

ARTICLE

<https://doi.org/10.1038/s42003-019-0384-8>

OPEN

# Computational modeling of the olfactory receptor Olfr73 suggests a molecular basis for low potency of olfactory receptor-activating compounds

Shuguang Yuan<sup>1,2,3</sup>, Thamani Dahoun<sup>2</sup>, Marc Brugarolas<sup>2</sup>, Horst Pick<sup>2</sup>, Slawomir Filipek<sup>3</sup> & Horst Vogel<sup>2</sup>

The mammalian olfactory system uses hundreds of specialized G-protein-coupled olfactory receptors (ORs) to discriminate a nearly unlimited number of odorants. Cognate agonists of most ORs have not yet been identified and potential non-olfactory processes mediated by ORs are unknown. Here, we used molecular modeling, fingerprint interaction analysis and molecular dynamics simulations to show that the binding pocket of the prototypical olfactory receptor Olfr73 is smaller, but more flexible, than binding pockets of typical non-olfactory G-protein-coupled receptors. We extended our modeling to virtual screening of a library of 1.6 million compounds against Olfr73. Our screen predicted 25 Olfr73 agonists beyond traditional odorants, of which 17 compounds, some with therapeutic potential, were validated in cell-based assays. Our modeling suggests a molecular basis for reduced interaction contacts between an odorant and its OR and thus the typical low potency of OR-activating compounds. These results provide a proof-of-principle for identifying novel therapeutic OR agonists.

<sup>1</sup>Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. <sup>2</sup>Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. <sup>3</sup>Laboratory of Biomodelling, Faculty of Chemistry & Biological and Chemical Research Centre, University of Warsaw, 02-093 Warsaw, Poland. Correspondence and requests for materials should be addressed to S.Y. (email: [shuguang.yuan@gmail.com](mailto:shuguang.yuan@gmail.com)) or to H.V. (email: [horst.vogel@epfl.ch](mailto:horst.vogel@epfl.ch))

The mammalian olfactory system has the intriguing capability of sensing and distinguishing a huge variety of chemical compounds, often with a considerable sensitivity and specificity<sup>1–3</sup>. The perception of smell of volatile molecules starts with the activation of olfactory receptors (ORs), translating the environmental chemical signals with the involvement of heterotrimeric G proteins into neuronal electrical responses. ORs are assumed to fold into a seven-transmembrane (7TM) helix structure establishing the largest class of G-protein-coupled receptors (GPCRs)<sup>2</sup>. More than 1000 ORs in mice and about 350 ORs in humans are dedicated to smell<sup>3</sup>, capable of detecting an enormous repertoire of chemical compounds by a combinatorial coding scheme. Typically, one OR recognizes multiple odorants and one odorant is recognized by multiple ORs, but different odorants are recognized by different combinations of ORs<sup>1,3,4</sup>. Although some progress in functional identification of ligands for ORs has been achieved in recent years, most ORs are orphan receptors, whose ligand repertoires remain to be determined<sup>5–8</sup>. This, together with the lack of high-resolution 3D structures of ORs, is the major reason why the fundamental mechanistic principles of odorant recognition and discrimination by ORs are largely unresolved at a molecular level. This is in contrast to non-olfactory GPCRs for which an increasing number of high-resolution 3D structures have fostered understanding of ligand/receptor interactions, of ligand-mediated transmembrane signaling, and of how they serve as critical templates for rational drug design<sup>9</sup>.

In the case of missing experimental structural and functional data, computer modeling can deliver reliable propositions on ligand structures, which may interact and activate a specific GPCR<sup>10–14</sup>. As ORs share similarities with the structure and function of class A GPCRs<sup>15</sup>, known structures of class A GPCRs have been used as templates for homology modelling of ORs. Combined with functional data on odorant-induced activation of wild-type and mutant ORs, computer modelling have generated reliable propositions for the structure of the ligand-binding sites and specific ligand-binding modes of particular ORs thereby gaining insight into odorant selectivity<sup>16–21</sup>.

The large number of ORs and their promiscuity in ligand-binding related to the combinatorial code opens a huge chemical range of OR specific agonists and antagonists<sup>7,22,23</sup>. Estimates under discussion range from thousands to billions of OR-activating compounds<sup>1,3,24–26</sup>. Decoding the large odorant library would be of interest for several reasons. First, to understand the fundamental principles of the combinatorial code of ligand binding to ORs, and how ligand binding eventually activates or suppresses signal transmission to the intracellular side of a particular olfactory receptor and finally induce neuronal responses. Second, to understand better olfaction related diseases and to find novel therapeutic strategies for their treatment. As an example, the correlation between the appearance of neurodegenerative diseases and dysfunction of the olfactory system is well known<sup>27</sup>. Interestingly, ORs are expressed in many non-olfactory tissues including, testis, tongue, heart, spleen, pancreas, lung, kidney and placenta<sup>28,29</sup>. This suggests that ORs influence non-olfactory processes and there are indications for that in sperm chemotaxis<sup>30</sup>, embryonic development and cell–cell recognition<sup>31</sup>, chemosensory functions in kidney<sup>32</sup>, proliferation of cancer cells<sup>33,34</sup>, and other metabolic processes intimately linked with the endocrine systems<sup>35</sup> that regulate the bodies energy balance, revealing a connection between the appearance of diabetes and the dysfunction of the olfactory system<sup>36–38</sup>. To date mostly synthetic odorants have been identified to activate ectopically expressed ORs, for instance, in wound healing processes<sup>39</sup> or by influencing gut motility through OR activation and the release of serotonin in human gastrointestinal enterochromaffin cells<sup>40</sup>. In pathologic

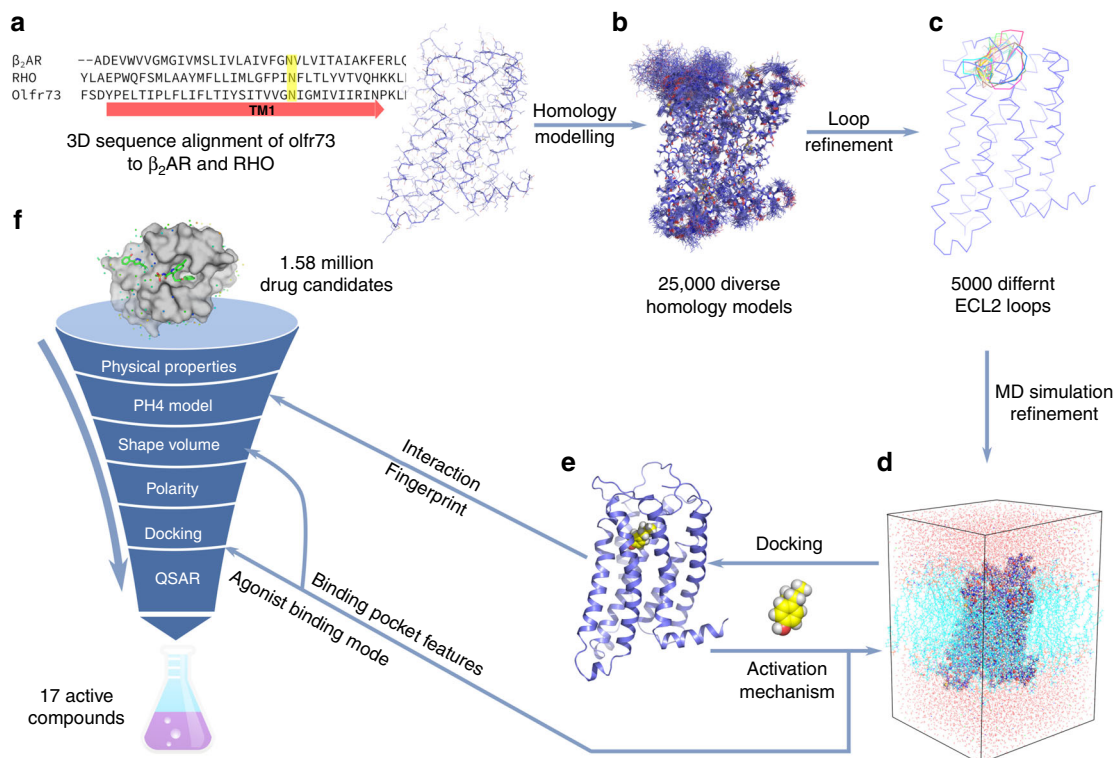
processes synthetic odorant molecules were shown to influence different stages of cancer development<sup>41,42</sup>. More recent findings suggest that ORs might be targetable not only by classical odorant molecules but even by therapeutic compounds. The anesthetic drug ketamine, which elicits various neuropharmacological effects, including sedation, analgesia, and anti-depressant activity, has been shown to activate odorant receptors in olfactory sensory neurons and in the mouse brain<sup>43</sup>. In the present study we find that Olfr73 can also be a target of drug-like molecules. Such findings are important to unravel the poly-pharmacology of therapeutic agents interacting with multiple intended but also unintended targets<sup>44</sup>.

