we observed the dihedral angle of these aromatic rings in the MD trajectories of OR9Q2 in the holo and apo states. The trajectory containing LSD No. 6 was the holo state, and the trajectory to which compound **1** did not bind was the apo state. Figure 8 shows the dihedral angle distribution of  $\chi_1$  in

**Fig. 8** Comparison of the orientation of residues in relation to the activation mechanism **A, B**. Distribution of rotation of sidechain rotamer angle of F251<sup>6.47</sup> and F252<sup>6.48</sup> in the molecular dynamics (MD) simulation of holo and apo states, respectively. **C** Close-up view of the aromatic residues in active OR51E2 and representative complex of OR9Q2 and compound **1**. **D–F** Close-up view of the aromatic residues in active OR51E2 and representative apo OR9Q2 of S1, S2 and S3, respectively



these aromatic rings in the OR51E2 experimental structure and MD trajectories in OR9Q2. For the MD trajectories in the holo state, these aromatic rings were stabilized at specific dihedral angles similar to the dihedral angles of OR51E2 (Fig. 8A). In contrast, the dihedral angles in the MD trajectories with the apo state transitioned in various orientations (Fig. 8B). S1 was a state in which F251<sup>6,47</sup> faced the plasma membrane from TM6–7 while F252<sup>6,48</sup> faced the plasma membrane from TM5–6 in state S2 (Fig. 8D–E). In S3, F251<sup>6,47</sup> and F252<sup>6,48</sup> opened toward the plasma membrane (Fig. 8F). Therefore, F251<sup>6,47</sup> and F252<sup>6,48</sup> may maintain a constant dihedral angle owing to its binding to compound 1. This suggests similar activation indicators for other ORs, including our model of OR9Q2.

### Evaluation of the OR model predicted using AF2

The number of known 3D structures of human ORs is limited, and 3D predicted structures are used in structure-activity relationship studies. In recent years, deep learning-based 3D structure prediction methods, such as AF2 and RoseTTAFold [70], have been developed, facilitating highly accurate protein 3D structural predictions [21, 71]. In general, the unbiased AF2 predicts the structure in an apo form. Recent studies have suggested that AF2 structures are often unsuitable for docking simulations [26–28]. Furthermore, membrane proteins, such as transporters, ion channels and GPCRs, exist in several conformations, and AI-based protein prediction can only generate one of these states. One proposed solution to this problem attempts to predict a specific active-state 3D structure using selected multiple sequence alignment data [30]. Other methods accurately predict alternative conformations by reducing the depth of the MSA and performing MD simulation-based structural optimization, focusing on sites related to the function of the OR for evaluation [72, 73]. In this study, we constructed an agonist-binding state, OR9Q2 (AF2<sup>tpi</sup>) with G(olf), predicted using an agonist template and a G protein-bound GPCR. Indeed, the AF2<sup>tpi</sup> model was more open towards the lipid membrane on the intracellular side of TM6 cells than the AF2<sup>ebi</sup> model (Fig. 6). Based on the conformational changes in TM6 upon activation of the general class A GPCR [61, 74], the AF2<sup>tpi</sup> model was closer to the active state than the AF2<sup>ebi</sup> model. Constructing an initial structure closer to a specific state for docking and MD simulations is preferable. To evaluate the advantages of this workflow, virtual screening was performed using the AF2<sup>ebi</sup> and AF2<sup>tpi</sup> models, both of which were generated without applying the workflow. The resulting AUROC values were 0.613 and 0.654, respectively (Figure S10). The AF2<sup>tpi</sup> model predicted higher performance of the agonist-binding state than the AF2<sup>ebi</sup> model. In OR structure sampling, focusing on the

initial structure of the receptor and use specific states, rather than general drug discovery, may be more effective. Furthermore, the AF2<sup>tpi</sup> models with MD simulations showed improved performance in the virtual screening (Fig. 7) compared to those of the AF2 structures. Therefore, our protocol is useful as a sampling method for virtual screening models.

