



HAL
open science

Structural basis of odorant recognition by a human odorant receptor

Christian B Billesbølle, Claire A de March, Wijnand J C van der Velden, Ning Ma, Jeevan Tewari, Claudia Llinas del Torrent, Linus Li, Bryan Faust, Nagarajan Vaidehi, Hiroaki Matsunami, et al.

► To cite this version:

Christian B Billesbølle, Claire A de March, Wijnand J C van der Velden, Ning Ma, Jeevan Tewari, et al.. Structural basis of odorant recognition by a human odorant receptor. *Nature*, 2023, 615 (7953), pp.742-749. 10.1038/s41586-023-05798-y . hal-04220923

HAL Id: hal-04220923

<https://hal.science/hal-04220923v1>

Submitted on 28 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Structural basis of odorant recognition by a human odorant receptor**

2

3 Christian B. Billesbølle^{1*}, Claire A. de March^{2*^}, Wijnand J. C. van der Velden^{3*}, Ning Ma³,
4 Jeevan Tewari², Claudia Llinas del Torrent^{1,4}, Linus Li¹, Bryan Faust¹, Nagarajan Vaidehi^{3#},
5 Hiroaki Matsunami^{2,5#}, Aashish Manglik^{1,6#}

6

7 1. Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, USA

8 2. Department of Molecular Genetics and Microbiology, Duke University, Durham, NC, USA

9 3. Department of Computational and Quantitative Medicine, Beckman Research Institute of the
10 City of Hope, Duarte, CA, USA

11 4. Laboratory of Computational Medicine, Biostatistics Unit, Faculty of Medicine, Universitat
12 Autònoma Barcelona, 08193 Bellaterra, Barcelona, Spain

13 5. Department of Neurobiology, Duke Institute for Brain Sciences, Duke University, Durham,
14 NC, USA

15 6. Department of Anesthesia and Perioperative Care, University of California, San Francisco,
16 CA, USA

17

18 *These authors contributed equally

19 ^Present address: Institut de Chimie des Substances Naturelles, UPR2301 CNRS, Université
20 Paris-Saclay, Gif- sur- Yvette, 91190, France

21 #Correspondence to: Nagarajan Vaidehi (NVaidehi@coh.org), Hiroaki Matsunami
22 (hiroaki.matsunami@duke.edu), or Aashish Manglik (aashish.manglik@ucsf.edu)

23 **Abstract**

24 Our sense of smell enables us to navigate a vast space of chemically diverse odor molecules.
25 This task is accomplished by the combinatorial activation of approximately 400 olfactory G
26 protein-coupled receptors (GPCRs) encoded in the human genome¹⁻³. How odorants are
27 recognized by olfactory receptors (ORs) remains mysterious. Here we provide mechanistic
28 insight into how an odorant binds a human olfactory receptor. Using cryogenic electron
29 microscopy (cryo-EM), we determined the structure of active human OR51E2 bound to the fatty
30 acid propionate. Propionate is bound within an occluded pocket in OR51E2 and makes specific
31 contacts critical to receptor activation. Mutation of the odorant binding pocket in OR51E2 alters
32 the recognition spectrum for fatty acids of varying chain length, suggesting that odorant
33 selectivity is controlled by tight packing interactions between an odorant and an olfactory
34 receptor. Molecular dynamics simulations demonstrate propionate-induced conformational
35 changes in extracellular loop 3 to activate OR51E2. Together, our studies provide a high-
36 resolution view of chemical recognition of an odorant by a vertebrate OR, providing insight into
37 how this large family of GPCRs enables our olfactory sense.

38 INTRODUCTION

39 Our sense of smell relies on our ability to detect and discriminate a vast array of volatile odor
40 molecules. The immense chemical diversity of potential odorants, however, poses a central
41 challenge for the olfactory system of all animals. In vertebrates, odorants are detected by
42 olfactory receptors (ORs), which are G protein-coupled receptors (GPCRs) expressed in
43 olfactory sensory neurons (OSNs) projecting from the olfactory epithelium to the olfactory bulb
44 in the brain^{1,3}. To detect and discriminate the vast diversity of potential odorants⁴, the OR gene
45 family has expanded dramatically in vertebrate genomes, with some species encoding
46 thousands of OR genes⁵. In humans, the approximately 400 functional ORs constitute half of the
47 broader class A GPCR family (**Fig. 1a**)^{6,7}.

48
49 Odorant stimulation of ORs activates signaling pathways via the stimulatory G protein G_{olf} ,
50 which ultimately leads to excitation of OSNs⁸. Each OR can only interact with a subset of all
51 potential odorants. Conversely, a single odorant can activate multiple ORs². This principle of
52 molecular recognition enables a central neural logic of olfaction where the perception of smell
53 arises from the combinatorial activity of multiple unique ORs that respond to an individual
54 odorant². Because each mature OSN expresses only a single OR gene⁹, understanding how an
55 individual OR is activated provides direct insight into the sensory coding of olfaction.