On the other hand, OR-activating compounds have been found to also activate non-olfactory receptors<sup>45</sup>. A better understanding of the odorant-OR structure–activity relationship will have impact beyond medical applications, including production of fragrances, perfumes, food and beverages, and offering new ways of rodent and non-insect pest control, to mention a few<sup>7,46</sup>.

Here we report on a computer-directed structure-function study to discover fundamental principles of how odorants bind to a prototypical olfactory receptor and to find subsequently new receptor-activating compounds beyond traditional chemical odorant libraries. For our study, we have chosen the mouse eugenol olfactory receptor Olfr73, as this receptor has been functionally expressed in heterologous mammalian cells for compound screening<sup>16,17,47</sup>. Figure 1 outlines the approach of our study. First, we built a homology model of the structure of Olfr73 and refined it by molecular dynamics simulations. Second, we used ligand docking to identify interaction fingerprints of various agonists in the receptor's binding pocket, which together with molecular dynamics simulations of the agonist-receptor complexes revealed structural details on the first steps of receptor activation following ligand binding. Finally, the structural models served as a reliable base to perform a virtual computer screening of the ZINC library of 1.6 million drug-like molecules. The *in silico* screen selected from the large compound library considerably reduced the number of potential OR-activating compounds out of which a high percentage showed functional activity in cell-based assays. This result underscores the capacity of our *in silico* approach for rapid identification of potentially activating compounds out of large chemical libraries. Combining the 25 already known agonists with the 17 newly identified compounds establishes one of the largest collections of active compounds for a specific OR from which substantially improved conclusions on the characteristics of ligand–OR interactions can be extracted.

## Results

**Homology models of Olfr73.** To obtain a reliable 3D model of Olfr73 through the initial homology modelling, we first compared the sequence of Olfr73 with sequences of other class A GPCRs in the PDB database. We found that it shared in the best case 19% sequence identity with beta-2-adrenergic receptor ( $\beta_2$ AR, pdb code: 4LDE)<sup>48</sup> and 16% sequence identity with rhodopsin (RHO, pdb code: 4BEY)<sup>49</sup>. Since multiple modelling templates can considerably improve the reliability of homology models<sup>50–52</sup>, we used the crystal structures of both receptors as templates for model building. 3D sequence alignment (Supplementary Fig. 1) indicates that most highly conserved residues/motifs in class A GPCRs<sup>9</sup> including N<sup>1,50</sup>, D<sup>2,50</sup>, DRY motif, W<sup>4,50</sup>, Y<sup>5,58</sup>, F<sup>6,44</sup> and NPxxY motif are also present in Olfr73<sup>53</sup>. However, residue P<sup>5,50</sup> and motif CWxP that are typically found in non-OR class A GPCRs are missing in Olfr73. Moreover, one gap in TM3 is observed between the sequence of C<sup>3,25</sup> and DRY motif of Olfr73 (Supplementary Fig. 1).



**Fig. 1** Workflow of virtual screening for new agonists of Olfr73. **a** Homology modelling of Olfr73 based on 3D structure–sequence alignment of Olfr73 to  $\beta_2$ AR and Rho. **b** Refinement of ECL2 loop. **c** Molecular dynamics (MD) simulations of agonist bound Olfr73. **d** Docking agonist molecule into the 3D structural model of Olfr73. **e** Interaction fingerprint analysis by docking 25 reported compounds. This information was subsequently used for PH4 model building in virtual screening. **f** Virtual screening for Olfr73 agonists in the ZINC compound library composed of 1.58 million drug candidates. Applying stepwise selection filtering based on shape volume, ionization penalty and polarity, downscaled the chemical library successively to 204 compounds. The shape volume features are deduced from the results of MD simulations. Finally, quantitative structure–activity relationship evaluations reduced the chemical library to 64 compounds with predicted potential to activate Olfr73. Out of this final list, agonist binding modes were verified manually based on the activation mechanism deduced from MD simulations, and 25 available compounds were tested by cellular functional assays yielding 17 active compounds

### Interaction fingerprints (IFP) between agonists and Olfr73.

The final refined homology model of Olfr73 (Fig. 2) shared many of the common features of non-OR class A GPCRs outlined elsewhere<sup>9,54</sup>. We then docked isoeugenol, a potent agonist for Olfr73<sup>47</sup>, into the predicted extracellular ligand-binding pocket to explore atomistic details of the interaction between the ligand and its receptor. As depicted in Fig. 2, the hydrophobic moiety of isoeugenol is surrounded by several aromatic residues including F102<sup>3,30</sup>, F105<sup>3,33</sup>, F182<sup>ECL2</sup>, F203<sup>5,42</sup>, Y260<sup>6,52</sup> which were also found by functional analysis of mutant receptors to play a crucial role in agonist binding<sup>47</sup>. Furthermore, the three non-aromatic hydrophobic residues L199<sup>5,38</sup>, L259<sup>6,51</sup> and V277<sup>7,39</sup> are contacting the agonist molecule. The hydroxyl group in isoeugenol forms an H-bond with Y260<sup>6,52</sup>, which in turn forms an H-bond with E208<sup>5,47</sup> and water mediated hydrogen bonding to S113<sup>3,41</sup>. Both Y260<sup>6,52</sup>, E208<sup>5,47</sup> and S113<sup>3,41</sup> had been shown elsewhere to be important for Olfr73 activation<sup>47</sup>.

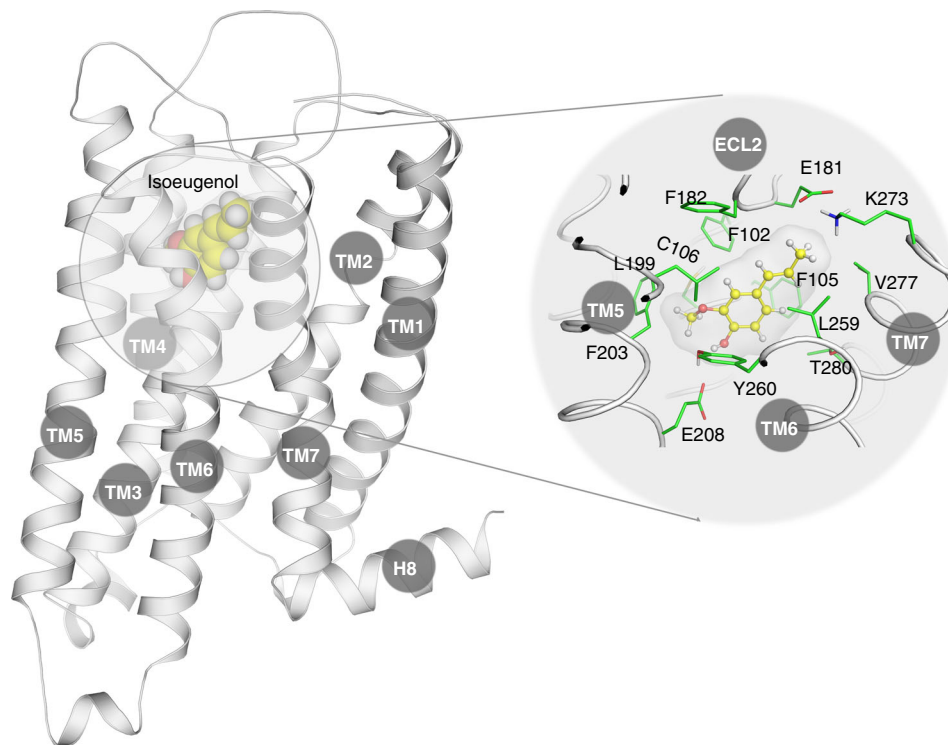
To further validate these observations, we performed an IFP analysis (Fig. 3a), which encodes specific interactions between a particular ligand and specific amino acids in the binding pocket. IFP analyses have been used for computational drug discovery for non-olfactory GPCRs<sup>55</sup>. Here we docked 25 previously reported Olfr73 agonist molecules<sup>16,45</sup> (Fig. 4) into the binding pocket of Olfr73 and obtained the interaction fingerprints of the different agonists with residues in the binding pocket (Fig. 3a). The IFP analysis showed that all docked agonists could interact with five residues in the receptor's binding pocket including F102<sup>3,30</sup>,

