Modeling the Orthosteric Binding Site of the G Protein-Coupled Odorant Receptor OR5K1

Alessandro Nicoli,[1‡] Franziska Haag,[1‡] Patrick Marcinek,[1] Ruiming He,[1,2] Johanna Kreißl,[1] Jörg Stein,[1] Alessandro Marchetto,[3,4] Andreas Dunkel,[1] Thomas Hofmann,[5] Dietmar Krautwurst,[1] Antonella Di Pizio[1*]*

[1] Leibniz Institute for Food Systems Biology at the Technical University of Munich, 85354 Freising, Germany

[2] Department of Chemistry, Technical University of Munich, D-85748 Garching, Germany

[3] Institute for Advanced Simulations (IAS)-5/Institute for Neuroscience and Medicine (INM)-9, Forschungszentrum Jülich, 52428 Jülich, Germany

[4] Faculty of Mathematics, Computer Science and Natural Sciences, RWTH Aachen University, 52062 Aachen, Germany

[5] Chair of Food Chemistry and Molecular Sensory Science, Technical University of Munich, 85354 Freising, Germany

18 **ABSTRACT**: With approximately 400 encoding genes in humans, odorant receptors (ORs) are

19 the largest subfamily of class A G protein-coupled receptors (GPCRs). Despite its high

20 relevance and representation, the odorant-GPCRome is structurally poorly characterized: no

21 experimental structures are available and the low sequence identity of ORs to experimentally

22 solved GPCRs is a major challenge for their modeling. Moreover, the receptive range of most

23 ORs is unknown. The odorant receptor OR5K1 was recently and comprehensively

24 characterized in terms of cognate agonists. Here we investigate the binding modes of identified

25 ligands into the OR5K1 orthosteric binding site using structural information both from AI-

26 driven modeling, as recently released in the AlphaFold Protein Structure Database, and from

27 template-based modeling. Induced-fit docking simulations were used to sample the binding site

28 conformational space for ensemble docking. Side chain residue sampling and model selection

29 were guided by mutagenesis data. We obtained models that could better rationalize the different

30 activity of active (agonist) versus inactive molecules with respect to starting models, and also

31 capture differences in activity related to small structural differences. We, therefore, provide a

32 model refinement protocol that can be applied to model the orthosteric binding site of ORs as

33 well as that of GPCRs with low sequence identity to available templates.

34

35

INTRODUCTION

G protein-coupled receptors (GPCRs) are the largest family of membrane proteins in the human genome. Through interaction with their modulators, GPCRs mediate the communication between the cell and the extracellular environment and are therefore involved in almost all physiological functions.[1-4] Commonly, GPCRs are grouped into six classes based on the phylogenetic analysis: A (rhodopsin-like), B (secretin-like), C (metabotropic glutamate receptors), D (pheromone receptors), E (cAMP receptors), and F (frizzled/smoothened receptors).[5-6] Class A GPCRs consist of over 80% of all GPCRs and are the targets of 34% of all drugs present in the market.[7-8]

Class A GPCRs share a basic architecture consisting of a bundle of seven transmembrane α-helices (TM1-TM7) connected by three intracellular loops (ICLs) and three extracellular loops (ECLs), a relatively short N-terminus in the extracellular region, and a short helix 8 connected to the C-terminus in the intracellular module. The ligand-binding domain of class A GPCRs, commonly referred to as the orthosteric binding site, is located in the EC part of the 7TM bundle (made up of residues belonging to TM3, TM5, TM6, and TM7), and has high structural diversity among different receptor subtypes. The 7TM bundle is the most structurally conserved component of the class A GPCR structures, presenting characteristic hydrophobic patterns and functionally important signature motifs.[9-10]

Odorant receptors (ORs), with approximately 400 encoding genes in humans, are the largest subfamily of class A GPCRs.[11-15] Mammalian odorant receptors are split into two phylogenetically distinct groups, class I and class II ORs, which can be distinguished by some characteristic features that are highly conserved within their sequences.[16-19] ORs present most of the class A GPCR signature motifs, despite an overall low sequence identity with the non-sensory class A GPCRs.[20-21] The orthosteric binding site of ORs was also found to coincide with that of non-sensory class A GPCRs.[20-25]

3

61   The olfactory system uses a combinatorial code of ORs to represent thousands of odorants: a

62   specific OR type may recognize more than one odorant, and each odorant may be recognized

63   by more than one OR.[26-31] Despite current efforts in assigning ORs to odorant molecules, or,

64   vice versa, in defining the chemical ligand space of individual ORs, only the molecular

65   recognition ranges of a few ORs have been investigated.[27, 32-37]

66   Structure-based virtual screening campaigns have been successfully applied for GPCR ligand

67   discovery and are always more in use with the recent extraordinary advances in GPCR structural

68   biology.[38] Currently, no experimental structures of human ORs are available, and homology

69   modeling techniques have been used to rationalize the binding modes of odorant compounds

70   into ORs and discover new OR ligands.[37, 39-43] AI-based methods are emerging as compelling

71   tools to predict the 3D structure of proteins.[44-46] During the CASP (Critical Assessment of

72   Structure Prediction) 14 competition, AlphaFold 2 (AF2) was shown to be able to predict the

73   structure of protein domains at an accuracy matching experimental methods.[47] A database of

74   over 360,000 protein models across 21 species was released and is scheduled to grow to cover

75   over 100 million proteins (https://alphafold.ebi.ac.uk/).[48-49] The database expands the coverage

76   for GPCR structures, including 4,192 proteins annotated as odorant receptors, 97% of which

77   are mammalian.[45]

78   In this paper, we used both AlphaFold 2 and template-based modeling methodologies for

79   OR5K1 structural prediction. OR5K1 has been recently characterized as the specialized OR for

80   the detection of pyrazine-based key food odorants and semiochemicals.[50] We investigated the

81   interaction of the set of identified agonists within the binding site of OR5K1 and used ligand

82   information and mutagenesis data to guide the model refinement process.
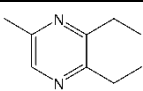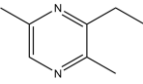
83
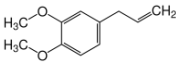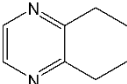
84

85

RESULTS AND DISCUSSION
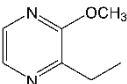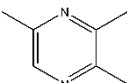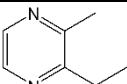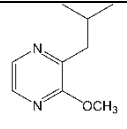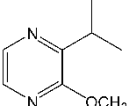
OR5K1 agonists. Pyrazines are known for contributing greatly to the aroma of roasted foods,[51-53] but they are also renowned as semiochemicals,[54-58] namely compounds that transfer chemical cues between individuals of the same and/or different species, most often eliciting a standardized behavior.[59] Recently, OR5K1 was characterized as a specialized odorant receptor for the detection of pyrazine-based key food odorants and semiochemicals.[50] The most potent compound is compound **1** (2,3-diethyl-5-methylpyrazine, $EC_{50} = 10.29 \mu M$). Compounds tested against OR5K1 include molecules with shorter or missing aliphatic chains to the pyrazine moiety (compounds **4**, **6**, **7**, **12**). We also know that the pyrazine itself does not activate this receptor.[50] Therefore, the activity of OR5K1 molecules is supposed to rely on the presence and position of the aliphatic chains (Table 1). Interestingly, in the screening of pyrazines, the mixture of isomers 2-ethyl-3,5(6)-dimethylpyrazine was found to activate OR5K1 with an EC_{50} of 21.18 μM.[50] In this work, 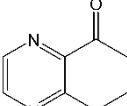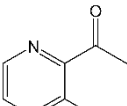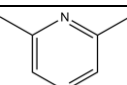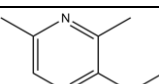we isolated the mixture and tested the individual isomers against OR5K1. We found that 2-ethyl-3,6-dimethylpyrazine (compound **2**) has an EC_{50} of 14.85 μM, while 2-ethyl-3,5-dimethylpyrazine (compound **13**) could not be measured to saturation with the concentration range available. This provides precise information on the contribution of the ethyl groups attached to the pyrazine ring.

Table 1. OR5K1 agonists and EC_{50} values. Data for compounds **1**, **3-12** are retrieved from literature,[50] while data for compounds **2** and **13** were tested in this work.

Compound	Name	Structure	CAS	EC_{50} (μM)
1	2,3-Diethyl-5-methylpyrazine		18138-04-0	10.29 ± 1.06

2	2-Ethyl-3,6-dimethylpyrazine		13360-65-1	14.85 ± 6.69
3	Methyl eugenol		93-15-2	62.21 ± 1.45
4	2,3-Diethylpyrazine		15707-24-1	94.36 ± 11.90
5	2-Ethyl-3-methoxypyrazine		25680-58-4	97.4 ± 15.59
6	2,3,5-Trimethylpyrazine		14667-55-1	139.04 ± 7.08
7	2-Ethyl-3-methylpyrazine		15707-23-0	537.87 ± 96.79
8	2-Isobutyl-3-methoxypyrazine		24683-00-9	177.94 ± 24.89
9	2-Isopropyl-3-methoxypyrazine		25773-40-4	145.63 ± 8.83
10	2-Acetyl-3-ethylpyrazine		32974-92-8	527.76 ± 167.17
11	2-Acetyl-3-methylpyrazine		23787-80-6	531.22 ± 27.59
12	2,6-Dimethylpyrazine		108-50-9	543.92 ± 19.50
13	2-Ethyl-3,5-dimethylpyrazine		13925-07-0	≥ 300*

* The last concentration that has been experimentally investigated is 300 µM. Concentration-response curves are shown in Figure 4.