56
57 To understand olfaction at a fundamental level, we need a structural framework describing how
58 odorants are recognized by ORs. Although recent structures of insect odorant-gated ion
59 channels have begun to decipher this molecular logic^{10,11}, the molecular rules that govern
60 odorant recognition in vertebrate ORs are likely distinct and remain obscure. Here, we used
61 cryogenic electron microscopy (cryo-EM) to determine the structure of a human OR activated by
62 an odorant. This structure reveals specific molecular interactions that govern odorant
63 recognition and provides a foundation for understanding how odorant binding activates ORs to
64 instigate cellular signaling.

65 66 **Structure of odorant bound OR51E2**

67 Several challenges have limited structural interrogation of vertebrate ORs, including low
68 expression levels in heterologous systems, low solubility of most volatile odorants, and
69 precipitous instability of purified ORs¹²⁻¹⁵. We therefore sought to identify a human OR that
70 overcomes these challenges. We prioritized a subset of ORs that are also expressed in tissues
71 outside of OSNs with chemoreceptive functions that are independent of olfaction¹⁶. The ability of

72 these ORs to function in non-olfactory tissue suggested that they may be more amenable to
73 expression in heterologous cell expression systems that lack olfactory-tissue specific
74 chaperones¹³. In a second line of reasoning, we prioritized Class I (so called “fish-like”) ORs as
75 these receptors generally recognize water-soluble odorants¹⁷. By contrast, Class II ORs tend to
76 respond to more hydrophobic odorants. Finally, we prioritized ORs that have significant
77 conservation across evolution, potentially because they recognize odorants that are critical for
78 animal survival across many species⁵. We reasoned that such ORs may be more constrained
79 by evolution for stability. With this approach, we identified human OR51E2 as an ideal candidate
80 for structure determination (**Supplementary Fig. 1**). OR51E2 is a Class I OR that responds to
81 the short chain fatty acid propionate¹⁸ (**Fig. 1a,b**). In addition to its olfactory function, OR51E2
82 and its mouse homolog Olfr78 are expressed in several other tissues to enable chemoreception
83 of short chain fatty acids¹⁹⁻²⁴. Consistent with our reasoning, OR51E2 emerged as one of the
84 most highly expressed ORs in HEK293T cells among hundreds of human and mouse ORs that
85 we have previously tested¹².

86
87 To further stabilize OR51E2, we aimed to isolate OR51E2 in a complex with a heterotrimeric G
88 protein. ORs couple with the two highly homologous stimulatory G proteins $G\alpha_{olf}$ and $G\alpha_s$. In
89 mature OSNs, ORs activate $G\alpha_{olf}$ to stimulate cAMP production via adenylyl cyclase⁸. In
90 immature OSNs, ORs activate adenylyl cyclase via $G\alpha_s$ to drive accurate anterior-posterior axon
91 targeting²⁵. Furthermore, OR51E2 signals via $G\alpha_s$ outside of the olfactory system in tissues
92 lacking $G\alpha_{olf}$ ²⁰. The ability of OR51E2 to signal physiologically via $G\alpha_s$, combined with the
93 availability of a nanobody (Nb35) that stabilizes GPCR- $G\alpha_s$ complexes²⁶, prompted us to focus
94 on purifying an OR51E2- G_s complex. To do so, we generated an OR51E2 construct with a C-
95 terminally fused “mini $G\alpha_s$ ” protein. The mini $G\alpha_s$ protein is engineered to trap the receptor-
96 interacting conformation of $G\alpha_s$ in the absence of any guanine nucleotide²⁷. Fusion of the mini G_s
97 to OR51E2 fully blocked propionate stimulated cAMP signaling in HEK293T cells
98 (**Supplementary Fig. 2b**). We surmised that mini $G\alpha_s$ tightly engages the 7TM core of OR51E2
99 to preclude endogenous $G\alpha_s$ coupling and cAMP production.

100
101 We purified OR51E2-mini G_s in the presence of 30 mM propionate, and then further generated a
102 complex with recombinantly purified $G\beta_1\gamma_2$ and Nb35 (**Supplementary Fig. 2a and c**). The
103 resulting preparation was vitrified and analyzed by single particle cryogenic electron microscopy
104 (cryo-EM) (**Supplementary Fig 3 and Supplementary Table 1**), which yielded a 3.1 Å
105 resolution map of OR51E2 bound to the G_s heterotrimer. We additionally generated a map with

106 focused refinement on only the 7TM domain of OR51E2, which afforded improved map
107 resolution of the binding site and extracellular loops of the receptor (**Supplementary Fig. 3e**).
108 The resulting reconstructions allowed us to model the OR51E2 7TM domain, the propionate
109 ligand, and the G_s heterotrimer (**Fig. 1c,d** and **Supplementary Fig. 4a-c**).