F105<sup>3,33</sup>, L199<sup>5,38</sup>, L259<sup>6,51</sup> and Y260<sup>6,52</sup>. Furthermore, C106<sup>3,34</sup> (80%), V109<sup>3,37</sup> (96%), E181<sup>ECL2</sup> (80%), F182<sup>ECL2</sup> (64%), F203<sup>5,42</sup> (60%), E208<sup>5,47</sup> (88%), V277<sup>7,39</sup> (52%) and T280<sup>7,42</sup> (52%) are also found frequently (percentage in parentheses) contacting the agonists (Fig. 3a, top histogram). In addition, the three residues V110<sup>3,38</sup> (5%), F179<sup>ECL2</sup> (8%) and K273<sup>7,35</sup> (12%) were found sometimes in contact with the agonists. Each particular ligand was found interacting with at least 80% of all the residues in the binding pocket (Fig. 3a, right histogram). Most of these mentioned residues were found by functional analysis of mutant receptors to play a crucial role in agonist binding<sup>47</sup>.

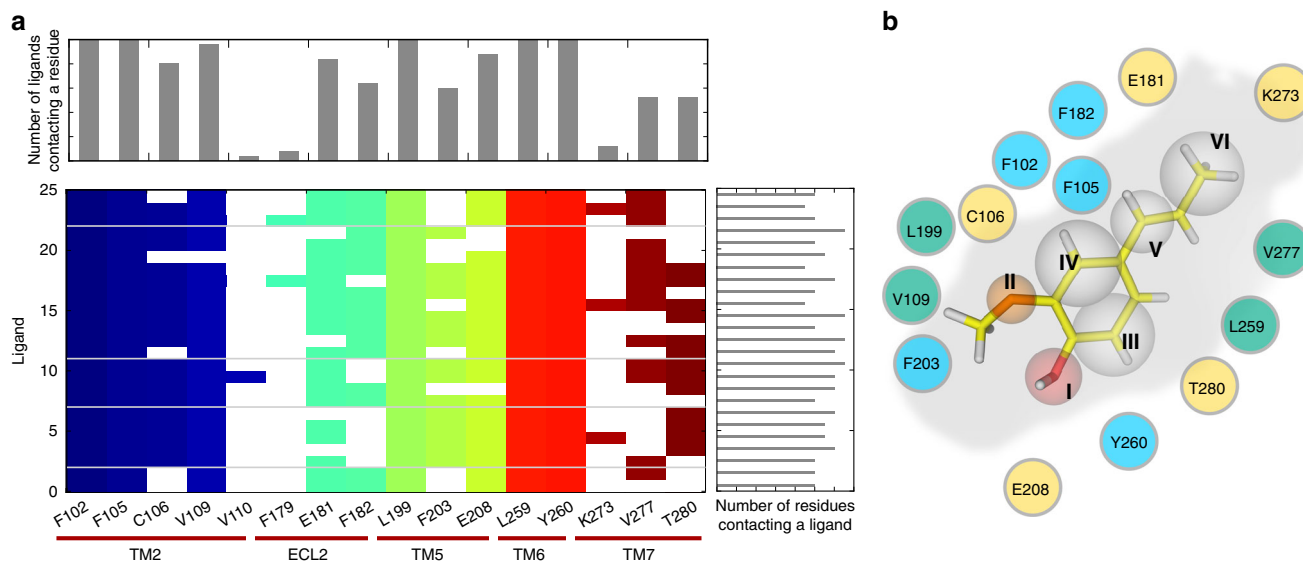
### Structural characteristics of Olfr73 from molecular dynamics simulations.

We modelled the Olfr73 with crystal structures of activated GPCRs. To explore reliable atomic details, we performed  $2 \times 500$  ns all-atom molecular dynamics simulations for both the apo form of the receptor (apo-Olfr73) and the receptor with agonist isoeugenol (iEG-Olfr73) (Fig. 2). The molecular dynamics simulations showed that the volumes of the binding pocket for apo-Olfr73 and iEG-Olfr73 were  $190 \pm 3 \text{ \AA}^3$  and  $220 \pm 3 \text{ \AA}^3$ , respectively. This is probably because of the induced fit effect (IFD), which has been widely observed in GPCR system and many others<sup>56,57</sup>.

Since the Olfr73 was modelled with low sequence identity templates, it was necessary to restrain the backbone of the modeled OR structure during molecular dynamics simulations to



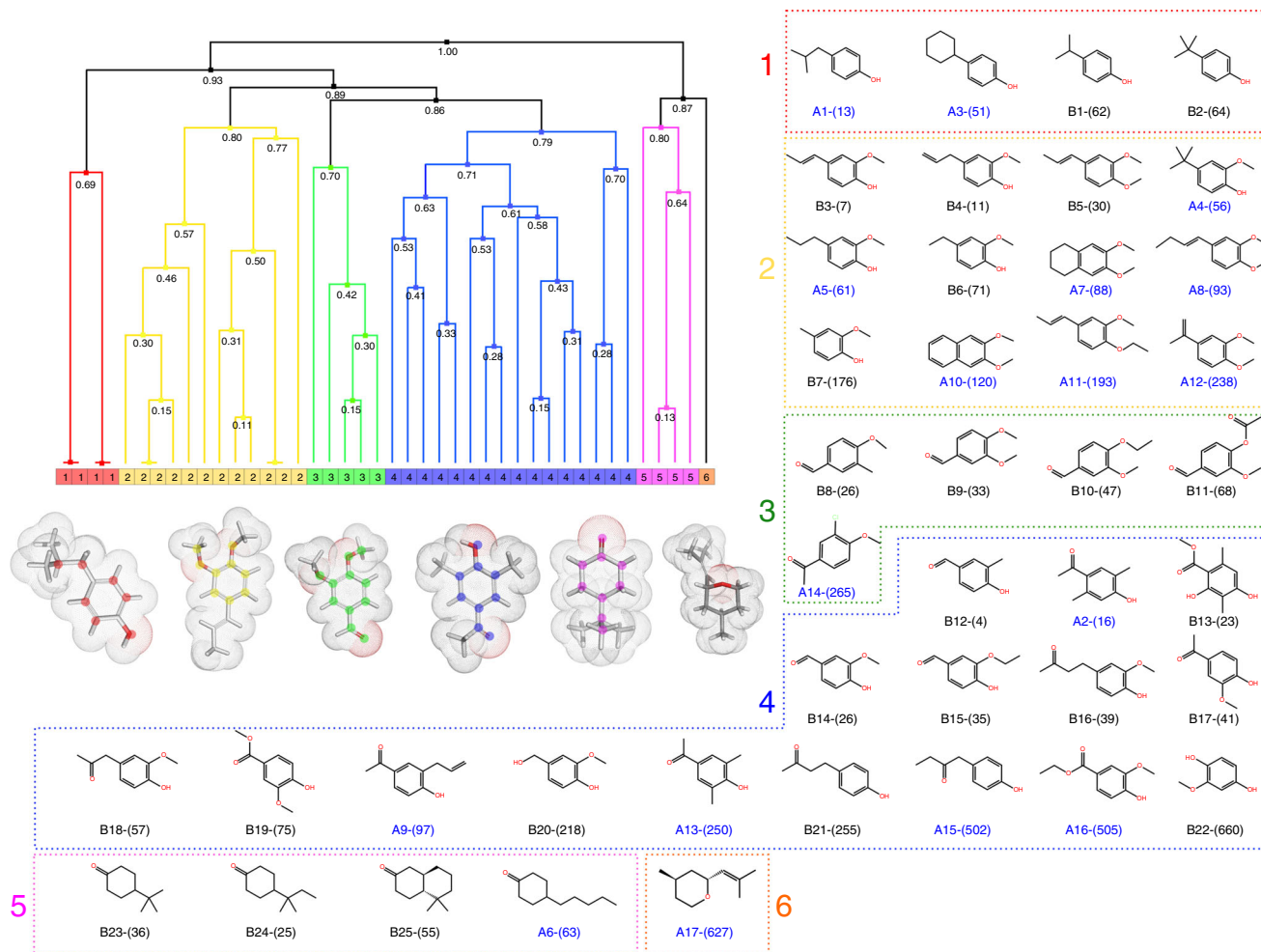
**Fig. 2** The 3D structural model of Olfr73 (left) and enlarged view of the binding mode of the agonist isoeugenol (right). Amino acid side chains in contact with bound isoeugenol are shown in green