6

111   **OR5K1 structure prediction.** ORs and chemosensory GPCRs share low sequence similarity

112   (below 20%) with experimentally solved GPCRs.[20, 60] The accuracy of 3D structures obtained

113   by homology modeling is highly dependent on the templates. Good models of membrane

114   proteins can be obtained for template sequence identities higher than 30%.[61] A multi-template

115   homology modeling approach has been used for successfully modeling different ORs, including

116   OR51E1 and OR7D4.[23, 62] In this approach, conserved motifs were used to guide the sequence

117   alignment of odorant receptors; bovine Rhodopsin (bRho), human β2-adrenergic (hβ2AR),

118   human Adenosine-2A (hA2A), and human Chemokine-4 (hCXCR4) receptors were used as
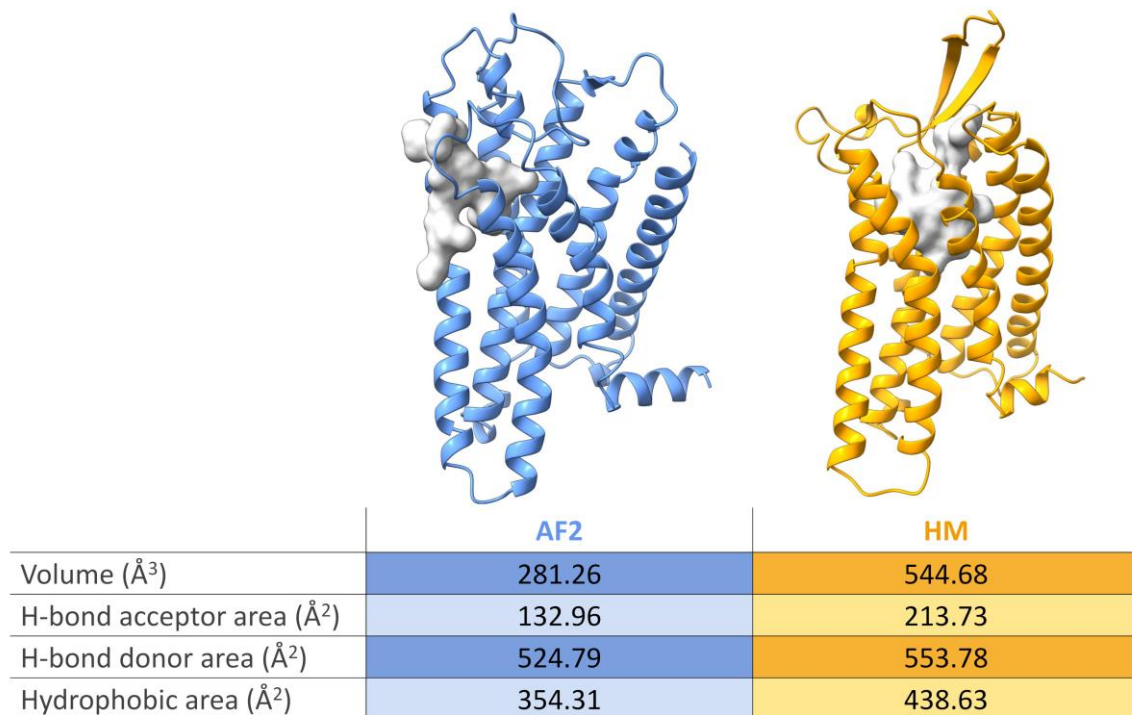
119   templates.[21]

120   OR5K1 shares 15-19% sequence identity with these templates (Figure S1). Considering that we

121   aimed to use the model to investigate the binding modes of agonists, we built the 3D structure

122   of OR5K1 using bRho, hβ2AR, and hA2A in their active state, while hCXCR4 is only available

123   in its inactive state.[38] The extracellular loop 2 (ECL2) of the templates is much shorter than the

124   ECL2 of OR5K1 (Figure S2). ECL2 is the largest and most structurally diverse extracellular

125   loop of GPCRs,[63] and those of ORs are among the longest ECL2 in class A GPCR.[64] Loop

126   modeling is highly challenging when sequence length reaches the size of the ECL2.[65-67] We

127   remodeled this region using templates with higher similarities in terms of length and sequence

128   composition (Figures S2 and S3). Specifically, we used the ECL2 of NPY2 and CCK1 receptors

129   as templates for the segment before the conserved $Cys^{45.50}$ ($S157^{4.57}$- $Y179^{ECL2}$) and the Apelin

130   receptor for the segment after the $Cys^{45.50}$ ($C180^{45.50}$- $L188^{5.37}$).

131   We   then   downloaded   the   Alphafold   2   (AF2)   structure   of   OR5K1

132   (https://alphafold.ebi.ac.uk/entry/Q8NHB7) to compare it with our homology model (HM).

133   Except for the N-Terminus and the ECL3, the per-residue confidence score (average predicted

134   local distance difference test, pLDDT) of all regions of the model is >90 (very high) or between

135   70 and 90 (confident) (Figure S3). The OR5K1 AF2 model is also among the high confidence

7

136 AF2 GPCR models, as assessed by the per-model $pLDDT_{80}$ score, which was suggested as a

137 potential criterion to assess the quality of AF2 models for structure-based virtual screening.[68]

138 AF2 and HM models have a Root Mean Square Deviation (RMSD) of the alpha carbons of 3.26

139 Å. We observed a major difference in the TM5 conformation, which is closer to the orthosteric

140 binding site in the HM than in the AF2 model. We calculated the GPCR activation index of the

141 two models using the A100 tool,[69] confirming that the HM is in its active state with an activation

142 index of 68.46, but AF2 is an inactive model with an activation index of -21.30. This is because

143 the OR5K1 HM was modeled using most of the templates in the active state conformation,

144 instead, AF2 was generated with algorithms that do not necessarily take into consideration the

145 activity state.

146 To assess the predictive ability of the HM and AF2 models, we performed molecular docking

147 calculations of known ligands as actives (13 compounds, Table 1) and with all the compounds

148 that did not elicit receptor response with a defined chirality (131 compounds, the complete list

149 with SMILES is available at https://github.com/dipizio/OR5K1_binding_site) as inactives, and

150 we then evaluated the performance of each model through Receiver Operating Characteristic

151 (ROC) analysis.[70-71] The Area Under the Curve (AUC) values are similar for HM (0.67) and

152 AF2 (0.68), and the enrichment factor in the top 15% of the sorted screened molecules ($EF_{15\%}$)

153 is very low in both cases, 0.11 and 0.24 for HM and AF2, respectively ($EF_{15\% \, max} = 1.63$) (Figure

154 S4). The AF2 model is not able to dock the most potent agonists in our set. The only highly

155 ranked agonist in both HM and AF2 models is compound **9** ($EC_{50} = 527.76$ μM), with docking

156 scores of -5.68 and -4.91 kcal/mol, respectively. As expected, HM and AF2 models have

157 different residue arrangements in the binding site, but, surprisingly, also the location of the

158 predicted binding pocket is different (Figure 1). The orthosteric binding site of AF2 is not

159 accessible, the pocket calculated with Sitemap (Schrödinger Release 2021-3: SiteMap,

160 Schrödinger, LLC, New York, NY, 2021)[72-73] is located between TM5 and TM6 and extends

161 towards the membrane bilayer (Figure 1). Indeed, the location of the orthosteric binding site is

162 partially occluded by the ECL2. The ECL2 folding is the most evident difference between the

163 two models: we modeled the HM as an anti-parallel β-sheet, instead AF2 carries out an

164 unstructured loop with a small α–helix that enters the orthosteric binding site.

	AF2	HM
Volume (Å³)	281.26	544.68
H-bond acceptor area (Å²)	132.96	213.73
H-bond donor area (Å²)	524.79	553.78
Hydrophobic area (Å²)	354.31	438.63

165

166 **Figure 1. OR5K1 starting models from AF2 (in blue) and HM (in orange).** SiteMap volume, H-bond acceptor

167 area, H-bond donor area, and hydrophobic area are reported for both models.

168

169 Moreover, the secondary structure of the terminal region of TM6 is not well defined in the AF2,

170 this portion is classified with local prediction confidence pLDDT between 70 and 90 for the

171 helix part and lower than 70 for the ECL3 part (Figure S3). The initial part of TM7 is also

172 different between the two models, there is a shift of one position in the helix and therefore

173 different residue arrangements.

174   **OR5K1 model refinement.** AF2 and HM models propose two different ligand positions and

175   binding poses. We performed induced-fit docking (IFD) simulations (Schrödinger Release

176   2021-3: Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2021; Prime,

177   Schrödinger, LLC, New York, NY, 2021)[74] with the most active compounds (compound **1**) for

178   both AF2 and HM, allowing the flexibility of the binding site side chains to explore the

179   conformational space of the orthosteric binding site of the two models. 44 models were

180   generated starting from the AF2 model and 57 from HM. The ROC curves of these models

181   show an improvement in their performance, the best models have AUC values of 0.81 and 0.85,

182   and $EF_{15\%}$ of 0.24 and 0.50 for AF2 and HM, respectively (Figure S5). The binding modes of

183   compound **1** in the best models of AF2 and HM are different but the ligand is now located in

184   the core of the orthosteric binding site in both models (Figure S5). Interestingly, we noticed that

185   two leucine residues, $L104^{3.32}$ and $L255^{6.51}$, are predicted to be in the binding pocket by both

186   models (Figure S5). Odorant molecules are typically small organic compounds of less than 300

187   Da with high-to-moderate hydrophobicity and their binding to ORs is driven by shape

188   complementarity and mostly hydrophobic interactions.[64, 75]

189    $L104^{3.32}$ is conserved in 98% of orthologs investigated across 51 species, except for the receptor

190   of the new world monkey *Aotus nancymaae* (XP_012332612.1), where a rather conservative

191   amino acid exchange replaced the leucine at position 104 by an isoleucine (Figure S7, Table

192   S5). Similarly, $L255^{6.51}$ of OR5K1 is conserved in 96% of all orthologs, except for the receptors

193   of *Aotus nancymaae*, *Loxodonta africana* (African elephant, XP_003418985.1), and

194   *Urocitellus parryii* (Arctic ground squirrel, XP_026258216.1). In all three orthologs and in the

195   human paralog OR5K2, again, a rather conservative amino acid exchange replaced the leucine

196   at position 255 by an isoleucine (Figure S7, Table S5). Single nucleotide missense variations

197   have been reported for both amino acid positions, $L104^{3.32}I$ (rs777947557) and $L255^{6.51}F$

198   (rs1032366530) in human OR5K1, albeit with frequencies way below 0.01. Moreover, both

10

199   positions $L104^{3.32}$ and $L255^{6.51}$ are part of a set of 22 amino acids that have been suggested

200   previously to constitute a generalized odorant binding pocket in ORs.[76] Both amino acid

201   positions have been identified also experimentally as odorant interaction partners in different

202   receptors by several independent studies.[24, 36, 62, 77-82] Therefore, these leucine residues are likely

203   to play a relevant role in the ligand recognition of OR5K1 agonists. We mutated these residues

204   to alanine ($L104^{3.32}$A, $L255^{6.51}$A) and found that there is a shift in EC_{50} values for both mutants

205   when stimulated with compound **1**: EC_{50} of 525.28 ± 92.28 μM for $L104^{3.32}$A and EC_{50} of

206   478.36 ± 185.10 μM for OR5K1 $L255^{6.51}$A (Figure 2a). Monitoring the distance between the

207   centroid of the ligand and the center between the Cα atoms of the two leucine residues on the

208   poses obtained with IFD simulations, we observed that, while for the HM, this distance reaches

209   the 0.2 nm, for the AF2 model it is above 0.4 nm (Figure 2b).

210   To improve the conformational rearrangement around the ligand, we performed a second round

211   of IFD simulations, allowing the flexibility of the binding site side chains around compound **1**.