110

111 **Odorant binding pocket**

112 We identified cryo-EM density for propionate in a region bounded by transmembrane helices
113 (TM) 3, 4, 5, and 6 in OR51E2 (**Fig. 2a** and **Supplementary Fig. 4b,d**). The propionate odorant
114 binding pocket in OR51E2 is in a similar general region as ligand binding pockets in two
115 prototypical Class A GPCRs: the adrenaline binding site in the β 2-adrenergic receptor (β 2-AR)²⁸
116 and all-trans retinal in rhodopsin²⁹ (**Fig. 2a-c**). Compared to the β 2-AR and rhodopsin, the
117 odorant binding pocket in OR51E2 is smaller and does not engage TM2 and TM7. Extensive
118 packing of the OR51E2 N-terminus with extracellular loops 1 and 2 (ECL1 and ECL2)
119 diminishes the potential size of the odorant binding pocket. Notably, unlike many class A
120 GPCRs with diffusible agonists, the binding pocket for propionate is fully occluded from the
121 extracellular milieu (**Fig. 2d**).

122

123 Propionate makes several contacts within the OR51E2 odorant binding pocket. The carboxylic
124 acid of propionate engages R262^{6.59} (superscript numbers indicate conserved Ballesteros-
125 Weinstein numbering for GPCRs^{30,31}) in TM6 as a counter-ion. The same propionate functional
126 group also engages in hydrogen bonding interactions with S258^{6.55} and Q181 in ECL2 (**Fig. 2e**).
127 We used molecular dynamics (MD) simulations to understand whether these interactions are
128 stable. We performed five 1 μ s simulations of OR51E2 bound to propionate, but in the absence
129 of the G_s heterotrimer. During these simulations, we observed that the carboxylic group of
130 propionate forms a persistent interaction with R262^{6.59}, with an average distance that is identical
131 to that observed in the cryo-EM structure (**Fig. 2f** and **Supplementary Fig. 5**). Simulations also
132 supported persistent interactions between the propionate carboxylic group and S258^{6.55}, with
133 additional contacting residues outlined in **Fig 2g**. Indeed alanine mutations for these carboxylic
134 group coordinating residues, with the exception of Q181^{ECL2}, abolished propionate induced
135 activation of OR51E2 (**Fig. 2h**).

136

137 The van-der Waals contacts between the propionate aliphatic group and OR51E2 are governed
138 by tight packing interactions. The aliphatic portion of propionate contacts residues in TM3
139 (H104^{3.33}), TM4 (F155^{4.57} and L158^{4.60}), and TM5 (G198^{5.39} and I202^{5.43}). Unlike the persistent

140 contacts observed for the oxygens in the carboxylic acid group, interactions between specific
141 propionate carbon atoms and aliphatic residues in OR51E2 were more dynamic in simulations
142 (**Fig. 2g**) and showed minimal contact with F155^{4.57}. However, alanine mutations to G198^{5.39},
143 I202^{5.43} and H104^{3.33} significantly decreased propionate activity at OR51E2, suggesting that
144 there are specific spatial requirements for propionate to bind and activate the receptor. By
145 contrast, propionate is only moderately less efficacious at OR51E2 with the L158^{4.60}A mutation
146 (**Fig. 2h**), likely because this residue only engages the distal C γ carbon of propionate. OR51E2
147 therefore recognizes propionate with specific ionic and hydrogen bonding interactions combined
148 with more distributed van der Waals interactions with tight shape complementarity.

149

150 **Tuning olfactory receptor selectivity**

151 Many ORs are capable of responding to a wide diversity of chemically distinct odorants^{2,18}.
152 Class I ORs, by contrast, are generally more restricted to carboxylic acid odorants³². We tested
153 the selectivity of fatty acid odorants of various chain lengths at OR51E2 to understand how
154 structural features in the receptor lead to odorant specificity. Consistent with previous
155 reports^{23,33}, we identified that acetate (C2) and propionate (C3) activate OR51E2 with millimolar
156 potency (**Fig. 3a,b**). By contrast, longer chain length fatty acids (C4-C10) were either poorly or
157 not active at OR51E2.

158

159 We speculated that the selectivity of OR51E2 for short chain fatty acids arises from the
160 restricted volume of the occluded binding pocket (31 Å³), which would accommodate short chain
161 fatty acids like acetate and propionate but would preclude binding of fatty acids with longer
162 aliphatic chain lengths (**Fig. 3c**). We therefore hypothesized that the volume of the binding
163 pocket acts as a selectivity determinant for fatty acid chain length. To directly test this
164 hypothesis, we designed two mutations that are predicted to result in increased binding pocket
165 volumes while maintaining the specific contacts with R262^{6.59} important for fatty acid activation
166 of OR51E2. More specifically, we mutated two residues that are proximal to the carbon chain of
167 propionate: F155^{4.57} and L158^{4.60}. Computational modeling of the F155^{4.57}A and L158^{4.60}A
168 mutations predicted pocket volumes of 90 Å³ and 68 Å³ respectively, suggesting that both
169 mutants should sufficiently accommodate fatty acids with longer chain length (**Fig. 3c**). Indeed
170 in cAMP assays, both the F155^{4.57}A and L158^{4.60}A OR51E2 mutants were broadly responsive to
171 longer chain fatty acids (**Fig. 3d, Supplementary Table 2 and 3**). The size of each binding
172 pocket was correlated with the maximum chain length tolerated and, additionally, which chain
173 length has the greatest potency. For example, F155^{4.57}A is responsive to a range of fatty acids