**Fig. 3** Interaction fingerprints of 25 known Olfr73 agonists grouped in classes 1-5 according to Fig. 4. **a** In the interaction histogram, each contact of a particular residue with the ligand is indicated by a color. The color code distinguishes the residue location in a particular TM helix. Each class of compounds is separated by a horizontal gray line. **b** The pharmacophore model based on 25 known Olfr73 agonists. As a prototypical example, the position of isoeugenol in the Olfr73 binding pocket showing the interaction fingerprint. Assignments: H-bond donor (I), H-bond acceptor (II), hydrophobic moiety (III, IV, V, VI); polar residues (yellow), aromatic residues (cyan), hydrophobic residues (green)

keep correct secondary structure<sup>50,52</sup>. Thus, we added a small force constrain during all our molecular dynamics simulations (see methods section). Transmembrane (TM) movements are a hallmark of GPCR activation. Since the templates used for the molecular dynamics simulations are based on receptors in activated states, the cytoplasmic TM regions of Olfr73 have been kept in the active open conformation by the end of molecular dynamics simulations (Supplementary Fig. 2).

**Virtual agonist screening.** We established a refined 3D structural homology model of Olfr73, the agonist-receptor interaction fingerprint and the structural framework explaining the mechanism of receptor activation. To validate these findings, we performed a virtual screen on a large chemical compound library to find new candidates of agonists for Olfr73 (Fig. 1) beyond classical odorant compound libraries, which finally will be tested by cellular functional assays.



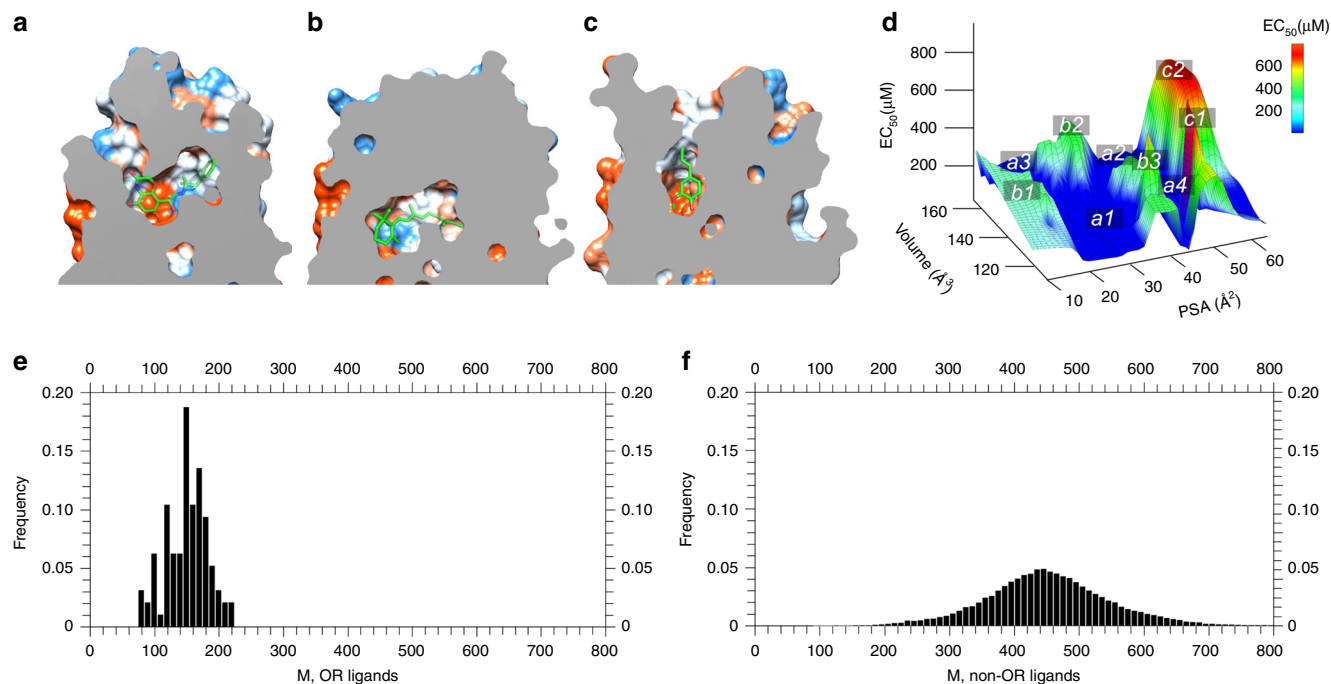
**Fig. 4** Hierarchical clustering of Olfr73 agonist molecules. Six different classes of agonists are identified (distinguished by a color code) according to their PH4 features. In the Hierarchical diagram, the links between the chemical compounds are represented as branched vertical lines. The height of the lines, coupled with merging distance (numbers showed in each node), indicate the normalized dissimilarity distance between the adjacent compounds. A higher line or a larger merging distance denotes a larger dissimilarity. A typical representative molecular structure of each class is shown below the dendrogram together with their molecular surfaces indicating hydrophobic moieties in grey and polar moieties in red. The commonly shared atoms within a certain class of molecules are labeled with colored dots accordingly. The molecular structures of the six classes of agonists are grouped in boxes. The 17 newly found agonists are represented as A1-A17 in blue. The 25 previously reported agonists are represented as B1-B25 in black. The agonist isoeugenol is B3 and p-isobutyphenol is A1. In all cases the corresponding micromolar  $EC_{50}$  values are indicated in brackets. Names of A- and B-compounds are listed in Supplementary Tables 2 and 3

First, we evaluated the physicochemical properties of all reported compounds (see Methods section on physical properties filtering) and used them for setting the conditions for an initial filter according to which 312,800 compounds were selected from the initial 1.58 million drug-like compounds of the ZINC library (Supplementary Table 1). For details about this procedure, please see the Methods section on virtual screening.

We applied the next round of selection criteria to our downsized chemical compound library using pharmacophore search (PH4)<sup>58</sup>, a screening method selecting compounds according to their chemical shape (Fig. 3b). The PH4 screen heavily relies on our results obtained from the IFP analysis and the molecular dynamics simulations. According to the molecular dynamics simulations, the oxygen at site I is crucial for agonist binding, forming distinct H-bonds with Y260<sup>6,52</sup> and E208<sup>5,47</sup>. IFP analysis further confirmed that the interaction with Y260<sup>6,52</sup> in this position is highly conserved (Fig. 3a). Since the -OH group could be either H-bond donor or H-bond acceptor, it was featured allocated preferentially the PH4 selection filter further downsized the library to 266,000 compounds.

Interestingly, the empty ligand-binding pocket of Olfr73 has a volume of 190 Å<sup>3</sup> and is noticeably smaller than that of other GPCRs including A<sub>2A</sub>R (270 Å<sup>3</sup>)<sup>59</sup>, rhodopsin (260 Å<sup>3</sup>)<sup>11</sup> (Fig. 5), the 5-HT<sub>1A</sub><sup>60</sup> receptor (360 Å<sup>3</sup>), or the μ-opioid receptor<sup>57</sup> (510 Å<sup>3</sup>), and therefore acts as a size-selection filter for potential binders. This explains why all currently reported agonists of Olfr73 are small (MW = 130–220) and the corresponding  $EC_{50}$  values are relatively high, due to the limited interactions in such small binding pocket. On this basis we created a volume counter along the 3D space of the sixteen superimposed ligands further reducing the library of potential Olfr73 binders to 493 compounds. We then continued selection filtering applying first an ionization penalty and then a molecular polarity counter, which narrowed the library further down first to 371 and then to 204 compounds.