212   With the second round of simulations, there is a better sampling for HM conformations and an

213   enrichment of poses in close contact with $L104^{3.32}$ and $L255^{6.51}$ for the AF2 model (Figure 2b).
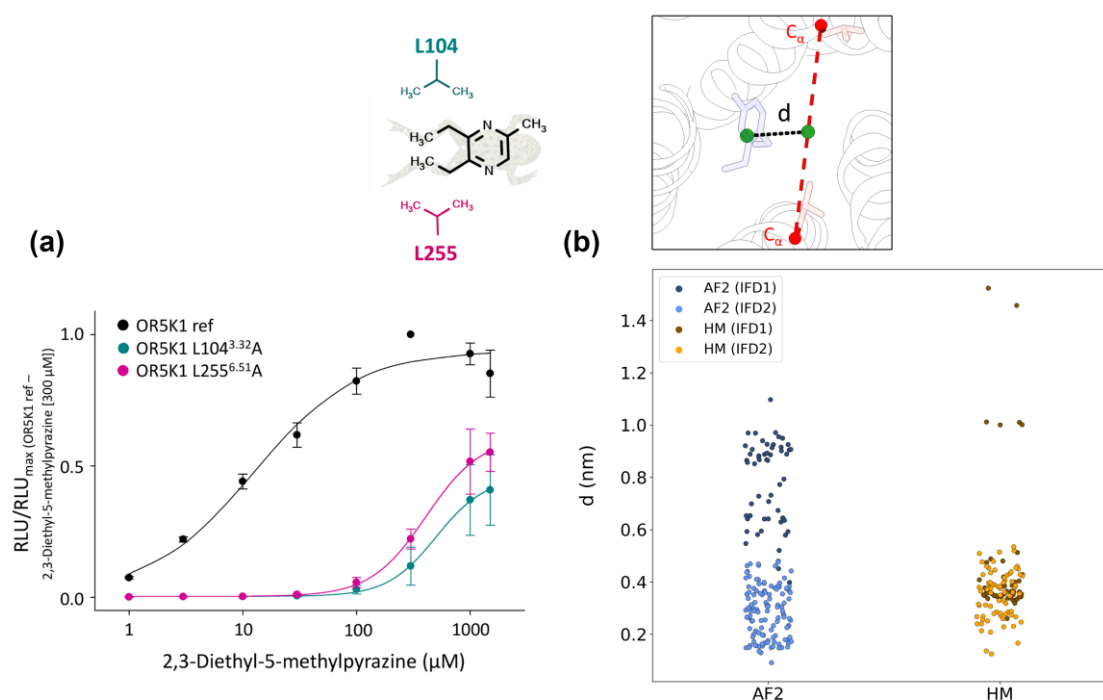

214

215 **Figure 2**: Concentration−response relations of compound **1** (2,3-diethyl-5-methylpyrazine) on OR5K1 ref (black),

216 OR5K1 L104$^{3.32}$A (turquois), and OR5K1 L255$^{6.51}$A (pink). Data were mock control-subtracted, normalized to the

217 response of OR5K1 ref to 2,3-diethyl-5-methylpyrazine (300 μM) and displayed as mean ± SD (n = 4). RLU =

218 relative luminescence units. **(b)** Distance between the ligand centroid and the center between L104$^{3.32}$ and L255$^{6.51}$

219 alpha carbons in the first and second IFD simulation rounds.

220

221 Then we analyzed all the poses where the ligand is close to L104$^{3.32}$ and L255$^{6.51}$ (with a distance

222 below 0.4 nm): 106 structures for AF2 (1 from the first round of IFD and 105 from the second

223 round) and 110 for HM (39 from the first round of IFD and 71 from the second round). We

224 clustered the complexes into 31 and 34 possible binding poses for AF2 and HM, respectively.

225 The distribution of the clusters is reported in Figure S6. Among all the potential binding modes,

226 6 models from the refinement of AF2 model and 12 structures from the refinement of HM have

227 an AUC higher than 0.8 (Table S1). These may be considered the most predictive binding site

228 conformations and were submitted to a third round of IFD simulations for the extensive

229 sampling of the conformational space of L104$^{3.32}$ and L255$^{6.51}$. This generates 555 structures

230 from the model refined from AF2 and 431 structures from the model refined from HM with

231 AUC greater than 0.8 and distance between the ligand centroid and the center between L104$^{3.32}$

232 and L255$^{6.51}$ alpha carbons lower than 0.4 nm. Despite the high similarity of generated

233 structures, we could appreciate different sampled binding modes (37 clusters from HM and 30

234 clusters from AF2, Figure S8). The best performing structures for each cluster are available at

235 https://github.com/dipizio/OR5K1_binding_site. Considering the performance, the shape of the

236 ROC curves and the contribution to the binding of L104$^{3.32}$ and L255$^{6.51}$, we selected the binding

237 poses shown in Figure 3.

238 The starting models obtained from AF2 and HM have different conformations of the TM helices

239 that prevent reaching convergence when sampling only the side chain conformations. As an

12

240   example, in Figure 3, it is possible to appreciate the difference in the shift of TM7 residues in

241   the two models: position 7.42 is F278 in the model from AF2 and T279 in the model from HM.

242
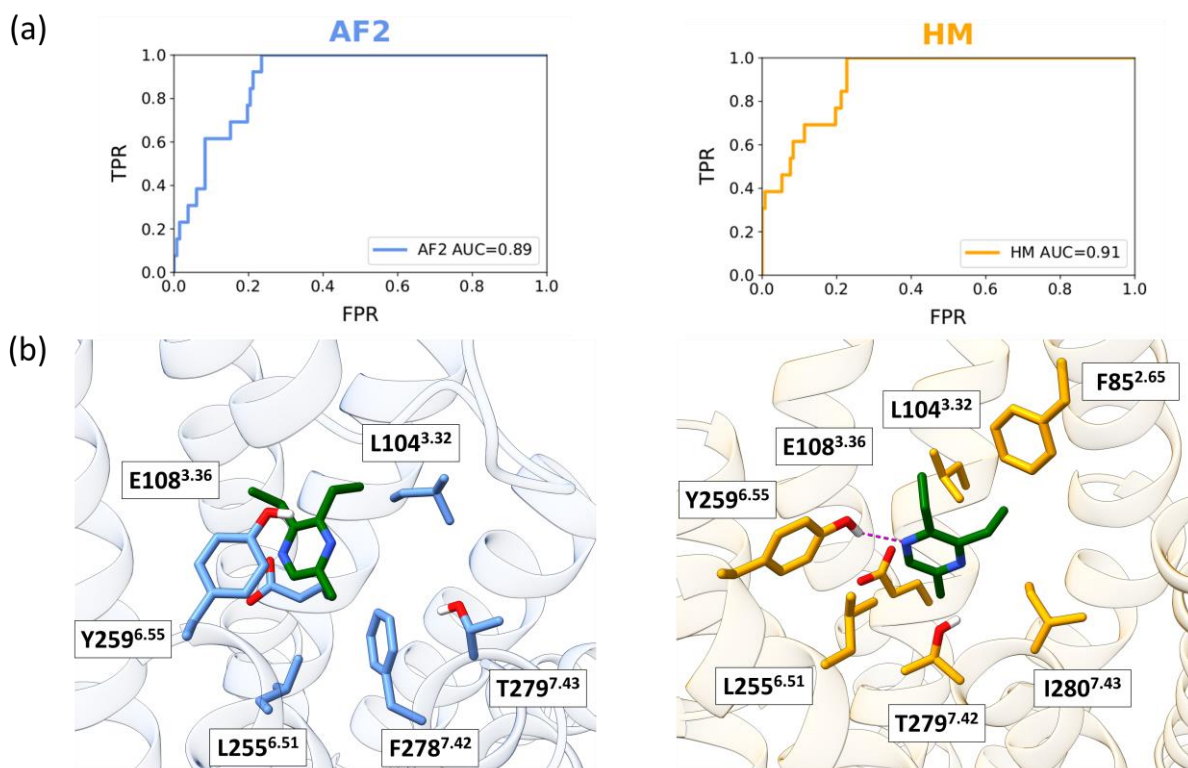
Figure 3. (a) ROC curves and **(b)** binding modes of compound **1** into the OR5K1 binding site of the best AF2 and HM models obtained after the extensive sampling of the conformational space of L104$^{3.32}$ and L255$^{6.51}$. We show as stick residues in the binding site positions that are in common between the two models. Residue F85$^{2.65}$ is only reported for the HM model, because TM2 in the AF2 model is not pointing to the binding site (the Cα atoms of F85 in the two models are 8.85 Å distant).

248

249   However, the ligand in both models is oriented in a similar position and interacts with L104$^{3.32}$

250   and L255$^{6.51}$. L104$^{3.32}$ and L255$^{6.51}$ interact with the aliphatic chains attached to the pyrazine

251   moiety and might play a relevant role on ligand selectivity. Indeed, we have shown that even

252   isomers, such as compounds **2** and **13**, elicit different receptor activation (Figure 4).

253

13
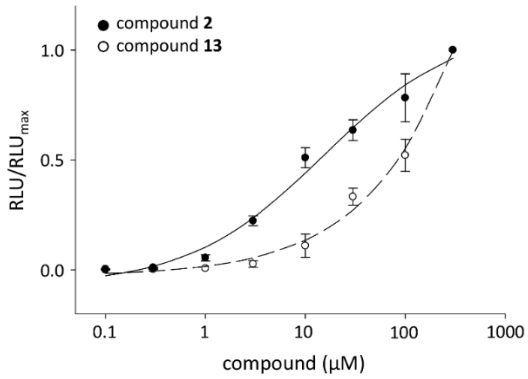
254

255 **Figure 4.** Concentration-response relations of 2-ethyl-3,6-dimethylpyrazine (compound **2**) and 2-ethyl-3,5-

256 dimethylpyrazine (compound **13**) on OR5K1. Data were mock control-subtracted, normalized to the OR5K1 signal

257 of each ligand, and displayed as mean $\pm$ SD of independent transfection experiments (n = 4). RLU = relative

258 luminescence units.

259

260 We computationally mutated L104$^{3.32}$ and L255$^{6.51}$ to alanine residues in these two models.

261 Interestingly, the docking scores correlate with the drop in activation values observed

262 experimentally and are highly influenced by the van der Waals (vdW) contribution of the

263 leucine residues (Figure 5). Also, the docking scores of compound **2** in both models are lower

264 than those of compound **13**. Therefore, both models seem to be able to capture most differences

265 in activity related to small structural differences either at the ligand or receptor side.