### Performance and potential of LSD

GPCRs have a common topology, comprising even trans-membrane helices; however, the sequence identities among these GPCRs are low, and each GPCR recognizes multiple ligands [75]. In general, there are many OR genes in class A GPCRs, and a particular OR type may recognize more than one odor molecule or each odor molecule may be recognized by several ORs [18–21]. Furthermore, odorous molecules are lipophilic, volatile, and relatively small [76]. This background suggests that OR-odor affinity is low and does not follow the classical lock-and-key model that governs many receptor-ligand interactions [77, 78]. Thus, we performed complex sampling to search for binding poses using MD simulations with pre-complex structures from docking simulations. Although docking simulations, clustering analyses, and MD simulations are widely used standard approaches for exploring ligand binding modes, the novelty of our protocol lies in the use of the LSD method to reliably extract representative binding poses from MD trajectories. The LSD analysis enabled the extraction of trajectories with stable poses, even if the target had low ligand-binding or fragment stability. The LSD method can subsequently be applied to these trajectories to extract pockets and poses where active compounds are potentially stable and can be developed for ligand recognition. Araki et al. extracted conformations maintained stably from complex MD simulations, and free energy simulations were performed using representative conformations [79]. These results suggest that LSD as a suitable analysis method for ORs with low protein-ligand binding affinity and few homologous known 3D structures. The present workflow applies to specific ORs and fragment-based drug design in drug discovery research. However, before this method can be applied in practice, it must be validated to ensure its suitability for specific targets and ligand chemotypes.

### Practical considerations in using LSD

The first consideration concerns the parameters used in LSD, namely the window time of VF (25 ns), the threshold ( $VF(t) < 0.2 \text{ \AA}^2$ ), the duration (100 ns) and the total production MD time (12.5  $\mu$ s). Regarding the window time, a provisional value of 50 ns (501 frames) was used in this study. As the criteria depend on the factors such as molecular size

and ligand potency, appropriate reference value must be validated for different ligands. In the validation of docking simulations, re-docking performance is commonly evaluated based on the RMSD of ligand poses. In general, a docking result is considered successful when the RMSD between the predicted and reference ligand poses is within 1.5–2.0 Å, depending on the ligand size [80, 81]. In our study, a VF threshold of 0.2 Å<sup>2</sup> corresponds to a standard deviation of approximately 0.45 Å, representing a deviation of  $\pm 0.5$  Å from the mean RMSD. In other words, this threshold allows for a positional change of roughly 1.0 Å, which we consider sufficiently strict, even for small-molecule ligands. However, this threshold should be adjusted according to specific conditions, such as ligand size and the affinity of the protein–ligand complex. The duration time was set to 100 ns based on prior research employing short molecular dynamics simulations for refinement and rescaling in docking analyses, which supports this duration as a practical and meaningful choice [82].

As a final consideration regarding parameters, we address the total time of production MD time. It should be emphasized that elucidating the binding and activation mechanisms of odorant molecules by MD would ultimately require ultra-long simulations on the millisecond timescale or beyond, rather than a modest extension of trajectories to 1 microsecond. Such extended timescales are necessary to capture rare events and slow conformational transitions, but they represent a highly resource-intensive challenge that exceeds the scope of the present work. In this study, we aimed to demonstrate the utility of the LSD-based protocol within a computationally feasible framework, while recognizing that future advances in computational power will be essential for exploring the odorant–receptor system at such extended timescales.

It is also important to recognize that odorant molecules typically exhibit weaker binding affinities than conventional drug-like ligands, and thus the classical MD simulation schemes developed for stable protein–ligand complexes may not necessarily be directly applicable. A more definitive understanding of odorant–receptor binding mechanisms will require the accumulation of experimentally determined structures of odorant–receptor complexes, obtained for example by cryo-electron microscopy or X-ray crystallography. Such structural information will be crucial for benchmarking computational models and for guiding the refinement of simulation-based approaches to odorant recognition.

It should be noted that, in this study, the validation of the LSD protocol through virtual screening test was performed using the same agonist series that had been employed for the derivation of LSDs. This design choice was intentional, as binding poses are known to vary substantially across

different chemotypes. Our primary objective was therefore to assess the explanatory capacity of the LSD protocol within a single chemotype framework. We reasoned that if the protocol failed to demonstrate validity even within the same dataset, its extension to other chemotypes would be unlikely to succeed. Accordingly, the present results should be interpreted as an evaluation of binding pose validity confined to one chemotype. Future work will focus on extending the application of this protocol to independent datasets of confirmed active compounds with distinct chemotypes, thereby enabling broader generalization of the approach.