Finally, we selected potential agonists from the remaining 204 compounds using quantitative structure–activity relationships (QSAR) based on comparative molecular field analysis (CoMFA) methods<sup>61</sup>. We docked the top 100 ranked compounds by QSAR into the MD refined homology model and found that 64



**Fig. 5** Cross-section through several GPCRs along the membrane normal showing the vertical part of the ligand-binding pocket of (a) A<sub>2A</sub>R in complex with ZMA, (b) rhodopsin in complex with retinal, and (c) Olfr73 in complex with isobutylphenol. **d** Plot of EC<sub>50</sub> values versus agonist volumes and agonist polar surface areas (PSA) for Olfr73 based on all reported agonists. Highly potent agonists are located in regions a1, a2, a3 and a4; agonists with medium potency are in the regions of b1, b2 and b3; agonists with lower potency are found in regions c1 and c2. **e** Molecular mass distribution of OR ligands. **f** Molecular mass distribution of non-olfactory GPCR ligands

compounds fitted into the ligand-binding pocket of Olfr73, close to the activation trigger F105<sup>3,33</sup>. However, only 25 out of the 64 selected compounds have been commercially available for testing biological activity.

**Cell-based functional tests.** Next, we used the SEAP reporter assay, monitoring changes in cyclic adenosine monophosphate (cAMP) second messenger signalling as a read-out for cellular responses of odorant-induced receptor activation and found ligands capable of activating Olfr73 in Hana3A cells<sup>47</sup>. We tested 25 compounds of the molecules predicted from virtual screening and identified 17 (Fig. 4, blue labelled compounds; Supplementary Table 1 and Supplementary Fig. 3) inducing a noticeable SEAP signal in a concentration-dependent activation of Olfr73. It would be interesting to test by additional experiments whether of the eight compounds, which did not show agonist activity, there are antagonists for Olfr73.

**The diversity of OR agonists.** In the following, we used a hierarchical agglomerative clustering method<sup>62</sup> to classify both the newly found and the previously known Olfr73 agonists based on their PH4 characteristics. As shown in Fig. 4, the 42 compounds can be grouped into 6 different classes. The four agonists of class-1 comprise a common phenol group with bulky hydrophobic groups (cyclohexyl or branched methyl containing alkyl chains) in *para* position. The EC<sub>50</sub> values of class-1 agonists range from 13 to 64 μM. The agonists of class-2 share a central modified pyrocatechol structure (primarily in form of monomethoxy-phenol or dimethoxy-phenol) with an additional linear, branched or cyclic hydrophobic group attached. The EC<sub>50</sub> values of these agonists range from 7 to 240 μM. The agonists of class-3 contain a central benzaldehyde structure. The *para* positions carry primarily a methoxy- or ethoxy-substitute; the *meta* positions are substituted mostly by methoxy groups or for one case by a methyl

group. The EC<sub>50</sub> of class-3 agonists range from 26 to 270 μM. The 16 agonists of class-4, share a central phenol structure with oxygen carrying groups in the *para* and sometimes also in the ortho position. The class-4 agonists are the most polar ones in our collection; ten of them show EC<sub>50</sub> values in the range of 4–100 μM, the remaining six have EC<sub>50</sub> values between 200 and 660 μM. The class-5 agonists are quite different from the initial four classes; they do not contain an aromatic ring but instead carry a central cyclohexanone structure preferentially with a linear or branched alkyl substituent at the *para* position. The four class-5 agonists show EC<sub>50</sub> values from 36 to 63 μM. Only one agonist is listed in class-6. It is composed of a tetrahydro-2H-pyran structure carrying two hydrophobic substitutions in the ring and has an EC<sub>50</sub> value of 630 μM.

**Therapeutic potential of the newly discovered agonists.** We found p-isobutylphenol (4-isobutylphenol) as the most potent ligand activating Olfr73 in our functional assay (Fig. 4). It is a known degradation product of Ibuprofen which is widely used as analgesic anti-inflammatory drug but p-isobutylphenol has also been shown to exhibit antibiotic activity<sup>63</sup>. The estrogenic activity of the compound 4-cyclohexylphenol has been documented by in-vitro assays<sup>64</sup>. The Olfr73 activating compound 4'-hydroxy-3',5'-dimethoxyacetophenone (Acetosyringone) has anti-asthmatic and anti-inflammatory properties<sup>65</sup>. And finally, 4'-hydroxypropiophenone is a predicted inhibitor of metalloproteinase 10, which has an active role in lung cancer development (Kiresee et al., 2016). Thus, our results have revealed some insights regarding the potential poly-pharmacological profile of these drugs acting not only on a defined medicinal target but also activating an OR. Similar observations of unintended interactions and activation by medicinal drugs have also been documented for the bitter taste receptor TAS2R14<sup>66</sup>. The complete list of newly discovered compounds can be found in

Supplementary Table 2. The full list of reported compounds is in Supplementary Table 3.

**Limited volume of OR binding pocket.** ORs in general and Olfr73 in particular show some interesting structural and functional differences to their class A GPCR relatives. Most of the known OR-agonists are smaller in size than the typical agonists of non-olfactory class A GPCRs. Considering a large panel of reported OR ligands<sup>67</sup> together with the new ligands (105 compounds in total) from this work shows that the molecular mass (M) of OR ligands distribute between 80 and 220 Da (maximum around 150 Da) (Fig. 5e). In contrast, non-olfactory GPCR ligands<sup>68</sup> (161,083 compounds in total) distribute primarily between 300 and 600 Da (maximum around 450 Da) (Fig. 5f). Additionally, EC<sub>50</sub> values for OR-agonists are usually much higher than those of the agonists of the non-olfactory class A GPCRs<sup>69</sup>. In our present study, this can be explained by the volume of the ligand-binding pocket of the Olfr73 which in the apo form is considerably smaller than comparable regions of the non-olfactory class A GPCRs reducing the number of interaction points between ligand and receptor (Fig. 5). Obviously, the ligand-binding pocket of a particular receptor acts as a size exclusion filter for potential ligands. To test this hypothesis and to probe the flexibility of the ligand-binding pocket for the Olfr73-agonists, we submitted Olfr73 with bound compounds A1, A2 and A3 (Fig. 4) to additional 2 × 500 ns all-atom molecular dynamics simulations. Similar to the agonist isoeugenol, the volume of the binding pocket of Olfr73 increased from 190 ± 3 Å<sup>3</sup> in the empty state to 220 ± 3 Å<sup>3</sup> for A1 (MW = 149) to 225 ± 5 Å<sup>3</sup> for A2 (MW = 166), and 240 ± 2 Å<sup>3</sup> for A3 (MW = 171), respectively. In general, larger ligands induce a larger volume increase in the occupied binding pocket<sup>56</sup>. Obviously, the binding pocket of Olfr73 is within a certain range quite flexible and adjusts perfectly to the size of the bound ligand with volume changes between 15 and 25%. We previously made similar observations for non-olfactory class A GPCRs such as the P2Y<sub>1</sub> receptor<sup>70</sup> changing the volume of the binding pocket from 230 ± 4 to 280 ± 5 Å<sup>3</sup> (22% change), the 5-HT<sub>1A</sub><sup>60</sup> receptor from 360 ± 5 to 425 ± 3 Å<sup>3</sup> (18%), the A<sub>2A</sub>R<sup>59</sup> from 270 ± 2 to 315 ± 4 Å<sup>3</sup> (17%), and the μ-opioid receptor<sup>57</sup> from 510 ± 3 to 575 ± 5 Å<sup>3</sup> (13%). In the next step, we performed an interaction fingerprint analysis for our newly found 17 agonists and compared the outcome with that in Fig. 3 of the 25 known agonists (Supplementary Fig. 4). The IFPs of both sets of agonists are quite similar. Moreover, IFPs indicated that there are only two hydrogen bond interactions between Olfr73 and its agonists, whereas there are much more polar interactions in other GPCRs<sup>71,72</sup>. These results further confirm our conclusions that the volume of Olfr73 is limited which is responsible for the small size of its agonists and weak EC<sub>50</sub> values.

## Discussion

Olfactory receptors represent almost 50% of the human GPCRs and may have additional physiological and pathological functions in the human body beyond their role in olfaction. A critical step allowing studies of the different functional roles of ORs relies on the discovery of their activating ligands. By employing a combination of homology modelling, interaction fingerprint analysis and molecular dynamics simulations to finally find novel agonists beyond classical odorant compounds by in silico screening a large drug compound library. We find that even drug-like molecules can target Olfr73, and compared their OR-activation mechanism to that mediated by classical odorant molecules.