266

267

14
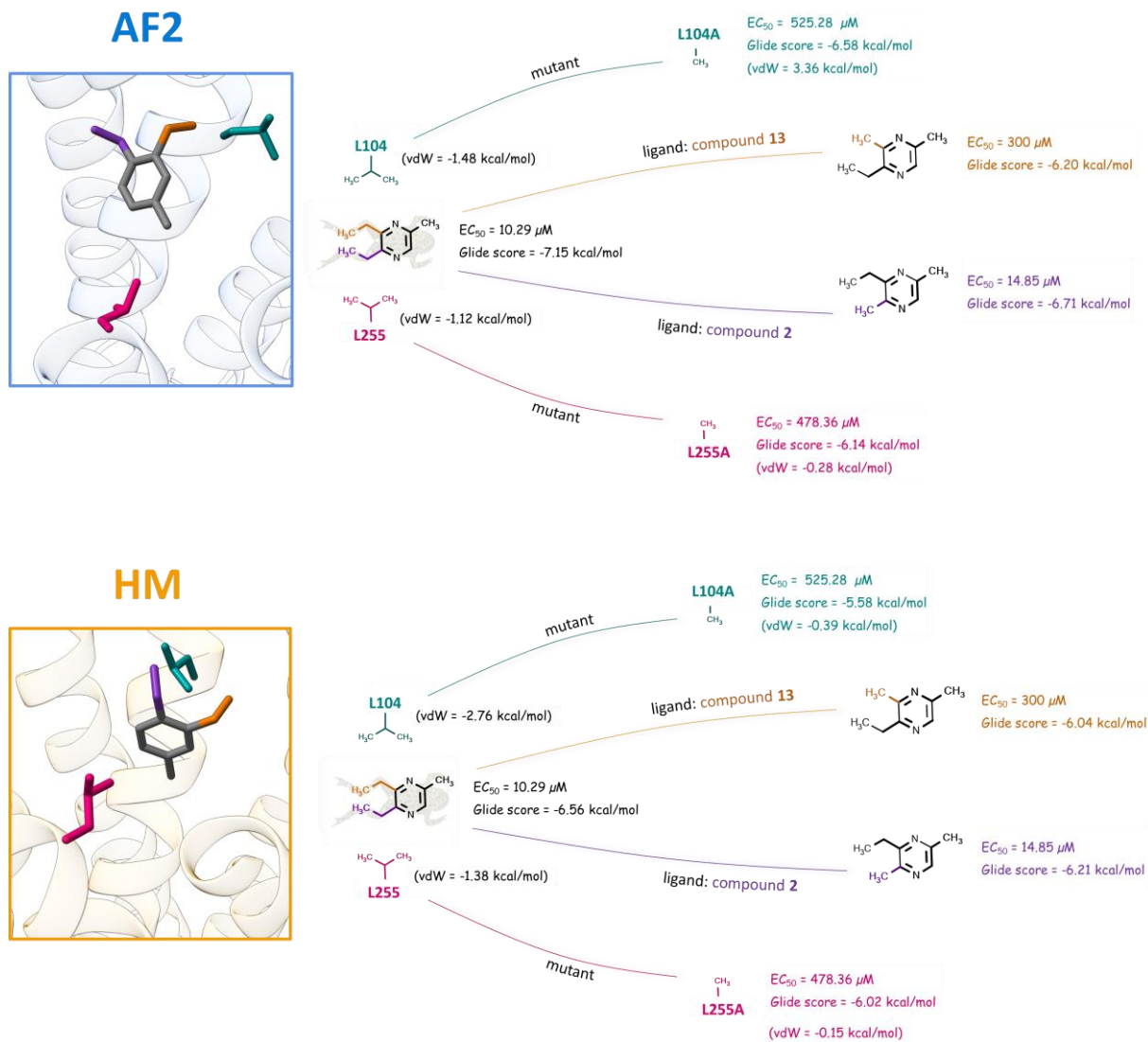
Figure 5. Schematic representation of binding mode of pyrazines **1**, **2** and **13** in the OR5K1 binding site of selected models. For compound **1**, we report also docking scores of the mutant models.

CONCLUSIONS

ORs are class A GPCRs for which we do not have experimental structures and that share a very low sequence identity with non-sensory GPCRs. The small size of OR modulators and the low resolution of the structure modeling represent a major challenge for the investigation of the molecular recognition mechanisms of this important class of receptors. Most ORs are still orphan and the receptive range of a few ORs has been characterized until now. In this paper, we used the recently published ligand information on OR5K1[50] to model and refine the OR5K1 orthosteric binding site. We used a multi-template homology modeling approach, as previously suggested to be a successful strategy for OR modeling.[20-21, 23, 62] Moreover, we further refined the ECL2 loop, which we previously identified to be a necessary procedure for low resolution GPCR modeling.[70, 83-84]

We also used the AlphaFold 2 model of OR5K1 for our analyses. This allowed us to evaluate the use of AlphaFold2 OR structures for ligand-protein interaction studies. AF2 and HM models have differences in the backbone that unavoidably affect the binding site conformations. A difference between HM and AF2 models is the activation state. The prevalence of GPCR models in the inactive state has been addressed in a recent paper by Heo et al.,[85] and the authors found that this may also affect the accuracy of binding site predictions and proposed multi-state models of GPCRs.

Altogether, we found that the refinement was a necessary step for both HM and AF2 models. The refinement process of AF2 model was needed not only to improve the performance, as for HM, but also to open the orthosteric binding site and allow docking of agonists. Through the modeling, we could identify relevant residues for the activity of OR5K1 agonists, namely, L104$^{3.32}$ and L255$^{6.51}$. These positions are highly conserved in OR5K1 orthologs across 51 species and have an extremely low frequency of SNP-based missense variations according to the 1000

16

297   Genomes Project. The support of mutagenesis experiments furnished precious experimental

298   information for model refinement.

299   In summary, we propose here an iterative experimental-computational workflow that allowed

300   us to explore the conformational space of OR5K1 binding site and can be used to model the

301   orthosteric binding site of ORs as well as that of GPCRs with low sequence identity to available

302   templates.

17

MATERIALS AND METHODS

Synthesis of 2-ethyl-3,5(6)-dimethylpyrazine. 2-ethyl-3,5(6)-dimethylpyrazines were synthesized according to Czerny et al.[86] by a Grignard-type reaction. Briefly, a solution of ethylmagnesium bromide in tetrahydrofuran (20 mL; 1.0 M; 20 mmol) was placed in a three-necked flask (100 mL) equipped with a reflux condenser, a dropping funnel and an argon inlet. While stirring at 40 °C a small portion of the respective reactant (2.2 g; 20 mmol) solved in 20 mL THF was added dropwise via the dropping funnel. 2,5-dimethylpyrazine was used for the synthesis of 2-ethyl-3,6-dimethylpyrazine and 2,6-isomere was taken as starting material for 2-ethyl-3,5-dimethylpyrazine. After the mixture was refluxed (73°C) the residual 2,5(6)-dimethylpyrazine solution was added over a period of 30 min. The mixture was stirred under refluxed for 2 h, cooled to room temperature, and water (20 mL) was added dropwise. The emulsion was extracted with diethyl ether (3 ×50 mL) and dried over anhydrous Na_2SO_4. The compounds were purified by means of flash column chromatography. For this purpose, the concentrated extract (1.0 mL) was placed on the top of a water-cooled glass column (33 × 2.5 cm) filled with a slurry of silica gel 60 (with the addition of 7 % water, 40 – 63 μm, Merck, Darmstadt, Germany, # 1.09385.2500) and n-pentane. The target compounds were eluted with n-pentane/diethyl ether (100 ml, 40:60, v/v). The purity of each target compound was analyzed by gas chromatography-mass spectrometry (GC-MS) and nuclear magnetic resonance (NMR). For determining the concentration of each 2-ethyl-3,5(6)-dimethylpyrazine, quantitative NMR (qNMR) was applied. For the NMR experiments, the solvent was distilled off and the residue was solved in $CDCl_3$.

2-ethyl-3,5-dimethylpyrazine: MS (EI): *m/z* (%) 135 (100), 136 (M[+], 81), 42 (18), 108 (17), 107 (15), 56 (12). ^{1}H-NMR ($CDCl_3$, 400 MHz, 25 °C) δ (ppm) 8.15 (s, 1 H, H-C6), 2,80 (q, *J*=7.6, 2H, H-C7), 2.53 (s, 3 H, H-C9/10), 2.49 (s, 3 H, H-C9/10), 1,27 (t, *J*=7.6, 3H, H-C8).

327 2-ethyl-3,6-dimethylpyrazine: MS (EI): *m/z* (%) 135 (100), 136 (M$^+$, 92), 56 (24), 108 (16), 42

328 (12), 107 (11). ^{1}H-NMR (400 MHz, CDCl$_3$) δ (ppm) 8.20 (s, 1 H, H-C6), 2.81 (q, *J*=7.5, 2H,

329 H-C7), 2.54 (s, 3 H, H-C9/10), 2.49 (s, 3 H, H-C9/10), 1,28 (t, *J*=7.5, 3H, H-C8).

330

331 **Nuclear magnetic resonance (NMR).** NMR experiments were performed using an Avance III

332 400 MHz spectrometer equipped with a BBI probe (Bruker, Rheinstetten, Germany). Topspin

333 software (version 3.2) was used for data acquisition. For structure elucidation the compounds

334 were solved in chloroform-d (CDCl$_3$). Chemical shifts were referenced against solvent signal.

335 Quantitative ^{1}H-NMR (qNMR) was done according to Frank et al.[87] For this, an aliquot (600

336 µL) of the dissolved solutions was analyzed in NMR tubes (5 × 178 mm, Bruker, Faellanden,

337 Switzerland).

338

339 **Gas chromatography – mass spectrometry (GC-MS).** Mass spectra of the synthesized

340 pyrazines in the electron ionization mode were recorded using a GC-MS system consisting of

341 a Trace GC Ultra gas chromatograph coupled to a single quadrupole ISQ mass spectrometer

342 (Thermo Fisher Scientific, Dreieich, Germany) as described more detailed by Porcelli et al.[88]

343 A DB-1701 coated fused silica capillary column (30 m × 0.25 mm i.d., 0.25 µm film thickness;

344 Agilent, Waldbronn, Germany) was taken for chromatographic separation using the following

345 temperature program: 40°C held for 2 min, then it was raised at 10 °C/min to 230°C (held for

346 4 min). Mass spectra were acquired at a scan range of 40–300 m/z at an ionization energy of 70

347 eV. The mass spectra were evaluated using Xcalibur 2.0 software (Thermo Fisher Scientific).

348

349 **Molecular cloning of OR5K1.** The protein-coding region of human OR5K1

350 (NM_001004736.3) derived from our previously published OR library.[89] Amplification was

19

351   carried out in a touchdown approach using gene-specific primers (Table S2): an initial

352   denaturation (98 °C, 3 min) and ten cycles consisting of denaturation (98 °C, 30 s), annealing

353   (60 °C, decreasing 1 °C per cycle down to 50 °C, 30 s), and extension (72 °C, 1 min), followed

354   by 25 cycles of denaturation (98 °C, 30 s), annealing (50 °C, 30 s), and extension (72 °C, 1

355   min), finishing with a final extension step in the end (72 °C, 7 min). Insertion of nucleotides

356   into expression vectors was done with T4-DNA ligase (#M1804, Promega, Madison, USA) via

357   EcoRI/NotI (#R6017/#R6435, Promega, Madison, USA) into the expression plasmid

358   pFN210A,[90] and verified by Sanger sequencing using internal primers (Table S3) (Eurofins

359   Genomics, Ebersberg, Germany).

360

361   **PCR-based site-directed mutagenesis.** Mutants L104$^{3.32}$ and L255$^{6.51}$ were generated by PCR-

362   based site-directed mutagenesis in two steps. Utilized mutation primers were designed

363   overlapping and are listed in Table S4. Step one PCR was performed in two amplifications, one

364   with the forward vector-internal primer and the reverse mutation-primer, the other with the

365   forward mutation-primer and the reverse vector-internal primer. Amplification was performed

366   with the touchdown approach described above. Both PCR amplicons were then purified and

367   used as template for step two. The two overlapping amplicons were annealed using the

368   following touchdown program: denaturation (98 °C, 3 min), ten cycles containing denaturation

369   (98 °C, 30 s), annealing (start 60 °C, 30 s), and extension (72 °C, 2 min). After this, vector-

370   internal forward and reverse primers were added and 25 further cycles of denaturation (98 °C,

371   30 s), annealing (50 °C, 30 s), and extension (72 °C, 1 min) were carried out, finishing with a

372   final extension step in the end (72 °C, 7 min). The amplicons were then sub-cloned as described

373   above.