## Conclusion

In conclusion, we constructed ligand-binding complexes of OR9Q2 using our molecular simulation-based protocol, which included four steps: structure prediction, docking simulation, molecular dynamics simulation with LSD analysis, and structural verification. Consequently, promising complex structures that could discriminate between active and decoy compounds were obtained. These promising complex structures exhibited a behavior of activation, as suggested in previous studies, and maintained the specific orientation of key residues, as well as those in the active state OR51E2 determined using cryo-EM. In the case of models predicted using ColabFold with the template of an active GPCR, using an initial structure close to a specific state for docking and MD simulations is better. We suggest this protocol as a suitable analysis method for ORs and LSD analysis applies to fragment-based drug design in drug discovery research.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10822-025-00706-0>.

**Acknowledgements** Computational experiments were performed using the Cygnus and Pegasus computational resources provided by the Multidisciplinary Cooperative Research Program at the Center for Computational Sciences (Project Code: CADD) at the University of Tsukuba.

**Author contributions** T. Hirao, Y.I., C.I., G.K., R.Y., and T. Hirokawa conceived the study. Y.I. and C.I. performed the Luciferase reporter gene assay. T. Hirao performed molecular simulations and analyzed the data. The manuscript was written with contributions from all authors. All the authors approved the final version of the manuscript.

**Funding** This research was supported by Ajinomoto Co., Inc., the Research Support Project for Life Science and Drug Discovery [Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)] (grant number JP23ama121029j0002) of the Japan Agency for Medical Research and Development (AMED), KAKENHI (grant numbers 23K16987) from the Japan Society for the Promotion of Science (JSPS), and Program for Promoting Researches on the Supercomputer Fugaku (JPMXP1020230120).

**Data availability** The 3D structure of OR9Q2 was predicted using ColabFold (ver. 1.3.0), and downloaded from the AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk/>). H++ server (Ver. 4.0) was used to determine the protonation state of OR9Q2. MOE (Chemical Computing Group, LLC) was used for 3D structure construction, protonation, and minimization of compound **1**. We used AmberTools21 and Gaussian 16 Revision C.01 to prepare compound **1**. DOCK6.9 was used for docking simulation. Membrane Builder in CHARMM-GUI was used as the system setting for the MD simulation. GROMACS 2021.5 was used as the MD engine. We used a decoy from DUD-E to prepare the decoy compounds. PyMOL (Ver. 2.5.2, Schrödinger, & DeLano [2020]) was used for visualization. The structure of representative LSD No. 3, the structure of representative LSD No. 6, the structure of representative LSD No. 7, the initial structure of OR9Q2, SMILES of decoys used for docking, the Python script of LSD analysis, the parameter file of colabfold, the output file of CHARMM-GUI, the parameter file of virtual screening test and the parameter file of production MD are deposited in Zenodo (<https://doi.org/10.5281/zenodo.13328860>). The Supporting Information is available free of charge.