174 (C2-C9), with octanoate (C8) displaying maximal potency and efficacy. By contrast, hexanoate
175 (C6) is the most efficacious agonist at the L158^{4,60}A mutant. For both of these mutations, the
176 potency of acetate and propionate is reduced compared to OR51E2, suggesting that tight
177 packing interactions with the aliphatic chain is an important determinant of agonist potency.

178
179 We next examined the conservation of selectivity-determining residues in both human Class I
180 and Class II ORs. Reflecting its importance in fatty acid recognition, arginine is highly conserved
181 in the 6.59 position in most human Class I ORs (Class I 71% vs Class II 7%) (**Supplementary**
182 **Fig. 6**). Positions 4.57 and 4.60 in all human Class I ORs are constrained to aliphatic amino
183 acids of different size (V/I/L/M/F, Class I >80% vs Class II <15%). By contrast, none of these
184 positions have similar constraints in Class II ORs. We surmise that the conserved residue R^{6,59}
185 may anchor odorants in many Class I OR binding pockets, while diversity in the 4.57 and 4.60
186 positions tunes the binding pocket to enable selective recognition of the remainder of the
187 molecule. Two features may therefore drive odorant recognition for Class I ORs: 1) hydrogen-
188 bonding or ionic interactions that anchor polar features of odorants to conserved OR binding
189 pocket residues, and 2) van-der Waals interactions of diverse aliphatic residues in the OR
190 binding pocket that define a closed volume having a geometry that closely matches the shape of
191 cognate odorants.

192

193 **Activation mechanisms of OR51E2**

194 Odorant binding to ORs is predicted to cause conformational changes in the receptor that
195 enable G protein engagement. Our strategy to stabilize OR51E2 with miniG_s precluded structure
196 determination of inactive OR51E2 in the absence of an odorant. We therefore turned to
197 comparative structural modeling, mutagenesis studies, and molecular dynamics simulations to
198 understand the effect of propionate binding on the conformation of OR51E2.

199

200 Comparison of active OR51E2 to G_s-coupled, active state β 2-adrenergic receptor (β 2-AR)
201 demonstrated that both receptors engage the G protein with a similar overall orientation of the
202 7TM domain and G α_s (**Fig. 4a** and **Supplementary Fig. 7**). A central hallmark of GPCR
203 activation is an outward displacement of TM6 in the cytoplasmic side of the receptor, which is
204 accompanied by more subtle movement of the other TM helices³⁴. These conformational
205 changes create a cavity for the G protein C-terminal α -helix. Prior structural biology studies have
206 identified two regions conserved in Class A GPCRs that are critically important for allosteric
207 communication between the agonist binding site and the G protein-binding site: a connector

208 region that is adjacent to the ligand binding site and a G protein-coupling region adjacent to the
209 $G\alpha_s$ C-terminal α -helix³⁴ (**Fig. 4a**). We aimed to understand how propionate binding to OR51E2
210 stabilizes these regions in an active conformation. Although the overall conformation of OR51E2
211 and β 2-AR are similar (root-mean-square deviation (RMSD) of 3.2 Å), the specific sequences
212 that define the G protein-coupling and connector regions are distinct between ORs and non-
213 olfactory Class A GPCRs. Comparison of sequence conservation in TM6 between human ORs
214 and non-olfactory Class A GPCRs revealed a highly conserved motif (KAFSTCxSH^{6.40}) in the G
215 protein coupling region in ORs that is absent in non-olfactory receptors (**Fig. 4b**). By contrast,
216 the highly conserved CWxP^{6.50} motif in the connector region of Class A GPCRs is absent in
217 ORs. Instead ORs contain the previously described FYGx^{6.50} motif in the connector region³⁵
218 (**Fig. 4f,g**).