374

20

375   **Cell culture and transient DNA transfection.** We utilized HEK-293 cells,[91] a human

376   embryonic kidney cell-line, as a test cell system for the functional expression of ORs.[92] Cells

377   were cultivated at 37 °C, 5% CO_2, and 100% humidity in 4.5 g/L D-glucose containing DMEM

378   with 10% fetal bovine serum, 2 mM L-glutamine, 100 U/mL penicillin, and 100 U/mL

379   streptomycin. Cells were cultured in a 96-well format (Nunclon™ Delta Surface, #136102;

380   Thermo Fisher Scientific, Schwerte, Germany) at 12,000 cells/well overnight. Then, cells were

381   transfected utilizing 0.75 µL/well ViaFect™ (#E4981, Promega, USA) with the following

382   constructs: 100 ng/well of the respective OR construct, 50 ng/well of chaperone RTP1S,[93] 50

383   ng/well of the G protein subunit $G\alpha_{olf}$,[94-95] olfactory G protein subunit Gγ13,[96] and 50 ng/well

384   of pGloSensor™-22F (Promega, Madison, USA).[97] The utilized pGloSensor™-22F is a

385   genetically engineered luciferase with a cAMP-binding pocket, allowing for measurements of

386   a direct cAMP-dependent luminescence signal. All measurements were mock-controlled, i.e.

387   pFN210A without OR was transfected in parallel.

388

389   **Luminescence assay.** Concentration-response assays were measured 42 hours post-

390   transfection as described previously.[92] In short, supernatant was removed and cells were loaded

391   with a physiological salt buffer (pH 7.5) containing 140 mmol/L NaCl, 10 mmol/L HEPES, 5

392   mmol/L KCl, 1 mmol/L CaCl2, 10 mmol/L glucose, and 2% of beetle luciferin sodium salt

393   (Promega, Madison, USA). For luminescence measurements, the GloMax® Discover

394   microplate reader (Promega, Madison, USA) was used. After an incubation for 50 minutes in

395   the dark, the basal luminescence signal of each well was recorded thrice. Then the odorant,

396   serially diluted in the physiological salt buffer with added Pluronic PE-10500 (BASF,

397   Ludwigshafen, Germany), was applied to the cells and luminescence was measured thrice after

398   ten minutes of incubation time. The final Pluronic PE-10500 concentration on the cells was

399   0.05%.

21

400   **Data analysis of the cAMP-luminescence measurements.** The raw luminescence data

401   obtained from the GloMax® Discover microplate reader detection system were analyzed for

402   concentration/response assays by averaging both data points of basal levels and data points after

403   odorant application. For a given luminescence signal, the respective basal level was subtracted

404   and the now corrected data set was normalized to the maximum amplitude of the reference. The

405   data set for the mock control was subtracted and EC_{50} values and curves were derived from

406   fitting the function:

407
$$f(x) = \left(\frac{(min - \max)}{(1 + (\frac{x}{EC_{50}})^{Hillslope})} \right) + max$$

408   to the data by nonlinear regression (SigmaPlot 14.0, Systat Software).[98] Data are presented as

409   mean ± SD.

410

411   **Phylogenetic analysis.** NCBI[99] was used as database for the retrieval of genetic information on

412   *Homo sapiens* (human) odorant receptor genes as well as orthologous receptor genes of OR5K1

413   (for accession numbers see Table S5). The phylogenetic reconstruction of ORs was performed

414   with QIAGEN CLC Genomics Workbench 21.0 (https://digitalinsights.qiagen.com/) and

415   MEGA X software.[100] Therefore, in a first step, all sequences were aligned using ClustalW

416   algorithm.[101] The evolutionary history was inferred using the Neighbor-Joining method [102]

417   followed by 500 bootstrap replications.[103] Scale bar refers to the evolutionary distances,

418   computed using the Poisson correction method.[104] Evolutionary analyses were conducted in

419   MEGA X.[100] For rooting the constructed tree, human rhodopsin (NCBI entry: NP_000530.1)

420   was used as an out-group.

421

22

Homology Modeling. Rhodopsin receptor (PDB ID: 4X1H), β2- adrenergic receptor (PDB ID: 6MXT), CXCR4 receptor (PDB ID: 3ODU), and A2A receptor (PDB ID: 2YDV) were used as templates for modeling the 3D structure of OR5K1, following the template selection from de March et al. 2015.[20] The structures were downloaded from GPCRdb,[105] and their sequences were aligned to the OR5K1 sequence (residues 20-292) with the Protein Structure Alignment module available in Maestro (Schrödinger Release 2021-3, Maestro, Schrödinger, LLC, New York, NY, 2021). The sequence alignment was then manually adjusted, ensuring that conserved GPCR residues were correctly aligned (Figure S1). OR5K1 shares a sequence identity of 19% with 6MXT.pdb, of 15% with 4X1H.pdb, of 15% with 3ODU.pdb and of 16% with 2YDV.pdb. We modeled the ECL2 region ($S157^{4.57}$- $L188^{5.37}$) using as templates NPY2 (PDB ID: 7DDZ) and CCK1 (PDB ID: 7MBY) for the before-$Cys^{45.50}$ segment, and apelin (PDB ID: 6KNM) for the after-$Cys^{45.50}$ segment (Figures S2 and S3). We also remodeled the region between $P81^{2.58}$ and $L105^{3.32}$ with the NPY2 to ensure the correct orientation of the ECL2 towards TM3 and ECL1, and the formation of the conserved disulfide bridge between $C^{3.25}$ and $C^{45.50}$. 100 homology models were generated using MODELLER v9.23.[106] Four models were selected based on the DOPE score and visual inspection of the ECL2 and the most predictive model, based on ROC AUC (see the paragraph Molecular Docking) was chosen for the following analysis.

Protein preparation and binding site analysis. OR5K1 AF2 model was downloaded from the AlphaFold 2 database (https://alphafold.ebi.ac.uk/entry/Q8NHB7). OR5K1 AF2 and HM were superimposed through the Protein Structure Alignment module available in Maestro (Schrödinger Release 2021-3, Maestro, Schrödinger, LLC, New York, NY, 2021). Hydrogen atoms and side chains of both models were optimized with the Protein Preparation Wizard tool at physiological pH (Schrödinger Release 2021-3, Maestro, Schrödinger, LLC, New York, NY, 2021). Ramachandran plots were generated to verify the reliability of the backbone dihedral

23

448   angles of amino acid residues in the models. The A100 tool was used to investigate the

449   activation state of the models.[69]

450   SiteMap tool (Schrödinger Release 2021-3: SiteMap, Schrödinger, LLC, New York, NY, 2021)

451   was used to characterize the binding cavities of both models.

452

453   **Molecular Docking.** The compounds used in the screening by Marcinek et al. were used for

454   the model evaluation.[50] However, we excluded from this set 54 molecules employed as a

455   mixture of isomers. Indeed, the measured activity of the mixture may not correspond to the

456   activity of the individual stereoisomers (e.g., only one stereoisomer is active) and compromise

457   our validation. Among the subset of molecules with defined stereochemistry, we selected 11

458   agonists with EC_{50} values below 600 μM and compounds characterized in this work were

459   included in the list of active molecules (Table 1). 131 compounds that did not elicit receptor

460   response   were   used   as   inactives   (the   list   of   compounds   is   available   at

461   https://github.com/dipizio/OR5K1_binding_site).

462   3D structures of ligands and inactive molecules were retrieved from PubChem through CAS

463   numbers and prepared for docking through the generation of stereoisomers and protonation

464   states at pH 7.2 ± 0.2 with LigPrep, as implemented in the Schrödinger Small-Molecule Drug

465   Discovery Suite 2021 (LigPrep, Schrödinger, LLC, New York, NY, 2021). Glide Standard

466   Precision (Glide, Schrödinger, LLC, New York, NY, 2021)[107-108] was used for docking all

467   compounds to the OR5K1 models. The grid box was centroid of SiteMap grid points for HM

468   and AF2 binding pockets combined together for the models obtained after the first round of

469   IFD, and instead was centroid of the docked 2,3-diethyl-5-methylpyrazine (compound **1**) for

470   the models obtained after the second round of IFD simulations.

24

471   The docking poses of compound **1** within OR5K1 mutants were performed using the *in-place*

472   docking (Glide Standard precision), generating the grid from the centroid of the docked

473   compound. Mutants were generated with the 'Mutate residue' tool available in Maestro.

474   An in-house python script based on Scikit-learn (v0.24.2) package was used for the ROC curve

475   analysis,[109] and the data were plotted with Matplotlib Python library.[110] AUC and $EF_{15\%}$ of the

476   training library were used to evaluate the performance of each model in discriminating between

477   active and inactive compounds.

478   The ROC curves were obtained plotting False Positive Rate (FPR) vs. True Positive Rate (TPR).

479   TPR and FPR values are calculated by the following equations:

480
$$TPR = \frac{TP}{(TP + FN)}$$

481   where TP is the number of true positive compounds, and FN is the number of false negative

482   compounds.

483
$$FPR = \frac{FP}{(TN + FP)}$$

484   where FP is the number of false positive compounds, and TN is the number of true negative

485   compounds.

486    $EF_{15\%}$ values are calculated by the following equation:

487
$$EF_{15\%} = \frac{N_{actives\ (15\%)}}{N_{inactives\ (15\%)}}$$

488   where $N_{actives\ (15\%)}$ and $N_{inactives\ (15\%)}$ represent the number of actives and inactives, respectively,

489   in the 15% of ranked screened compounds.

490

491   **Induced-fit docking simulations.** In the first round of simulations, HM and AF2 starting

492   models were used for IFD simulations using Schrödinger Suite 2021 Induced Fit Docking

493 protocol (Glide, Schrödinger, LLC, New York, NY, 2021; Prime, Schrödinger, LLC, New

494 York, NY, 2021).[111] 2,3-diethyl-5-methylpyrazine was used as ligand and the flexibility of the

495 side chains at 3 Å from the SiteMap grid points was allowed. The best structures based on AUC

496 values and visual inspection from IFD1 (4 structures after refinement of HM and 7 after

497 refinement of AF2 model) underwent to a second round of simulations (IFD2). In the second

498 round of simulations, the residues at 4 Å from the ligand (2,3-diethyl-5-methylpyrazine) were

499 allowed to move. The most predictive structures from IFD2 (Table S1) were submitted to a

500 third round of IFD simulations (IFD 3), in which only the side chains of L104$^{3.32}$ and L255$^{6.51}$

501 and the ligand were treated as flexible. For an extensive sampling of the leucine residues, we

502 used as ligand both compound **1** and **2**.