## Declarations

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

## References

- Hilger D, Masureel M, Kobilka BK (2018) Structure and dynamics of GPCR signaling complexes. *Nat Struct Mol Biol* 25:4–12. <https://doi.org/10.1038/s41594-017-0011-7>
- Wacker D, Stevens RC, Roth BL (2017) How ligands illuminate GPCR molecular Pharmacology. *Cell* 170:414–427
- Congreve M, de Graaf C, Swain NA, Tate CG (2020) Impact of GPCR structures on drug discovery. *Cell* 181:81–91
- Davies MN, Secker A, Halling-Brown M et al (2008) GPCRTree: online hierarchical classification of GPCR function. *BMC Res Notes*. <https://doi.org/10.1186/1756-0500-1-67>. 1:
- Katritch V, Cherezov V, Stevens RC Structure-Function of the G-protein-Coupled Receptor Superfamily. <https://doi.org/10.1146/annurev>
- Kooistra AJ, Mordalski S, Pándy-Szekeres G et al (2021) GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res* 49:D335–D343. <https://doi.org/10.1093/nar/gkaa1080>
- Alhosaini K, Azhar A, Alonazi A, Al-Zoghaibi F (2021) GPCRs: the most promiscuous druggable receptor of the mankind. *Saudi Pharm J* 29:539–551
- Tautermann CS (2014) GPCR structures in drug design, emerging opportunities with new structures. *Bioorg Med Chem Lett* 24:4073–4079
- Kobilka B (2013) The structural basis of G-protein-coupled receptor signaling (nobel lecture). *Angewandte Chemie - Int Ed* 52:6380–6388. <https://doi.org/10.1002/anie.201302116>
- Deupi X (2014) Relevance of rhodopsin studies for GPCR activation. *Biochim Biophys Acta Bioenerg* 1837:674–682
- Kolb P, Kenakin T, Alexander SPH et al (2022) Community guidelines for GPCR ligand bias: IUPHAR review 32. *Br J Pharmacol* 179:3651–3674
- Hillenbrand M, Schori C, Schöppe J, Plückthun A (2015) Comprehensive analysis of heterotrimeric G-protein complex diversity and their interactions with GPCRs in solution. *Proc Natl Acad Sci U S A* 112:E1181–E1190. <https://doi.org/10.1073/pnas.1417573112>
- Santos R, Ursu O, Gaulton A et al (2016) A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* 16:19–34. <https://doi.org/10.1038/nrd.2016.230>
- Hauser AS, Attwood MM, Rask-Andersen M et al (2017) Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov* 16:829–842. <https://doi.org/10.1038/nrd.2017.178>
- Malnic B, Godfrey PA, Buck LB (2004) Howard § ‡ The human olfactory receptor gene family
- Niimura Y Evolutionary dynamics of olfactory receptor genes in chordates: Interaction between environments and genomic contents
- Billesbølle CB, de March CA, van der Velden WJC et al (2023) Structural basis of odorant recognition by a human odorant receptor. *Nature*. <https://doi.org/10.1038/s41586-023-05798-y>
- Geithe C, Noe F, Kreissl J, Krautwurst D (2017) 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
- Kajiya K, Inaki K, Tanaka M et al (2001) Molecular bases of odor discrimination. Reconstitution of Olfactory Receptors that Recognize Overlapping Sets of Odorants
- Nara K, Saraiva LR, Ye X, Buck LB (2011) A large-scale analysis of odor coding in the olfactory epithelium. *J Neurosci* 31:9179–9191. <https://doi.org/10.1523/JNEUROSCI.1282-11.2011>
- Nicoli A, Haag F, Marcinek P et al (2022) Modeling the orthosteric binding site of the G Protein-Coupled odorant receptor OR5K1. *J Chem Inf Model*. <https://doi.org/10.1021/acs.jcim.2c00752>
- Jumper J, Evans R, Pritzel A et al (2021) Highly accurate protein structure prediction with alphafold. *Nature* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Borkakoti N, Thornton JM (2023) AlphaFold2 protein structure prediction: implications for drug discovery. *Curr Opin Struct Biol* 78
- Varadi M, Anyango S, Deshpande M et al (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 50:D439–D444. <https://doi.org/10.1093/nar/gkab1061>
- Bryant P, Pozzati G, Elofsson A (2022) Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* 13. <https://doi.org/10.1038/s41467-022-28865-w>
- Scardino V, Di Filippo JI, Cavasotto CN (2023) How good are alphafold models for docking-based virtual screening? *iScience* 26:. <https://doi.org/10.1016/j.isci.2022.105920>
- Díaz-Rovira AM, Martín H, Beuming T et al (2022) Are deep learning structural models sufficiently accurate for virtual screening? Application of Docking algorithms to AlphaFold2 predicted