219
220 Closer inspection of the G protein-coupling region in OR51E2 revealed a unique hydrogen-
221 bonding network between the highly conserved residues R121^{3.50} in TM3, H243^{6.40} in TM6 and
222 Y217^{5.58} in TM5 that is not observed in other Class A GPCRs (**Fig. 4c,d**). Activation of the β 2-
223 AR is associated with an inward movement of TM7 that positions Y316^{7.53} within a water-
224 mediated hydrogen bonding distance of Y219^{5.58}; this movement leads to outward movement of
225 TM6 by displacing the aliphatic I278^{6.40} residue (**Fig. 4d**). Given the high conservation of R^{3.50},
226 H^{6.40} and Y^{5.58} across all ORs (89%, 97% and 93%, respectively, **Supplementary Fig. 7**), we
227 propose that this contact is important in stabilizing the OR active conformation. Indeed alanine
228 mutagenesis of OR51E2 residues in the G protein coupling region show a dramatic loss of
229 activity for H243^{6.40}, Y217^{5.58}, and R121^{3.50} associated with poor receptor expression (**Fig. 4e**
230 and **Supplementary Table 2**). Mutation of Y291^{7.53} in OR51E2, by contrast, has a more modest
231 effect on propionate activity.

232
233 We next examined the connector region of OR51E2 directly adjacent to the propionate binding
234 site (**Fig. 4f**). Activation of the β 2-AR is associated with a rearrangement of the PIF motif
235 between positions I^{3.40} (TM3), P^{5.50} (TM5), and F^{6.44} (TM6), which leads to an outward
236 displacement of TM6. This coordinated movement has been observed for the majority of class A
237 GPCRs for which both active and inactive state structures have been obtained³⁴. Conservation
238 at the PIF positions is low in ORs, suggesting an alternative mechanism. In OR51E2, we
239 observe an extended hydrogen bonding network between Y251^{6.48} of the OR-specific FYGx
240 motif and residues in TM3 (S111^{3.40}), TM4 (R150^{4.52}), and TM5 (D209^{5.50}). Notably, the
241 intramembrane ionic interaction between D209^{5.50} and R150^{4.52} is likely only conserved in Class

242 I ORs (Class I: D^{5.50}-82%, R^{4.52}-88%, Class II: D^{5.50}-0.3%, R^{4.52}-0%, **Supplementary Fig. 7**).
243 Alanine mutagenesis of most residues in this connector region of OR51E2 abolishes response
244 to propionate (**Fig. 4h**), in part because mutations in this region dramatically decrease receptor
245 expression (**Supplementary Table 2**). More conservative substitutions to F250^{6.47} or Y251^{6.48}
246 also show impairment in OR51E2 function, suggesting that the specific contacts observed in
247 active OR51E2 are important for robust receptor activation.

248
249 We turned to molecular dynamics simulations to examine how ligand binding influences the
250 conformation of the connector region. After removing the G protein, we simulated OR51E2 with
251 and without propionate in the binding site. For each condition, we performed five 1 μ s
252 simulations. OR51E2 simulated with propionate remains in a conformation similar to the cryo-
253 EM structure. In the absence of propionate, the connector region of OR51E2 displays
254 significantly more flexibility in simulations (**Fig. 4i** and **Supplementary Fig. 8**). We observed
255 two motions in the FYGx motif associated with this increased conformational heterogeneity: a
256 rotameric flexibility of F250^{6.47} between the experimentally observed conformation and
257 alternative rotamers and a disruption of a hydrogen bond between Y251^{6.48} and S111^{3.40} (**Fig.**
258 **4j,k** and **Supplementary Fig. 8**). Simulations without propionate show that the distance
259 between the hydroxyl groups of Y251^{6.48} and S111^{3.40} is >4 Å, indicating the loss of a hydrogen
260 bond that was observed in both the cryo-EM structure of OR51E2 and the MD simulations with
261 propionate (**Fig. 4k** and **Supplementary Fig. 8**). Based on structural comparison to other Class
262 A GPCRs, mutagenesis studies, and molecular dynamics simulations, we therefore propose
263 that odorant binding stabilizes the conformation of an otherwise dynamic FYGx motif to drive
264 OR activation.

265 266 **Structural dynamics of ECL3 in OR function**

267 Olfactory receptors display significant sequence variation in extracellular loop 3 (ECL3), a
268 region previously shown to be critical for recognition of highly diverse odorants^{36,37}. We
269 therefore aimed to understand the involvement of ECL3 in propionate binding to OR51E2, and
270 more generally, how ECL3 may drive the conformational changes in TM6 necessary for OR
271 activation (**Fig. 5**). In our structure of OR51E2, ECL3 is directly coupled to odorant binding via a
272 direct interaction between the carboxylic acid moiety of propionate and the ECL3 adjacent
273 residue R262^{6.59} (**Fig. 5a**). In order to investigate the role of R262^{6.59} in maintaining the
274 conformation of ECL3 by binding the odorant, we analyzed simulations of OR51E2 performed
275 without propionate. In the absence of coordination with the carboxylic acid group of propionate,

276 R262^{6.59} showed a marked increase in flexibility, with an outward movement of up to 8 Å away
277 from the ligand binding site (**Fig. 5b** and **c**). This movement is accompanied by displacement of
278 ECL3 away from the odorant binding pocket.