503

504 **Clustering of docking poses.** For all poses from IFD1, IFD2, and IDF3 we monitored the

505 distance between the ligand centroid and the center between L104$^{3.32}$ and L255$^{6.51}$ alpha

506 carbons. The centroids and distances were calculated using PLUMED (version 2.7).[112-114] The

507 docking poses from IDF1 and IDF2 with a distance below 0.4 nm were clustered using the

508 conformer_cluster.py from Schrödinger (https://www.schrodinger.com/scriptcenter). First, a

509 pair-wise RMSD matrix was calculated for compound **1** and the residues within 7 Å of its

510 centroid (for HM, residues 104, 105, 108, 159, 199, 202, 206, 255, 256, 276, 279, 280; for AF2,

511 residues: 101, 104, 105, 108, 178, 180, 181, 199, 255, 258, 259, 275, 278, 279), and then the

512 complexes were clustered using the hierarchical cluster method (average group linkage). The

513 number of clusters was set to 31 for AF2 and 34 for HM based on the second minimum of the

514 Kelly-Penalty score. Docking poses obtained from IDF3 were filtered by distance (below 0.4

515 nm), AUC (greater than 0.8) and the conformations of the binding site were clustered using the

516 conformer_cluster.py from Schrödinger. RMSD matrices of best performing structures from

517 the different clusters were calculated with rmsd.py from Schrödinger (Figure S8).

518   ChimeraX (v1.3) was used to render the protein images.[115]

519

DATA AND SOFTWARE AVAILABILITY

521   The   dataset   and   refined   models   can   be   downloaded   from

522   https://github.com/dipizio/OR5K1_binding_site.

523

524   **Supporting Information**. Multiple Sequence Alignment of OR5K1 with templates (Figure
525   S1); ECL2 for OR5K1 and experimental class A GPCRs (Figure S2); OR5K1 models built with
526   AlphaFold 2 and homology modeling (Figure S3); ROC analysis of the starting OR5K1 AF2
527   and HM models (Figure S4); Binding modes and ROC analyses of the OR5K1 AF2 and HM
528   models after the first IFD simulation round (Figure S5); Distribution of the clusters binding
529   poses of compound 1 in proximity to L104$^{3.32}$ and L255$^{6.51}$ (Figure S6); Leucine residues
530   L104$^{3.32}$ and L255$^{6.51}$ are highly conserved in OR5K1 homologs (Figure S7); RMSD matrices
531   for the orthosteric binding site of the best performing models obtained after clustering IFD3
532   models (Figure S8); Models from IFD1 and IFD2 with $d < 0.4$, $AUC > 0.8$ (Table S1);
533   Oligonucleotides for molecular cloning of OR5K1 (Table S2); Vector internal oligonucleotides
534   (Table S3); Oligonucleotides for Homo sapiens OR5K1 site directed mutagenesis (Table S4);
535   NCBI reference sequences of olfactory receptor genes investigated (Table S5).

536 **AUTHOR INFORMATION**

537 Corresponding Authors

538 Antonella Di Pizio – Molecular Modeling Group, Leibniz Institute for Food Systems Biology

539 at the Technical University of Munich, 85354 Freising, Germany; ORCID:

540 https://orcid.org/0000-0002-8520-5165 ; Email: a.dipizio.leibniz-lsb@tum.de

541 Dietmar Krautwurst - Taste and Odor Systems Reception Group, Leibniz Institute for Food

542 Systems Biology at the Technical University of Munich, 85354 Freising, Germany; Email:

543 d.krautwurst.leibniz-lsb@tum.de

544

545 Authors

546 Alessandro Nicoli - Molecular Modeling Group, Leibniz Institute for Food Systems Biology at

547 the Technical University of Munich, 85354 Freising, Germany, ORCID: https://orcid.org/0000-

548 0001-6177-9749

549 Franziska Haag - Taste and Odor Systems Reception Group, Leibniz Institute for Food Systems

550 Biology at the Technical University of Munich, 85354 Freising, Germany, ORCID:

551 https://orcid.org/0000-0003-4210-0475

552 Patrick Marcinek - Taste and Odor Systems Reception Group, Leibniz Institute for Food

553 Systems Biology at the Technical University of Munich, 85354 Freising, Germany

554 Ruiming He - Molecular Modeling Group, Leibniz Institute for Food Systems Biology at the

555 Technical University of Munich, 85354 Freising, Germany; Department of Chemistry,

556 Technical University of Munich, D-85748 Garching

557   Johanna Kreißl – Analytical Technologies, Leibniz Institute for Food Systems Biology at the

558   Technical University of Munich, 85354 Freising, Germany; ORCID: https://orcid.org/0000-

559   0002-5252-0409

560   Jörg Stein - Food Metabolome Chemistry Group, Leibniz Institute for Food Systems Biology

561   at the Technical University of Munich, 85354 Freising, Germany

562   Alessandro Marchetto - Computational Biomedicine group, Institute for Advanced Simulations

563   (IAS)-5/Institute for Neuroscience and Medicine (INM)-9, Forschungszentrum Jülich, 52428

564   Jülich, Germany; Faculty of Mathematics, Computer Science and Natural Sciences, RWTH

565   Aachen University, 52062 Aachen, Germany, ORCID: https://orcid.org/0000-0001-9857-4376

566   Andreas Dunkel - Integrative Food Systems Analysis Group, Leibniz Institute for Food Systems

567   Biology at the Technical University of Munich, 85354 Freising, Germany

568   Thomas Hofmann - Chair of Food Chemistry and Molecular Sensory Science, Technical

569   University of Munich, 85354 Freising, Germany

570

571   Author Contributions

572   The manuscript was written through the contributions of all authors. All authors have approved

573   the final version of the manuscript.

574   ‡These authors contributed equally.

575

576

577

REFERENCES

(1) Hall, R. A.; Premont, R. T.; Lefkowitz, R. J., Heptahelical receptor signaling: beyond the G protein paradigm. *J Cell Biol* **1999,** *145* (5), 927-32.

(2) Pierce, K. L.; Premont, R. T.; Lefkowitz, R. J., Seven-transmembrane receptors. *Nat Rev Mol Cell Biol* **2002,** *3* (9), 639-50.

(3) Weis, W. I.; Kobilka, B. K., The Molecular Basis of G Protein-Coupled Receptor Activation. *Annu Rev Biochem* **2018,** *87*, 897-919.

(4) Fredriksson, R.; Lagerstrom, M. C.; Lundin, L. G.; Schioth, H. B., The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **2003,** *63* (6), 1256-72.

(5) Davies, M. N.; Secker, A.; Halling-Brown, M.; Moss, D. S.; Freitas, A. A.; Timmis, J.; Clark, E.; Flower, D. R., GPCRTree: online hierarchical classification of GPCR function. *BMC Res Notes* **2008,** *1*, 67.

(6) Nordstrom, K. J.; Sallman Almen, M.; Edstam, M. M.; Fredriksson, R.; Schioth, H. B., Independent HHsearch, Needleman--Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families. *Mol Biol Evol* **2011,** *28* (9), 2471-80.

(7) Chan, H. C. S.; Li, Y.; Dahoun, T.; Vogel, H.; Yuan, S., New Binding Sites, New Opportunities for GPCR Drug Discovery. *Trends Biochem Sci* **2019,** *44* (4), 312-330.

(8) Congreve, M.; de Graaf, C.; Swain, N. A.; Tate, C. G., Impact of GPCR Structures on Drug Discovery. *Cell* **2020,** *181* (1), 81-91.

(9) Sanchez-Reyes, O. B.; Cooke, A. L. G.; Tranter, D. B.; Rashid, D.; Eilers, M.; Reeves, P. J.; Smith, S. O., G Protein-Coupled Receptors Contain Two Conserved Packing Clusters. *Biophys J* **2017,** *112* (11), 2315-2326.

(10) Hilger, D.; Masureel, M.; Kobilka, B. K., Structure and dynamics of GPCR signaling complexes. *Nat Struct Mol Biol* **2018,** *25* (1), 4-12.

(11) Niimura, Y., Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. *Hum Genomics* **2009,** *4* (2), 107-18.

(12) Olender, T.; Waszak, S. M.; Viavant, M.; Khen, M.; Ben-Asher, E.; Reyes, A.; Nativ, N.; Wysocki, C. J.; Ge, D.; Lancet, D., Personal receptor repertoires: olfaction as a model. *BMC Genomics* **2012,** *13*, 414.

(13) Buck, L.; Axel, R., A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **1991,** *65* (1), 175-87.

(14) Smith, S. O., Deconstructing the transmembrane core of class A G protein-coupled receptors. *Trends Biochem Sci* **2021,** *46* (12), 1017-1029.

(15) Malnic, B.; Godfrey, P. A.; Buck, L. B., The human olfactory receptor gene family. *Proc Natl Acad Sci U S A* **2004,** *101* (8), 2584-9.

(16) Glusman, G.; Bahar, A.; Sharon, D.; Pilpel, Y.; White, J.; Lancet, D., The olfactory receptor gene superfamily: data mining, classification, and nomenclature. *Mamm Genome* **2000,** *11* (11), 1016-23.

(17) Zhang, X.; Firestein, S., The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* **2002,** *5* (2), 124-33.

(18) Niimura, Y.; Nei, M., Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci U S A* **2005,** *102* (17), 6039-44.

(19) Kotthoff, M.; Bauer, J.; Haag, F.; Krautwurst, D., Conserved C-terminal motifs in odorant receptors instruct their cell surface expression and cAMP signaling. *FASEB J* **2021,** *35* (2), e21274.

(20) de March, C. A.; Kim, S. K.; Antonczak, S.; Goddard, W. A., 3rd; Golebiowski, J., G protein-coupled odorant receptors: From sequence to structure. *Protein Sci* **2015,** *24* (9), 1543-8.

624 (21) de March, C. A.; Yu, Y.; Ni, M. J.; Adipietro, K. A.; Matsunami, H.; Ma, M.; Golebiowski, J.,
625 Conserved Residues Control Activation of Mammalian G Protein-Coupled Odorant Receptors. *J Am*
626 *Chem Soc* **2015,** *137* (26), 8611-8616.

627 (22) Man, O.; Gilad, Y.; Lancet, D., Prediction of the odorant binding site of olfactory receptor proteins
628 by human-mouse comparisons. *Protein Sci* **2004,** *13* (1), 240-54.

629 (23) Bushdid, C.; de March, C. A.; Topin, J.; Do, M.; Matsunami, H.; Golebiowski, J., Mammalian
630 class I odorant receptors exhibit a conserved vestibular-binding pocket. *Cell Mol Life Sci* **2019,** *76* (5),
631 995-1004.