- structures. *J Chem Inf Model*. <https://doi.org/10.1021/acs.jcim.2c01270>
28. Kersten C, Clower S, Barthels F (2022) Hic Sunt dracones: molecular Docking in uncharted territories with structures from AlphaFold2 and rosettafold. *J Chem Inf Model*. <https://doi.org/10.1021/acs.jcim.2c01400>
29. Abramson J, Adler J, Dunger J et al (2024) Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature* 630:493–500. <https://doi.org/10.1038/s41586-024-07487-w>
30. Heo L, Feig M (2022) Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins: Struct Function Bioinf* 90:1873–1885. <https://doi.org/10.1002/prot.26382>
31. Mirdita M, Schütze K, Moriwaki Y et al (2022) ColabFold: making protein folding accessible to all. *Nat Methods* 19:679–682. <https://doi.org/10.1038/s41592-022-01488-1>
32. Bateman A, Martin MJ, Orchard S et al (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49:D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
33. Liu C, Zheng C, Shen X et al (2023) Serum CRP interacts with SPARC and regulate immune response in severe cases of COVID-19 infection. *Front Immunol* 14. <https://doi.org/10.3389/fimmu.2023.1259381>
34. Zhang H, Chen LN, Yang D et al (2021) Structural insights into ligand recognition and activation of the melanocortin-4 receptor. *Cell Res* 31:1163–1175. <https://doi.org/10.1038/s41422-021-00552-3>
35. Gordon JC, Myers JB, Folta T et al (2005) H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* 33. <https://doi.org/10.1093/nar/gki464>
36. Anandakrishnan R, Aguilar B, Onufriev AV (2012) H++3.0: automating pK prediction and the Preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* 40. <https://doi.org/10.1093/nar/gks375>
37. Myers J, Grothaus G, Narayanan S, Onufriev A (2006) A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins: Struct Function Genet* 63:928–938. <https://doi.org/10.1002/prot.20922>
38. Molecular Operating Environment (MOE) (2022) 02 Chemical Computing Group ULC, 910–1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada, 2024
39. Allen WJ, Balius TE, Mukherjee S et al (2015) DOCK 6: impact of new features and current Docking performance. *J Comput Chem* 36:1132–1156. <https://doi.org/10.1002/jcc.23905>
40. De March CA, Yu Y, Ni MJ et al (2015) Conserved residues control activation of mammalian G Protein-Coupled odorant receptors. *J Am Chem Soc* 137:8611–8616. <https://doi.org/10.1021/jacs.5b04659>
41. De March CA, Kim SK, Antonczak S et al (2015) G protein-coupled odorant receptors: from sequence to structure. *Protein Sci* 24:1543–1548. <https://doi.org/10.1002/pro.2717>
42. Kozakov D, Grove LE, Hall DR et al (2015) The ftmap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* 10:733–755. <https://doi.org/10.1038/nprot.2015.043>
43. Halgren TA (2009) Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* 49:377–389. <https://doi.org/10.1021/ci800324m>
44. Halgren T (2007) New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des* 69:146–148. <https://doi.org/10.1111/j.1747-0285.2007.00483.x>
45. Maier JA, Martinez C, Kasavajhala K et al (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11:3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>
46. Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* 23:1623–1641. <https://doi.org/10.1002/jcc.10128>
47. Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. <https://doi.org/10.1002/jcc.20084>
48. Van Der Spoel D, Lindahl E, Hess B et al (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26:1701–1718
49. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem* 29:1859–1865. <https://doi.org/10.1002/jcc.20945>
50. Huang J, Rauscher S, Nawrocki G et al (2016) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 14:71–73. <https://doi.org/10.1038/nmeth.4067>
51. Jo S, Kim T, Im W (2007) Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0000880>. 2:
52. Miller BR, McGee TD, Swails JM et al (2012) MMPBSA.py: an efficient program for end-state free energy calculations. *J Chem Theory Comput* 8:3314–3321. <https://doi.org/10.1021/ct300418h>
53. Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E (2021) Gmx\_MMPBSA: A new tool to perform End-State free energy calculations with GROMACS. *J Chem Theory Comput* 17:6281–6291. <https://doi.org/10.1021/acs.jctc.1c00645>
54. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* 55:6582–6594. <https://doi.org/10.1021/jm300687e>
55. Triballeau N, Acher F, Brabet I et al (2005) Virtual screening workflow development guided by the receiver operating characteristic curve approach. Application to high-throughput Docking on metabotropic glutamate receptor subtype 4. *J Med Chem* 48:2534–2547. <https://doi.org/10.1021/jm049092j>
56. Ijichi C, Wakabayashi H, Sugiyama S et al (2019) Metabolism of odorant molecules in human nasal/oral cavity affects the odorant perception. *Chem Senses* 44:465–481. <https://doi.org/10.1093/chemse/bjz041>
57. Mainland JD, Li YR, Zhou T et al (2015) Human olfactory receptor responses to odorants. *Sci Data* 2:. <https://doi.org/10.1038/SDATA.2015.2>
58. Saito H, Chi Q, Zhuang H et al Odor coding by a mammalian receptor repertoire. <https://doi.org/10.1126/scisignal>
59. Mainland JD, Keller A, Li YR et al (2014) The missense of smell: functional variability in the human odorant receptor repertoire. *Nat Neurosci* 17:114–120. <https://doi.org/10.1038/nn.3598>
60. Ritz C, Baty F, Streibig JC, Gerhard D (2015) Dose-response analysis using R. *PLoS ONE* 10. <https://doi.org/10.1371/journal.pone.0146021>
61. Deupi X, Standfuss J (2011) Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr Opin Struct Biol* 21:541–551
62. Kobilka BK (2007) G protein coupled receptor structure and activation. *Biochim Biophys Acta Biomembr* 1768:794–807
63. Dalton JAR, Lans I, Giraldo J (2015) Quantifying conformational changes in gpcrs: glimpse of a common functional mechanism. *BMC Bioinformatics* 16. <https://doi.org/10.1186/s12859-015-0567-3>
64. Katritch V, Fenalti G, Abola EE et al (2014) Allosteric sodium in class A GPCR signaling. *Trends Biochem Sci* 39:233–244
65. Agasid MT, Sørensen L, Urner LH et al (2021) The effects of sodium ions on ligand binding and conformational States of G Protein-Coupled Receptors-Insights from mass spectrometry. *J Am Chem Soc* 143:4085–4089. <https://doi.org/10.1021/jacs.0c11837>