In summary, our study revealed a structural framework for ligand–receptor interactions. We found that the limited volume

of the binding pocket is responsible for the small size and weak EC<sub>50</sub> values of Olfr73 agonists, which is typical for many ORs. Molecular fingerprint analysis discovered the principal interactions between the agonists and the binding pocket comprising many aromatic and hydrophobic residues. Interestingly, when plotting the EC<sub>50</sub> values vs. ligand volume and ligand polar surface area (PSA), we identified four favorable regions for the most potent agonists with the following characteristics (Fig. 5d): (a1) small volume (120–140 Å<sup>3</sup>) and small PSA (20–35 Å<sup>2</sup>); (a2) large volume (140–160 Å<sup>3</sup>) and medium PSA (35–40 Å<sup>2</sup>); (a3) large volume (>160 Å<sup>3</sup>) and small PSA (10–20 Å<sup>2</sup>); (a4) small volume (120–130 Å<sup>3</sup>) and medium PSA (42–45 Å<sup>2</sup>). Three regions with moderate potencies were also identified: (b1) small or medium volume (120–150 Å<sup>3</sup>) and small PSA (10–20 Å<sup>2</sup>); (b2) large volume (150–160 Å<sup>3</sup>) and small PSA (20–30 Å<sup>2</sup>); (b3) medium volume (130–150 Å<sup>3</sup>) and medium PSA (35–40 Å<sup>2</sup>). Two regions with poor potencies were also found: (c1) small volume (<130 Å<sup>3</sup>) and large PSA (45–55 Å<sup>2</sup>); (c2) large volume (>145 Å<sup>3</sup>) and large PSA (>50 Å<sup>2</sup>).

Comparing with previous work on olfactory receptors<sup>15,43,73</sup>, we found several new aspects in this area including the structural principles leading to higher EC<sub>50</sub> values for OR activation and the molecular diversity of ligands and interactions with their OR. In conclusion, molecular dynamics simulations in combination with structure based in silico screening offer a promising way to deorphanize the mammalian OR repertoire and thereby contribute to a better understanding of the molecular basis of ligand–OR interactions in olfactory and non-olfactory processes.

## Methods

**Homology modelling of Olfr73.** The initial homology models of Olfr73 were obtained by Modeller 9.10<sup>74</sup> using the crystal structure of two GPCRs in the active state as templates, β<sub>2</sub>AR (pdb code: 4LDE)<sup>48</sup> and rhodopsin (pdb: 4BEY)<sup>49</sup>; β<sub>2</sub>AR shares 19% and rhodopsin 16% sequence identity with Olfr73. They share the highest sequence identity with Olfr73. 3D multiple sequence alignments were performed by Promals3D<sup>75</sup> with default settings and were adjusted manually for properly aligning conserved motifs and disulfide bridges. 25,000 structural models (5 × 5000 with different random seeds) were created from the two template structures for Olfr73 in Modeller with fully annealing protocol, and the optimal model was chosen for further study based on Discrete Optimized Protein Energy (DOPE) score. Models from Modeller were submitted to Rosetta for kinematic loop modelling refinement<sup>76</sup>. Over 20,000 structures for loop region were generated. The loop refinement was done in Rosetta<sup>76</sup>.

**Refinement of structural model and protein–ligand docking.** The initial Olfr73 structure models generated from Modeller were optimally aligned with the structure of β<sub>2</sub>AR (pdb code: 4LDE) using OPM (Orientations of Proteins in Membranes) database<sup>77</sup>. The pre-aligned Olfr73 structure models were imported into the Maestro v9.3 program<sup>78</sup>. Hydrogens were added to the structures corresponding to physiological pH 7.0. For details please see our previous work<sup>11</sup>.

**Molecular dynamics simulations.** We performed restrained molecular dynamics simulations to achieve local improvement of the homology models<sup>52,79</sup>. Using the g\_membed<sup>80</sup> tool in Gromacs<sup>81,82</sup>, the well-prepared Olfr73 structure model was embedded into a pre-equilibrated lipid bilayer of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) solvated in 0.15 M NaCl. All molecular dynamics simulations were performed in Gromacs<sup>81,82</sup>. For details please see our previous work<sup>11,83</sup>.

**Interaction fingerprint analysis and compound clustering.** Both the interaction fingerprint (IFP) analysis and compound clustering analysis were performed in Schrodinger software<sup>78</sup>. Each compound was first docked into the binding pocket of Olfr73 in Schrodinger, then the interaction fingerprint analysis was performed using the Canvas module<sup>84</sup>. Interaction fingerprints are calculated as a set of bits for the presence or absence of particular types of interactions between a set of ligands and the residues in the active site of a receptor. The IFP divides protein/ligand interactions into four different types: hydrophobic interaction, ion-lock interaction, and H-bond interactions. The frequency of each interaction was calculated by the sum of contacts over the whole frames during molecular dynamics simulations.

The 3D coordinates of each docked compound were imported into Canvas for compound clustering analysis. Canvas first calculates the pharmacophore features

(such as H-bond interaction and hydrophobic contact) of imported compounds, then performs hierarchical agglomerative clustering on a set of structures using a similarity matrix. The Hierarchical clustering was generated in Canvas. Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. The merges and splits are determined in a greedy manner<sup>85</sup>. The results of hierarchical clustering are usually presented in a dendrogram. In order to decide which clusters should be combined or split, a measure of dissimilarity between sets of observations is required. In the hierarchical clustering, this is achieved by use of “Average distance between all inter-cluster pairs” method in Schrodinger software<sup>85</sup>, and the Kelley criterion<sup>86</sup> was used for the linkage which specifies the dissimilarity of sets as a function of the pairwise distances of observations in the sets.

**Virtual screening.** We first considered the physicochemical properties of 25 known agonists of Olfr73<sup>17,47</sup> such as molecular mass (M), calculated logP (clogP), number of rotatable bonds (nRot) and bond valence. However, we extended the scale of each criterion. For instance, as the molecular mass M of the reported compounds are within 134–218 Da, we used a filter for M of 110–320 Da.

The pharmacophore (PH4) search was performed using the Phase<sup>58</sup> module in the Schrodinger software. PH4 is a fast and efficient tool for shape-based superposition and similarity searching of pharmacophores. A scoring function rank-orders potential pharmacophores by their performance in virtual screening and ligand alignment. The PH4 models were built according to the results obtained from both IFP analysis and molecular dynamics simulations. A H-bond donor/acceptor featured sphere (site I) was created next to the -OH group of Y260<sup>6,52</sup>. A H-bond acceptor descriptor (site II) was placed next to site I. Additional four hydrophobic featured spheres (site III, IV, V and VI) were created according to the IFP analysis.

On the basis of the outcome of the PH4 search, we created a volume counter (230 Å<sup>3</sup>) along the 3D space of the 25 docked ligands to further reduce the compound library. Then we continued the selection filtering applying first an ionization penalty and then a molecular polarity counter. All these steps were performed in the Schrodinger software package.

We selected potential agonists from the remaining compound library using quantitative structure–activity relationships (QSAR) based on comparative molecular field analysis (CoMFA) methods in SYBYL-X 1.3<sup>87</sup> software. CoMFA considers steric and electrostatic interactions, which would block ligand–receptor interactions<sup>88</sup>. As a result, each molecule is located within a three-dimensional grid of defined dimensions. A probe calculates the energy within the bound molecule and neighboring residues of the receptor, in all directions over the entire grid yielding thousands of interactions<sup>88</sup>. Here we used the 25 previously reported<sup>47</sup> active compounds as training sets. Good correlations ( $R^2 = 0.73$ ) were obtained between reported experimental results from functional assays and QSAR predicted ligands.

**Cell Culture and Transfection.** As described in detail elsewhere<sup>16,45,47</sup> HEK293T-derived Hana3A cells (provided by Prof. Matsunami, Duke University, USA) were grown in DMEM/F12 medium (Invitrogen, Netherlands) supplemented with 10% fetal calf serum (FCS) (Invitrogen, Netherlands), maintained under selective conditions with 1 µg/ml of puromycin (Sigma, Switzerland) and kept in the incubator at 37 °C and 5% CO<sub>2</sub>. Cells were transfected with plasmid DNA using Lipofectamine 2000 (Invitrogen, Netherlands).