632 (24) Gelis, L.; Wolf, S.; Hatt, H.; Neuhaus, E. M.; Gerwert, K., Prediction of a ligand-binding niche
633 within a human olfactory receptor by combining site-directed mutagenesis with dynamic homology
634 modeling. *Angew Chem Int Ed Engl* **2012,** *51* (5), 1274-8.

635 (25) Cong, X.; Ren, W.; Pacalon, J.; Xu, R.; Xu, L.; Li, X.; de March, C. A.; Matsunami, H.; Yu, H.;
636 Yu, Y.; Golebiowski, J., Large-Scale G Protein-Coupled Olfactory Receptor–Ligand Pairing. *ACS*
637 *Central Science* **2022,** *8* (3), 379-387.

638 (26) Chen, H.; Dadsetan, S.; Fomina, A. F.; Gong, Q., Expressing exogenous functional odorant
639 receptors in cultured olfactory sensory neurons. *Neural development* **2008,** *3*, 22.

640 (27) Geithe, C.; Noe, F.; Kreissl, J.; Krautwurst, D., The Broadly Tuned Odorant Receptor OR1A1 is
641 Highly Selective for 3-Methyl-2,4-nonanedione, a Key Food Odorant in Aged Wines, Tea, and Other
642 Foods. *Chemical senses* **2017,** *42* (3), 181-193.

643 (28) Kajiya, K.; Inaki, K.; Tanaka, M.; Haga, T.; Kataoka, H.; Touhara, K., Molecular bases of odor
644 discrimination: Reconstitution of olfactory receptors that recognize overlapping sets of odorants. *J*
645 *Neurosci* **2001,** *21* (16), 6018-25.

646 (29) Malnic, B.; Hirono, J.; Sato, T.; Buck, L. B., Combinatorial receptor codes for odors. *Cell* **1999,**
647 *96* (5), 713-23.

648 (30) Nara, K.; Saraiva, L. R.; Ye, X. L.; Buck, L. B., A Large-Scale Analysis of Odor Coding in the
649 Olfactory Epithelium. *Journal of Neuroscience* **2011,** *31* (25), 9179-9191.

650 (31) Saito, H.; Chi, Q.; Zhuang, H.; Matsunami, H.; Mainland, J. D., Odor coding by a Mammalian
651 receptor repertoire. *Sci Signal* **2009,** *2* (60), ra9.

652 (32) Adipietro, K. A.; Mainland, J. D.; Matsunami, H., Functional evolution of mammalian odorant
653 receptors. *PLoS Genet* **2012,** *8* (7), e1002821.

654 (33) Mainland, J. D.; Li, Y. R.; Zhou, T.; Liu, W. L.; Matsunami, H., Human olfactory receptor
655 responses to odorants. *Scientific data* **2015,** *2*, 150002.

656 (34) Noe, F.; Polster, J.; Geithe, C.; Kotthoff, M.; Schieberle, P.; Krautwurst, D., OR2M3: A Highly
657 Specific and Narrowly Tuned Human Odorant Receptor for the Sensitive Detection of Onion Key Food
658 Odorant 3-Mercapto-2-methylpentan-1-ol. *Chemical senses* **2017,** *42* (3), 195-210.

659 (35) Haag, F.; Di Pizio, A.; Krautwurst, D., The key food odorant receptive range of broadly tuned
660 receptor OR2W1. *Food Chem* **2022,** *375*, 131680.

661 (36) Charlier, L.; Topin, J.; Ronin, C.; Kim, S. K.; Goddard, W. A., 3rd; Efremov, R.; Golebiowski, J.,
662 How broadly tuned olfactory receptors equally recognize their agonists. Human OR1G1 as a test case.
663 *Cell Mol Life Sci* **2012,** *69* (24), 4205-13.

664 (37) Jabeen, A.; de March, C. A.; Matsunami, H.; Ranganathan, S., Machine Learning Assisted
665 Approach for Finding Novel High Activity Agonists of Human Ectopic Olfactory Receptors. *Int J Mol*
666 *Sci* **2021,** *22* (21).

667 (38) Ballante, F.; Kooistra, A. J.; Kampen, S.; de Graaf, C.; Carlsson, J., Structure-Based Virtual
668 Screening for Ligands of G Protein-Coupled Receptors: What Can Molecular Docking Do for You?
669 *Pharmacol Rev* **2021,** *73* (4), 527-565.

670   (39) Block, E., Molecular Basis of Mammalian Odor Discrimination: A Status Report. *J Agric Food Chem* **2018,** *66* (51), 13346-13366.

672   (40) Bushdid, C.; de March, C. A.; Fiorucci, S.; Matsunami, H.; Golebiowski, J., Agonists of G-Protein-Coupled Odorant Receptors Are Predicted from Chemical Features. *J Phys Chem Lett* **2018,** *9* (9), 2235-2240.

675   (41) Yuan, S.; Dahoun, T.; Brugarolas, M.; Pick, H.; Filipek, S.; Vogel, H., Computational modeling of the olfactory receptor Olfr73 suggests a molecular basis for low potency of olfactory receptor-activating compounds. *Commun Biol* **2019,** *2*, 141.

678   (42) Haag, F.; Ahmed, L.; Reiss, K.; Block, E.; Batista, V. S.; Krautwurst, D., Copper-mediated thiol potentiation and mutagenesis-guided modeling suggest a highly conserved copper-binding motif in human OR2M3. *Cell Mol Life Sci* **2019**.

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

684   (44) Hameduh, T.; Haddad, Y.; Adam, V.; Heger, Z., Homology modeling in the time of collective and artificial intelligence. *Comput Struct Biotechnol J* **2020,** *18*, 3494-3506.

686   (45) Akdel, M.; Pires, D. E. V.; Porta Pardo, E.; Jänes, J.; Zalevsky, A. O.; Mészáros, B.; Bryant, P.; Good, L. L.; Laskowski, R. A.; Pozzati, G.; Shenoy, A.; Zhu, W.; Kundrotas, P.; Ruiz Serra, V.; Rodrigues, C. H. M.; Dunham, A. S.; Burke, D.; Borkakoti, N.; Velankar, S.; Frost, A.; Lindorff-Larsen, K.; Valencia, A.; Ovchinnikov, S.; Durairaj, J.; Ascher, D. B.; Thornton, J. M.; Davey, N. E.; Stein, A.; Elofsson, A.; Croll, T. I.; Beltrao, P., **2021**.

691   (46) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millan, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D., Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021,** *373* (6557), 871-876.

697   (47) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596* (7873), 583-589.

703   (48) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Zidek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S., AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **2022,** *50* (D1), D439-D444.

709   (49) Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Zidek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; Velankar, S.; Kleywegt, G. J.; Bateman, A.; Evans, R.; Pritzel, A.; Figurnov, M.; Ronneberger, O.; Bates, R.; Kohl, S. A. A.; Potapenko, A.; Ballard, A. J.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Clancy, E.; Reiman, D.; Petersen, S.; Senior, A. W.; Kavukcuoglu, K.; Birney, E.; Kohli, P.; Jumper, J.; Hassabis, D., Highly accurate protein structure prediction for the human proteome. *Nature* **2021,** *596* (7873), 590-596.

715   (50) Marcinek, P.; Haag, F.; Geithe, C.; Krautwurst, D., An evolutionary conserved olfactory receptor for foodborne and semiochemical alkylpyrazines. *FASEB J* **2021,** *35* (6), e21638.

717 (51) Migita, K.; Iiduka, T.; Tsukamoto, K.; Sugiura, S.; Tanaka, G.; Sakamaki, G.; Yamamoto, Y.;
718 Takeshige, Y.; Miyazawa, T.; Kojima, A.; Nakatake, T.; Okitani, A.; Matsuishi, M., Retort beef aroma
719 that gives preferable properties to canned beef products and its aroma components. *Anim Sci J* **2017,** *88*
720 (12), 2050-2056.
721 (52) Hou, L.; Zhang, Y.; Wang, X., Characterization of the Volatile Compounds and Taste Attributes
722 of Sesame Pastes Processed at Different Temperatures. *J Oleo Sci* **2019,** *68* (6), 551-558.
723 (53) Henning, C.; Glomb, M. A., Pathways of the Maillard reaction under physiological conditions.
724 *Glycoconj J* **2016,** *33* (4), 499-512.
725 (54) Bohman, B.; Phillips, R. D.; Menz, M. H.; Berntsson, B. W.; Flematti, G. R.; Barrow, R. A.;
726 Dixon, K. W.; Peakall, R., Discovery of pyrazines as pollinator sex pheromones and orchid
727 semiochemicals: implications for the evolution of sexual deception. *New Phytol* **2014,** *203* (3), 939-52.
728 (55) Silva-Junior, E. A.; Ruzzini, A. C.; Paludo, C. R.; Nascimento, F. S.; Currie, C. R.; Clardy, J.;
729 Pupo, M. T., Pyrazines from bacteria and ants: convergent chemistry within an ecological niche. *Sci Rep*
730 **2018,** *8* (1), 2595.
731 (56) Osada, K.; Kurihara, K.; Izumi, H.; Kashiwayanagi, M., Pyrazine analogues are active components
732 of wolf urine that induce avoidance and freezing behaviours in mice. *PLoS One* **2013,** *8* (4), e61753.
733 (57) Osada, K.; Miyazono, S.; Kashiwayanagi, M., The scent of wolves: pyrazine analogs induce
734 avoidance and vigilance behaviors in prey. *Front Neurosci* **2015,** *9*, 363.
735 (58) Osada, K.; Miyazono, S.; Kashiwayanagi, M., Structure-Activity Relationships of Alkylpyrazine
736 Analogs and Fear-Associated Behaviors in Mice. *J Chem Ecol* **2017,** *43* (3), 263-272.
737 (59) Regnier, F. E., Semiochemical--structure and function. *Biol Reprod* **1971,** *4* (3), 309-26.
738 (60) Di Pizio, A.; Niv, M. Y., Computational Studies of Smell and Taste Receptors. *Israel Journal of*
739 *Chemistry* **2014,** *54* (8-9), 1205-1218.
740 (61) Forrest, L. R.; Tang, C. L.; Honig, B., On the accuracy of homology modeling and sequence
741 alignment methods applied to membrane proteins. *Biophys J* **2006,** *91* (2), 508-17.
742 (62) de March, C. A.; Topin, J.; Bruguera, E.; Novikov, G.; Ikegami, K.; Matsunami, H.; Golebiowski,
743 J., Odorant Receptor 7D4 Activation Dynamics. *Angew Chem Int Ed Engl* **2018,** *57* (17), 4554-4558.
744 (63) Nicoli, A.; Dunkel, A.; Giorgino, T.; de Graaf, C.; Di Pizio, A., Classification Model for the
745 Second Extracellular Loop of Class A GPCRs. *J Chem Inf Model* **2022,** *62* (3), 511-522.
746 (64) Yu, Y.; Pacalon, J.; Ma, Z.; Xu, L.; Belloir, C.; Topin, J.; Briand, L.; Golebiowski, J.; Cong, X.,
747 **2021**.
748 (65) Woolley, M. J.; Conner, A. C., Understanding the common themes and diverse roles of the second
749 extracellular loop (ECL2) of the GPCR super-family. *Mol Cell Endocrinol* **2017,** *449*, 3-11.
750 (66) Wink, L. H.; Baker, D. L.; Cole, J. A.; Parrill, A. L., A benchmark study of loop modeling methods
751 applied to G protein-coupled receptors. *J Comput Aided Mol Des* **2019,** *33* (6), 573-595.
752 (67) Won, J.; Lee, G. R.; Park, H.; Seok, C., GalaxyGPCRloop: Template-Based and Ab Initio
753 Structure Sampling of the Extracellular Loops of G-Protein-Coupled Receptors. *J Chem Inf Model* **2018,**
754 *58* (6), 1234-1243.
755 (68) Binder, J. L.; Berendzen, J.; Stevens, A. O.; He, Y.; Wang, J.; Dokholyan, N. V.; Oprea, T. I.,
756 AlphaFold illuminates half of the dark human proteins. *Curr Opin Struct Biol* **2022,** *74*, 102372.
757 (69) Ibrahim, P.; Wifling, D.; Clark, T., Universal Activation Index for Class A GPCRs. *J Chem Inf*
758 *Model* **2019,** *59* (9), 3938-3945.
759 (70) Di Pizio, A.; Waterloo, L. A. W.; Brox, R.; Lober, S.; Weikert, D.; Behrens, M.; Gmeiner, P.; Niv,
760 M. Y., Rational design of agonists for bitter taste receptor TAS2R14: from modeling to bench and back.
761 *Cell Mol Life Sci* **2020,** *77* (3), 531-542.