279
280 To test whether inward movement of R262^{6.59} is itself sufficient to activate OR51E2, we
281 designed a gain-of-function experiment. We hypothesized that introduction of an acidic residue
282 in the binding pocket with an appropriate geometry may substitute for the carboxylic acid of
283 propionate and coordinate R262^{6.59}. Indeed, substitution of Asp in position 181 (Q181D) of
284 OR51E2 yielded increased cAMP basal activity (**Fig. 5d**). By contrast, introduction of Glu in the
285 same position (Q181E) rendered OR51E2 largely inactive, suggesting the requirement for a
286 precise coordination geometry for R262^{6.59}. Substitution with the larger Gln (Q181N) rendered
287 OR51E2 completely unresponsive to propionate, either by sterically blocking R262^{6.59} or by
288 displacing propionate itself. In simulations of OR51E2 with the Q181D substitution, R262^{6.59} is
289 persistently engaged toward the ligand binding site (**Fig. 5b**). Furthermore, this inward
290 movement of R262^{6.59} and ECL3 is accompanied by activation-associated conformational
291 changes in the connector domain of OR51E2 (**Supplementary Fig. 9**), perhaps explaining the
292 basal activity of Q181D mutant. Inward movement of ECL3 is therefore sufficient to activate
293 OR51E2.

294
295 Because conformational changes in ECL3 are critical to OR51E2 activation, we speculated that
296 this region may provide a common activation mechanism across the OR family. To probe this
297 notion, we examined structural predictions of all human olfactory receptors by AlphaFold2³⁸. We
298 first compared the AlphaFold2 prediction for OR51E2 with the cryo-EM structure, which yielded
299 a high degree of agreement reflected in a RMSD of 1.3 Å for C α atoms. Importantly, the
300 AlphaFold2 predicted structure of OR51E2 appears to be in an intermediate or inactive
301 conformation characterized by outward displacement of R262^{6.59} and ECL3, a G protein-
302 coupling domain in the inactive conformation, and TM6 more inwardly posed compared to active
303 OR51E2 (**Fig 5e** and **Supplementary Fig. 10**). We next examined the predicted structures of all
304 human ORs, which revealed a largely shared topology for the extracellular region for the
305 broader family (**Fig. 5f**). Indeed, the per-residue confidence score from AlphaFold2 (predicted
306 local distance difference test, pLDDT) for the N terminus, ECL1, and ECL2 are predicted with
307 high confidence for the most ORs. By contrast, ECL3 shows significantly lower pLDDT scores.
308 Because low pLDDT scores correlate with disordered protein regions³⁸, we surmise that, in the
309 absence of odorant binding, the structure of ECL3 is less constrained compared to the rest of

310 the odorant binding pocket for the broader OR family. Similar to OR51E2, odorant binding may
311 therefore stabilize ECL3 to drive receptor activation for the broader OR family.

312

313 **Discussion**

314 We propose the following model for OR51E2 activation (**Fig. 5g**). In the unbound state, the
315 extracellular segment of TM6 is dynamic. Upon binding of propionate, TM6 rotates inward
316 towards the 7TM domain and is stabilized via a direct coordination of the propionate carboxylic
317 acid via R262^{6,59}. The conserved FYGx motif in TM6 acts as a structural pivot point around
318 which TM6 rotates to displace the intracellular end of TM6 from the TM-core and open the
319 canonical active G protein-binding site. Although specific interactions between the propionate
320 aliphatic chain and residues within the binding site are important for achieving full potency of the
321 odorant response, OR51E2 is constitutively active when an aspartate residue (Q181D) is
322 introduced in the binding pocket. This suggests that the observed rotation of TM6 mediated by
323 coordination of R262 with a stable anionic group in the binding site, in itself is sufficient for
324 receptor activation. While this model remains speculative due to the lack of an experimentally-
325 determined inactive-state structure of OR51E2, it integrates the findings from unique structural
326 features of ORs compared to other Class A GPCRs, molecular dynamics simulations, and
327 mutagenesis studies. A similar mechanism may be responsible for the activation of most Class I
328 ORs, a large majority of which recognize carboxylic acids and contain an arginine at position
329 6.59. The mechanism of activation of Class II ORs, which recognize a broader range of volatile
330 odorants and lack R^{6.59}, could be potentially distinct.