**Quantification of OR responses by SEAP reporter assay.** The assays were performed as described in detail elsewhere<sup>16,45,47</sup> twenty hours before transfection Hana3A cells were seeded into 96 wells (Greiner, Germany) at a concentration of  $3.5 \times 10^6$  cells per ml of medium. 75 ng of pRTP1S, 150 ng of the cAMP response element fused to the secreted alkaline phosphatase (pCRE-SEAP)<sup>89</sup> and either 75 ng of pOlfr73<sup>47</sup> or 75 ng of calf thymus DNA, used as a control, were co-transfected. Compounds to be tested (Sigma, Switzerland) were diluted in DMEM/F12 without FCS and added to the cells 7 h after transfection. Cells were incubated for 16 h at 37 °C in the incubator. For the SEAP reporter assay, the culture medium was mixed with an equal volume of 1 M diethanolamine-bicarbonate, pH 9.8, containing 20 mM para-nitrophenyl-phosphate (pNPP) (Sigma, Switzerland) and 1 mM MgCl<sub>2</sub> (Sigma, Switzerland). Absorbance was measured at 410 nm using a multiwell absorbance plate reader (Molecular Devices, USA) at 1- to 4-min intervals for a period of 5 min to determine SEAP expression dependent pNPP hydrolysis rates (A<sub>410</sub> /min). EC<sub>50</sub> values were determined from dose–response curves fitting experimental data with the Hill equation using IGOR Pro software (WaveMetrics):

$$f(x) = \frac{f(\max)}{1 + \left(\frac{EC_{50}}{x}\right)^n} f(x) \text{ is the background-corrected response signal at}$$

concentration  $x$  of the investigated ligand,  $f(\max)$  the maximal, background-corrected amplitude of the response signal, EC<sub>50</sub> the half maximal effective ligand concentration, and  $n$  the Hill coefficient. Experiments were performed in triplicate.

**Reporting Summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

## Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 22 September 2018 Accepted: 11 March 2019

Published online: 23 April 2019

## References

- Firestein, S. How the olfactory system makes sense of scents. *Nature* **413**, 211–218 (2001).
- Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
- Mombaerts, P. Genes and ligands for odorant, vomeronasal and taste receptors. *Nat. Rev. Neurosci.* **5**, 263–278 (2004).
- Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
- von der Weid, B. et al. Large-scale transcriptional profiling of chemosensory neurons identifies receptor–ligand pairs in vivo. *Nat. Neurosci.* **18**, 1455–1463 (2015).
- Jiang, Y. et al. Molecular profiling of activated olfactory neurons identifies odorant receptors for odors in vivo. *Nat. Neurosci.* **18**, 1446–1454 (2015).
- Dunkel, A. et al. Nature’s chemical signatures in human olfaction: a foodborne perspective for future biotechnology. *Angew. Chem.* **53**, 7124–7143 (2014).
- Peterlin, Z., Firestein, S. & Rogers, M. E. The state of the art of odorant receptor deorphanization: a report from the orphanage. *J. Gen. Physiol.* **143**, 527–542 (2014).
- Venkatakrishnan, A. J. et al. Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
- Nygaard, R. et al. The dynamic process of beta(2)-adrenergic receptor activation. *Cell* **152**, 532–542 (2013).
- Yuan, S., Filipek, S., Palczewski, K. & Vogel, H. Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nat. Commun.* **5**, 4733 (2014).
- Rose, A. S. et al. Position of transmembrane helix 6 determines receptor G protein coupling specificity. *J. Am. Chem. Soc.* **136**, 11244–11247 (2014).
- Kohlhoff, K. J. et al. Cloud-based simulations on Google Exacely reveal ligand modulation of GPCR activation pathways. *Nat. Chem.* **6**, 15–21 (2014).
- Lakkaraju, S. K., Lemkul, J. A., Huang, J. & MacKerell, A. D. Jr. DIRECT-ID: an automated method to identify and quantify conformational variations–application to beta2-adrenergic GPCR. *J. Comput. Chem.* **37**, 416–425 (2016).
- Park, J. H. et al. Opsin, a structural model for olfactory receptors? *Angew. Chem.* **52**, 11021–11024 (2013).
- Baud, O. et al. Exchanging ligand-binding specificity between a pair of mouse olfactory receptor paralogs reveals odorant recognition principles. *Sci. Rep.* **5**, 14948 (2015).
- Katada, S., Hirokawa, T., Oka, Y., Suwa, M. & Touhara, K. Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. *J. Neurosci.* **25**, 1806–1815 (2005).
- Di Pizio, A. & Niv, M. Y. Computational Studies of Smell and Taste Receptors. *Isr. J. Chem.* **54**, 1205–1218 (2014).
- Doszczak, L. et al. Prediction of perception: probing the hOR17-4 olfactory receptor model with silicon analogues of bourgeonal and linal. *Angew. Chem.* **46**, 3367–3371 (2007).
- Gelis, L., Wolf, S., Hatt, H., Neuhaus, E. M. & Gerwert, K. Prediction of a ligand-binding niche within a human olfactory receptor by combining site-directed mutagenesis with dynamic homology modeling. *Angew. Chem.* **51**, 1274–1278 (2012).
- Bavan, S., Sherman, B., Luetje, C. W. & Abaffy, T. Discovery of novel ligands for mouse olfactory receptor MOR42-3 using an in silico screening approach and in vitro validation. *PLoS ONE* **9**, e92064 (2014).
- Sell, C. S. *Chemistry and the Sense of Smell* (John Wiley & Sons, Inc., Hoboken, 2014).
- Ruddigkeit, L., Awale, M. & Reymond, J. L. Expanding the fragrance chemical space for virtual screening. *J. Cheminform.* **6**, 27 (2014).
- Gerkin, R. C. & Castro, J. B. The number of olfactory stimuli that humans can discriminate is still unknown. *eLife* **4**, e08127 (2015).
- Meister, M. On the dimensionality of odor space. *eLife* **4**, e07865 (2015).
- Bushdid, C., Magnasco, M. O., Vossball, L. B. & Keller, A. Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **343**, 1370–1372 (2014).
- Doty, R. L. Olfactory dysfunction in Parkinson disease. *Nat. Rev. Neurol.* **8**, 329–339 (2012).
- Flegel, C., Mantoniotis, S., Osthold, S., Hatt, H. & Gisselmann, G. Expression profile of ectopic olfactory receptors determined by deep sequencing. *PLoS ONE* **8**, e55368 (2013).