66. Pirona L, Ballabio F, Alfonso-Prieto M, Capelli R (2024) Calcium-Driven in Silico inactivation of a human olfactory receptor. *J Chem Inf Model* 64:2971–2978. <https://doi.org/10.1021/acs.jcim.4c00249>
67. Mafi A, Kim S-K, Iii WAG et al (2022) The mechanism for ligand activation of the GPCR-G protein complex. <https://doi.org/10.1073/pnas>
68. Ibrahim P, Wifling D, Clark T (2019) Universal activation index for class A GPCRs. *J Chem Inf Model* 59:3938–3945. <https://doi.org/10.1021/acs.jcim.9b00604>
69. Yu Y, Ma Z, Pacalon J et al (2022) Extracellular loop 2 of G protein-coupled olfactory receptors is critical for odorant recognition. *J Biol Chem* 298. <https://doi.org/10.1016/j.jbc.2022.102331>
70. Baek M, Dimaio F, Anishchenko I et al (2021) Accurate prediction of protein structures and interactions using a three-track neural network
71. Sakamoto K, Asano S, Ago Y, Hirokawa T (2022) AlphaFold version 2.0 elucidates the binding mechanism between VIPR2 and KS-133, and reveals an S–S bond (Cys25–Cys192) formation of functional significance for VIPR2. *Biochem Biophys Res Commun* 636:10–16. <https://doi.org/10.1016/j.bbrc.2022.10.071>
72. Del Alamo D, Sala D, McHaourab HS, Meiler J (2022) TITLE: sampling alternative conformational States of transporters and receptors with AlphaFold2. *Elife* 11:. <https://doi.org/10.7554/eLife.75751>
73. Alfonso-Prieto M, Capelli R (2023) Machine Learning-Based modeling of olfactory receptors in their inactive state: human OR51E2 as a case study. *J Chem Inf Model* 63:2911–2917. <https://doi.org/10.1021/acs.jcim.3c00380>
74. Dror RO, Arlow DH, Maragakis P et al Activation mechanism of the  $\beta$  2-adrenergic receptor.
75. Katritch V, Cherezov V, Stevens RC (2012) Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol Sci* 33:17–27
76. Ruddigkeit L, Awale M, Reymond JL (2014) Expanding the fragrance chemical space for virtual screening. *J Cheminform* 6. <https://doi.org/10.1186/1758-2946-6-27>
77. del Marmol J, Yedlin MA, Ruta V (2021) The structural basis of odorant recognition in insect olfactory receptors. *Nature* 597:126–131. <https://doi.org/10.1038/s41586-021-03794-8>
78. Genva M, Kemene TK, Deleu M et al (2019) Is it possible to predict the odor of a molecule on the basis of its structure? *Int J Mol Sci* 20
79. Araki M, Kamiya N, Sato M et al (2016) The effect of conformational flexibility on binding free energy Estimation between kinases and their inhibitors. *J Chem Inf Model* 56:2445–2456. <https://doi.org/10.1021/acs.jcim.6b00398>
80. Hevener KE, Zhao W, Ball DM et al (2009) Validation of molecular Docking programs for virtual screening against dihydropterotate synthase. *J Chem Inf Model* 49:444–460. <https://doi.org/10.1021/ci800293n>
81. Chen H, Lyne PD, Giordanetto F et al (2006) On evaluating molecular-docking methods for pose prediction and enrichment factors. *Journal of chemical information and modeling. American Chemical Society*, pp 401–415
82. Guterres H, Im W (2020) Improving Protein-Ligand Docking results with High-Throughput molecular dynamics simulations. *J Chem Inf Model* 60:2189–2198. <https://doi.org/10.1021/acs.jcim.0c00057>

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