331

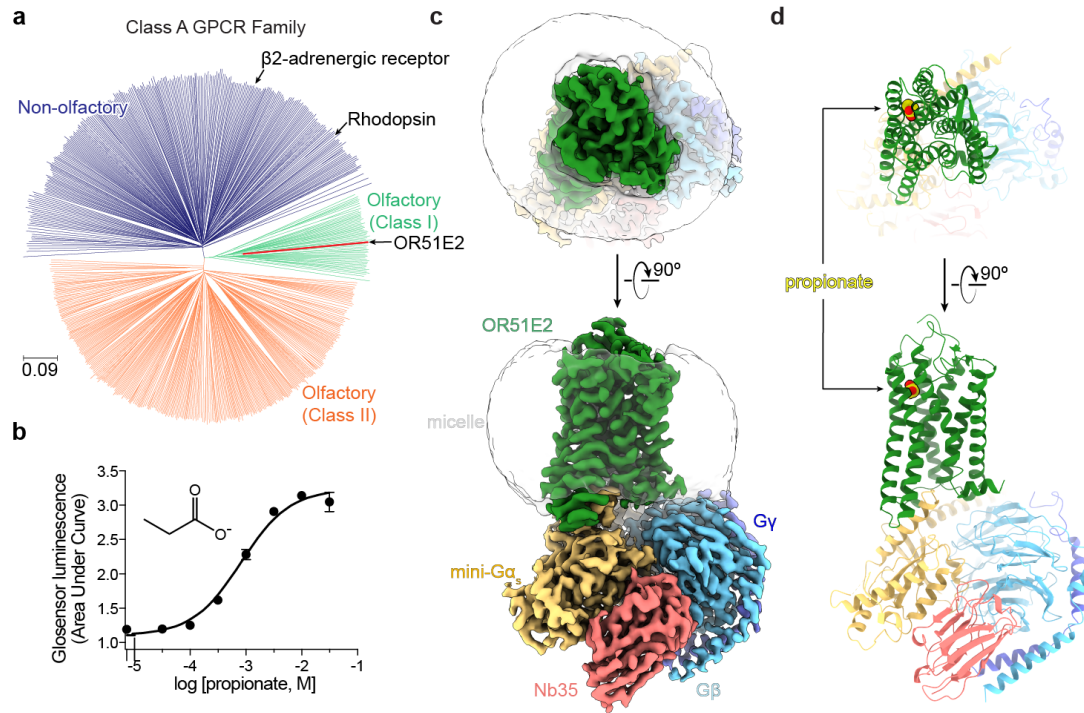
332 Our work illuminates the molecular underpinnings of odorant recognition in a vertebrate Class I
333 OR. While the full breadth of potential odorants that activate OR51E2 remains to be
334 characterized, profiling of known fatty acid odorants suggests that OR51E2 is narrowly tuned to
335 short chain fatty acids^{18,23}. Propionate binds OR51E2 with two types of interactions - specific
336 ionic and hydrogen bonding interactions that anchor the carboxylic acid, and more nonspecific
337 hydrophobic contacts that rely on shape complementarity with the aliphatic portion of the ligand.
338 We demonstrate that the specific geometric constraints imposed by the occluded OR51E2
339 odorant binding pocket are responsible, in part, for this selectivity. Molecular recognition in
340 OR51E2 is therefore distinct from the distributed hydrophobic interactions that mediate odorant
341 recognition at an insect odorant-gated ion channel¹¹. We anticipate that the molecular
342 mechanism we define here for OR51E2 is likely to extend to other Class I ORs that recognize
343 polar, water soluble odorants with multiple hydrogen bond acceptors and donors. Molecular

344 recognition by more broadly tuned ORs, and the larger Class II OR family, however, remains to
345 be defined.