29. Kang, N. & Koo, J. Olfactory receptors in non-chemosensory tissues. *BMB Rep.* **45**, 612–622 (2012).
30. Spehr, M. et al. Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science* **299**, 2054–2058 (2003).
31. Feldmesser, E. et al. Widespread ectopic expression of olfactory receptor genes. *BMC Genom.* **7**, 121 (2006).
32. Pluznick, J. L. et al. Functional expression of the olfactory signaling system in the kidney. *Proc. Natl Acad. Sci. USA* **106**, 2059–2064 (2009).
33. Wiese, H. et al. Quantitative phosphoproteomics reveals the protein tyrosine kinase Pyk2 as a central effector of olfactory receptor signaling in prostate cancer cells. *Biochim. Biophys. Acta* **1854**, 632–640 (2015).
34. Neuhaus, E. M. et al. Activation of an olfactory receptor inhibits proliferation of prostate cancer cells. *J. Biol. Chem.* **284**, 16218–16225 (2009).
35. Fleischer, J., Bumbalo, R., Bautze, V., Strotmann, J. & Breer, H. Expression of odorant receptor Olfr78 in enteroendocrine cells of the colon. *Cell Tissue Res.* **361**, 697–710 (2015).
36. Kang, N. et al. Olfactory receptor Olfr544 responding to azelaic acid regulates glucagon secretion in alpha-cells of mouse pancreatic islets. *Biochem. Biophys. Res. Commun.* **460**, 616–621 (2015).
37. Varkonyi, T., Korei, A., Putz, Z. & Kempler, P. Olfactory dysfunction in diabetes: a further step in exploring central manifestations of neuropathy? *Angiology* **65**, 857–860 (2014).
38. Palouzier-Paulignan, B. et al. Olfaction under metabolic influences. *Chem. Senses* **37**, 769–797 (2012).
39. Busse, D. et al. A synthetic sandalwood odorant induces wound-healing processes in human keratinocytes via the olfactory receptor OR2AT4. *J. Invest. Dermatol.* **134**, 2823–2832 (2014).
40. Braun, T., Voland, P., Kunz, L., Prinz, C. & Gratzl, M. Enterochromaffin cells of the human gut: sensors for spices and odorants. *Gastroenterology* **132**, 1890–1901 (2007).
41. Massberg, D. et al. Monoterpene (–)-citronellal affects hepatocarcinoma cell signaling via an olfactory receptor. *Arch. Biochem. Biophys.* **566**, 100–109 (2015).
42. Weber, L. et al. Activation of odorant receptor in colorectal cancer cells leads to inhibition of cell proliferation and apoptosis. *PLoS ONE* **12**, e0172491 (2017).
43. Ho, J. et al. Molecular recognition of ketamine by a subset of olfactory G protein-coupled receptors. *Sci. Signal.* **8**, ra33 (2015).
44. Keiser, M. J. et al. Predicting new molecular targets for known drugs. *Nature* **462**, 175–181 (2009).
45. Pick, H. et al. Dual activities of odorants on olfactory and nuclear hormone receptors. *J. Biol. Chem.* **284**, 30547–30555 (2009).
46. Gilbert, A. N. & Firestein, S. Dollars and scents: commercial opportunities in olfaction and taste. *Nat. Neurosci.* **5**, 1043–1045 (2002).
47. Baud, O. et al. The mouse eugenol odorant receptor: structural and functional plasticity of a broadly tuned odorant binding pocket. *Biochemistry* **50**, 843–853 (2011).
48. Ring, A. M. et al. Adrenaline-activated structure of beta2-adrenoceptor stabilized by an engineered nanobody. *Nature* **502**, 575–579 (2013).
49. Singhal, A. et al. Insights into congenital stationary night blindness based on the structure of G90D rhodopsin. *EMBO Rep.* **14**, 520–526 (2013).
50. Flohil, J. A., Vriend, G. & Berendsen, H. J. Completion and refinement of 3-D homology models with restricted molecular dynamics: application to targets 47, 58, and 111 in the CASP modeling competition and posterior analysis. *Proteins* **48**, 593–604 (2002).
51. Larsson, P., Wallner, B., Lindahl, E. & Elofsson, A. Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci.* **17**, 990–1002 (2008).
52. Fan, H. & Mark, A. E. Refinement of homology-based protein structures by molecular dynamics simulation techniques. *Protein Sci.* **13**, 211–220 (2004).
53. Isberg, V. et al. Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **36**, 22–31 (2015).
54. Granier, S. & Kobilka, B. A new era of GPCR structural and chemical biology. *Nat. Chem. Biol.* **8**, 670–673 (2012).
55. Vass, M. et al. Molecular interaction fingerprint approaches for GPCR drug discovery. *Curr. Opin. Pharmacol.* **30**, 59–68 (2016).
56. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **49**, 534–553 (2006).
57. Yuan, S. et al. The mechanism of ligand-induced activation or inhibition of mu- and kappa-opioid receptors. *Angew. Chem.* **54**, 7560–7563 (2015).
58. Sastry, G. M., Dixon, S. L. & Sherman, W. Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring. *J. Chem. Inf. Model.* **51**, 2455–2466 (2011).
59. Yuan, S., Hu, Z., Filipek, S. & Vogel, H. W246 opens a gate for a continuous intrinsic water pathway during activation of the adenosine receptor. *Angew. Chem.* **54**, 556–559 (2014).
60. Yuan, S., Peng, Q., Palczewski, K., Vogel, H. & Filipek, S. Mechanistic studies on the stereoselectivity of the serotonin 5-HT1A receptor. *Angew. Chem.* **55**, 8661–8665 (2016).
61. Parra-Delgado, H. et al. Synthesis and comparative molecular field analysis (CoMFA) of argentatin B derivatives as growth inhibitors of human cancer cell lines. *Bioorg. Med. Chem.* **14**, 1889–1901 (2006).
62. Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *J. Classif.* **31**, 274–295 (2014).
63. Caviglioli, G. et al. Identification of degradation products of ibuprofen arising from oxidative and thermal treatments. *J. Pharm. Biomed. Anal.* **30**, 499–509 (2002).
64. Ogawa, Y. et al. Estrogenic activities of chemicals related to food contact plastics and rubbers tested by the yeast two-hybrid assay. *Food Addit. Contam.* **23**, 422–430 (2006).
65. Muller, A. A., Reiter, S. A., Heider, K. G. & Wagner, H. Plant-derived acetophenones with antiasthmatic and anti-inflammatory properties: inhibitory effects on chemotaxis, right angle light scatter and actin polymerization of polymorphonuclear granulocytes. *Planta Med.* **65**, 590–594 (1999).
66. Levit, A. et al. The bitter pill: clinical drugs that activate the human bitter taste receptor TAS2R14. *FASEB J.* **28**, 1181–1197 (2014).
67. Modena, D., Trentini, M., Corsini, M., Bombaci, A. & Giorgetti, A. OlfactionDB: a database of olfactory receptors and their ligands. *Adv. Life Sci.* **1**, 5 (2011).
68. Gaulton, A. et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2012).
69. Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J. D. Odor coding by a Mammalian receptor repertoire. *Sci. Signal.* **2**, ra9 (2009).
70. Yuan, S. et al. The molecular mechanism of P2Y1 receptor activation. *Angew. Chem. Int. Ed.* **55**, 10331–10335 (2016).
71. Huang, W. et al. Structural insights into mu-opioid receptor activation. *Nature* **524**, 315–321 (2015).
72. Chung, K. Y. et al. Conformational changes in the G protein Gs induced by the beta2 adrenergic receptor. *Nature* **477**, 611–615 (2011).
73. Poivet, E. et al. Applying medicinal chemistry strategies to understand odorant discrimination. *Nat. Commun.* **7**, 11157 (2016).
74. Eswar, N., et al. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **15**, 5.6.1–5.6.30 (2006).
75. Pei, J. & Grishin, N. V. PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics* **23**, 802–808 (2007).
76. Mandell, D. J., Coutsiaris, E. A. & Kortemme, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **6**, 551–552 (2009).
77. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: orientations of proteins in membranes database. *Bioinformatics* **22**, 623–625 (2006).
78. Schrödinger Release 2018-2: Maestro, Schrödinger, LLC, New York, NY, USA (2018).
79. Thompson, J. D., Gibson, T. J. & Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics* Chapter 2, (2002).
80. Wolf, M. G., Hoefling, M., Aponte-Santamaria, C., Grubmüller, H. & Groenhof, G. g\_membed: efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *J. Comput. Chem.* **31**, 2169–2174 (2010).
81. Pronk, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
82. Hess, B. et al. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2009).
83. Chan, H. C. S. et al. Exploring a new ligand binding site of G protein-coupled receptors. *Chem. Sci.* **9**, 6480–6489 (2018).
84. Singh, J., Deng, Z., Narale, G. & Chuauqui, C. Structural interaction fingerprints: a new approach to organizing, mining, analyzing, and designing protein-small molecule complexes. *Chem. Biol. Drug. Des.* **67**, 5–12 (2006).
85. Schrödinger Suite 2012: Maestro, version 9.3.5. Schrödinger, LLC; New York, NY, USA.
86. Kelley, L. A., Gardner, S. P. & Sutcliffe, M. J. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* **9**, 1063–1065 (1996).
87. Sybyl-X Molecular Modeling Software Packages, Version 1.3. TRIPOS Associates, Inc; St. Louis, MO, USA (2012).
88. Ghasemi, J. B. & Tavakoli, H. Improvement of the prediction power of the CoMFA and CoMSIA models on histamine h3 antagonists by different variable selection methods. *Sci. Pharm.* **80**, 547–566 (2012).
89. Durocher, Y. et al. A reporter gene assay for high-throughput screening of G-protein-coupled receptors stably or transiently expressed in HEK293 EBNA cells grown in suspension culture. *Anal. Biochem.* **284**, 316–326 (2000).

## Acknowledgements

Most computational work has been done at the Interdisciplinary Centre for Mathema-

tical and Computational Modelling in Warsaw (grant no. GB71-3 and GB70-3). S.F. was funded by the National Center of Science, Poland (grant no. 2013/08/M/ST6/00788). H. V. was supported by the European Community (project SynSignal, grant no. FP7-KBBE-2013-613879) and funds of the EPFL.

### Author contributions

S.Y. and H.V. initiated and designed the project. S.Y. performed the modelling and molecular dynamics simulations. S.Y. and S.F. analyzed the results of modelling. T.D., M.C.B. and H.P. performed biological testing. All authors wrote the manuscript.

### Additional information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s42003-019-0384-8>.

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

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

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



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

© The Author(s) 2019