(71) Bender, B. J.; Gahbauer, S.; Luttens, A.; Lyu, J.; Webb, C. M.; Stein, R. M.; Fink, E. A.; Balius, T. E.; Carlsson, J.; Irwin, J. J.; Shoichet, B. K., A practical guide to large-scale docking. *Nat Protoc* **2021,** *16* (10), 4799-4832.

(72) Halgren, T. A., Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* **2009,** *49* (2), 377-89.

(73) Halgren, T., New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des* **2007,** *69* (2), 146-8.

(74) Sherman, W.; Beard, H. S.; Farid, R., Use of an induced fit receptor structure in virtual screening. *Chem Biol Drug Des* **2006,** *67* (1), 83-4.

(75) Di Pizio, A.; Behrens, M.; Krautwurst, D., Beyond the Flavour: The Potential Druggability of Chemosensory G Protein-Coupled Receptors. *Int J Mol Sci* **2019,** *20* (6).

(76) Man, O.; Gilad, Y.; Lancet, D., Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. *Protein Sci* **2004,** *13* (1), 240-54.

(77) Abaffy, T.; Malhotra, A.; Luetje, C. W., The molecular basis for ligand specificity in a mouse olfactory receptor: a network of functionally important residues. *J Biol Chem* **2007,** *282* (2), 1216-24.

(78) Ahmed, L.; Zhang, Y.; Block, E.; Buehl, M.; Corr, M. J.; Cormanich, R. A.; Gundala, S.; Matsunami, H.; O'Hagan, D.; Ozbil, M.; Pan, Y.; Sekharan, S.; Ten, N.; Wang, M.; Yang, M.; Zhang, Q.; Zhang, R.; Batista, V. S.; Zhuang, H., Molecular mechanism of activation of human musk receptors OR5AN1 and OR1A1 by (R)-muscone and diverse other musk-smelling compounds. *Proc Natl Acad Sci U S A* **2018,** *115* (17), E3950-E3958.

(79) Baud, O.; Etter, S.; Spreafico, M.; Bordoli, L.; Schwede, T.; Vogel, H.; Pick, H., The mouse eugenol odorant receptor: structural and functional plasticity of a broadly tuned odorant binding pocket. *Biochemistry* **2011,** *50* (5), 843-53.

(80) Baud, O.; Yuan, S.; Veya, L.; Filipek, S.; Vogel, H.; Pick, H., Exchanging ligand-binding specificity between a pair of mouse olfactory receptor paralogs reveals odorant recognition principles. *Sci Rep* **2015,** *5*, 14948.

(81) de March, C. A.; Yu, Y.; Ni, M. J.; Adipietro, K. A.; Matsunami, H.; Ma, M.; Golebiowski, J., Conserved Residues Control Activation of Mammalian G Protein-Coupled Odorant Receptors. *Journal of the American Chemical Society* **2015,** *137* (26), 8611-6.

(82) 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* **2005,** *25* (7), 1806-15.

(83) Di Pizio, A.; Shy, N.; Behrens, M.; Meyerhof, W.; Niv, M. Y., Molecular Features Underlying Selectivity in Chicken Bitter Taste Receptors. *Front Mol Biosci* **2018,** *5*, 6.

(84) Dunkel, A.; Hofmann, T.; Di Pizio, A., In Silico Investigation of Bitter Hop-Derived Compounds and Their Cognate Bitter Taste Receptors. *J Agric Food Chem* **2020,** *68* (38), 10414-10423.

(85) Heo, L.; Feig, M., Multi-State Modeling of G-protein Coupled Receptors at Experimental Accuracy. **2022**.

(86) Czerny, M.; Wagner, R.; Grosch, W., Detection of Odor-Active Ethenylalkylpyrazines in Roasted Coffee. *Journal of Agricultural and Food Chemistry* **1996,** *44* (10), 3268-3272.

(87) Frank, O.; Kreissl, J. K.; Daschner, A.; Hofmann, T., Accurate determination of reference materials and natural isolates by means of quantitative (1)h NMR spectroscopy. *J Agric Food Chem* **2014,** *62* (12), 2506-15.

(88) Porcelli, C.; Kreissl, J.; Steinhaus, M., Enantioselective synthesis of tri-deuterated (-)-geosmin to be used as internal standard in quantitation assays. *J Labelled Comp Radiopharm* **2020,** *63* (11), 476-481.

(89) Noe, F.; Frey, T.; Fiedler, J.; Geithe, C.; Nowak, B.; Krautwurst, D., IL-6-HaloTag((R)) enables live-cell plasma membrane staining, flow cytometry, functional expression, and de-orphaning of recombinant odorant receptors. *J Biol Methods* **2017,** *4* (4), e81.

(90) Noe, F.; Geithe, C.; Fiedler, J.; Krautwurst, D., A bi-functional IL-6-HaloTag((R)) as a tool to measure the cell-surface expression of recombinant odorant receptors and to facilitate their activity quantification. *J Biol Methods* **2017,** *4* (4), e82.

(91) Graham, F. L.; Smiley, J.; Russell, W. C.; Nairn, R., Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol* **1977,** *36* (1), 59-74.

(92) Geithe, C.; Andersen, G.; Malki, A.; Krautwurst, D., A Butter Aroma Recombinate Activates Human Class-I Odorant Receptors. *J Agric Food Chem* **2015,** *63* (43), 9410-20.

(93) Saito, H.; Kubota, M.; Roberts, R. W.; Chi, Q.; Matsunami, H., RTP family members induce functional expression of mammalian odorant receptors. *Cell* **2004,** *119* (5), 679-91.

(94) Shirokova, E.; Schmiedeberg, K.; Bedner, P.; Niessen, H.; Willecke, K.; Raguse, J. D.; Meyerhof, W.; Krautwurst, D., Identification of specific ligands for orphan olfactory receptors. G protein-dependent agonism and antagonism of odorants. *J Biol Chem* **2005,** *280* (12), 11807-15.

(95) Jones, D. T.; Reed, R. R., Golf: an olfactory neuron specific-G protein involved in odorant signal transduction. *Science* **1989,** *244* (4906), 790-5.

(96) Li, F.; Ponissery-Saidu, S.; Yee, K. K.; Wang, H.; Chen, M. L.; Iguchi, N.; Zhang, G.; Jiang, P.; Reisert, J.; Huang, L., Heterotrimeric G protein subunit Ggamma13 is critical to olfaction. *J Neurosci* **2013,** *33* (18), 7975-84.

(97) Binkowski, B.; Fan, F.; Wood, K., Engineered luciferases for molecular sensing in living cells. *Curr Opin Biotechnol* **2009,** *20* (1), 14-8.

(98) DeLean, A.; Munson, P. J.; Rodbard, D., Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. *Am J Physiol* **1978,** *235* (2), E97-102.

(99) NCBI Resource Coordinators, Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **2017,** *45* (Database issue), D12-D17.

(100) Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K., MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* **2018,** *35* (6), 1547-1549.

(101) Thompson, J. D.; Higgins, D. G.; Gibson, T. J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **1994,** *22* (22), 4673-80.

(102) Saitou, N.; Nei, M., The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **1987,** *4* (4), 406-25.

(103) Felsenstein, J., Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **1985,** *39*, 783-791.

(104) Zuckerkandl, E.; Pauling, L., *Evolutionary divergence and convergence in proteins*. 1965; p 97-166.

(105) Kooistra, A. J.; Mordalski, S.; Pandy-Szekeres, G.; Esguerra, M.; Mamyrbekov, A.; Munk, C.; Keseru, G. M.; Gloriam, D. E., GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res* **2021,** *49* (D1), D335-D343.

(106) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A., Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **2006,** *Chapter 5*, Unit-5 6.

(107) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: a new

855   approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med*
856   *Chem* **2004,** *47* (7), 1739-49.
857   (108) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.;
858   Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: docking and scoring incorporating a model of
859   hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **2006,** *49* (21), 6177-96.
860   (109) Pedregosa, F. a. V., G. and Gramfort, A. and Michel, V., Thirion, B. and Grisel, O., Blondel, M.,
861   Prettenhofer, P., Weiss, R. and Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M.,
862   Perrot, M., Duchesnay, E., Scikit-learn: Machine Learning in Python. *Journal of Machine Learning*
863   *Research* **2011,** *12*, 2825-2830.
864   (110) Hunter, J. D., Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*
865   **2007,** *9* (3), 90-95.
866   (111) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R., Novel procedure for modeling
867   ligand/receptor induced fit effects. *J Med Chem* **2006,** *49* (2), 534-53.
868   (112) consortium, P., Promoting transparency and reproducibility in enhanced molecular simulations.
869   *Nat Methods* **2019,** *16* (8), 670-673.
870   (113) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G., PLUMED 2: New feathers
871   for an old bird. *Computer Physics Communications* **2014,** *185* (2), 604-613.
872   (114) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.;
873   Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M., PLUMED: A portable plugin for free-energy
874   calculations with molecular dynamics. *Computer Physics Communications* **2009,** *180* (10), 1961-1972.
875   (115) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris,
876   J. H.; Ferrin, T. E., UCSF ChimeraX: Structure visualization for researchers, educators, and developers.
877   *Protein Science* **2020,** *30* (1), 70-82.

878