346

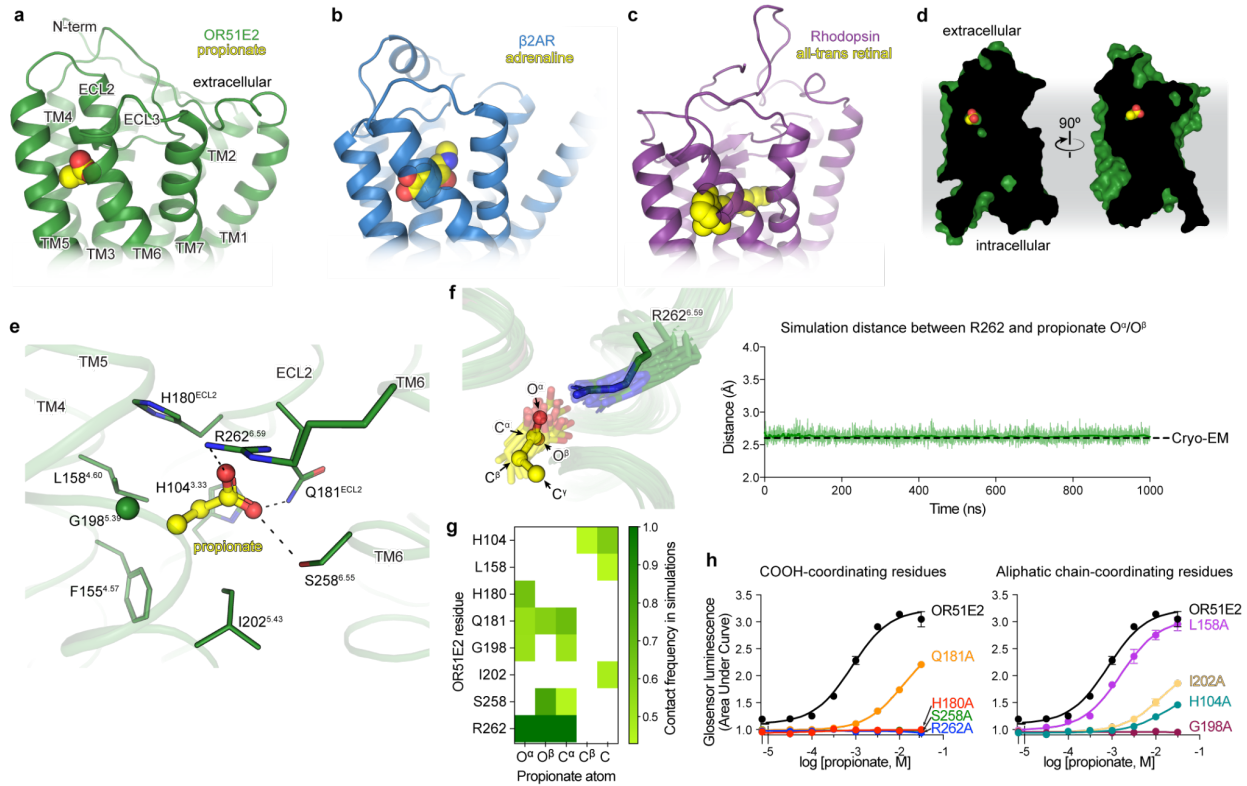
347 The structural basis of ligand recognition for OR51E2 also provides insight into evolution of the
348 OR family. Unlike most vertebrate OR genes that have evolved rapidly via gene duplication and
349 diversification, OR51E2 is one of a few ORs with strong evolutionary conservation within
350 different species⁵. This constraint may result from recognition of odorants important for survival
351 or from vital non-olfactory roles of OR51E2 activity detecting propionate and acetate, the main
352 metabolites produced by the gut microbiota. Molecular recognition of propionate by OR51E2
353 may therefore represent a unique example of specificity within the broader OR family. While
354 future work will continue to decipher how hundreds of ORs sense an immensely large diversity
355 of odorants, our structure and mechanistic insight into OR51E2 function provides a new
356 foundation to understand our sense of smell at an atomic level.

357 **Main text figures:**



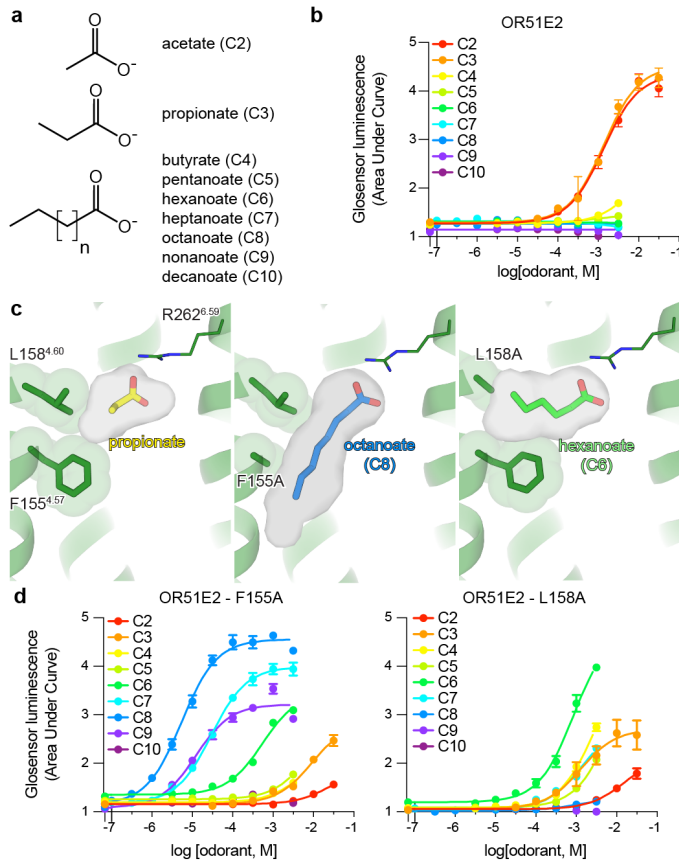
358

359 **Figure 1. Structure of human olfactory receptor OR51E2.** **a)** Phylogenetic tree of human
360 Class A GPCRs, including both non-olfactory (blue) and olfactory receptors. Olfactory receptors
361 are further divided into Class I (green) and Class II (orange). OR51E2 is a Class I OR. The
362 phylogenetic distance scale is represented on the left bottom corner (the distance represents
363 9% differences between sequences). **b)** Real-time monitoring of cAMP concentration assay
364 showing that human OR51E2 responds to the odorant propionate. Data points are mean \pm
365 standard deviation from $n = 4$ replicates. Cryo-EM density map (**c**) and ribbon model (**d**) of
366 active human OR51E2 bound to propionate (yellow spheres) in complex with G_s heterotrimer
367 and stabilizing nanobody Nb35.



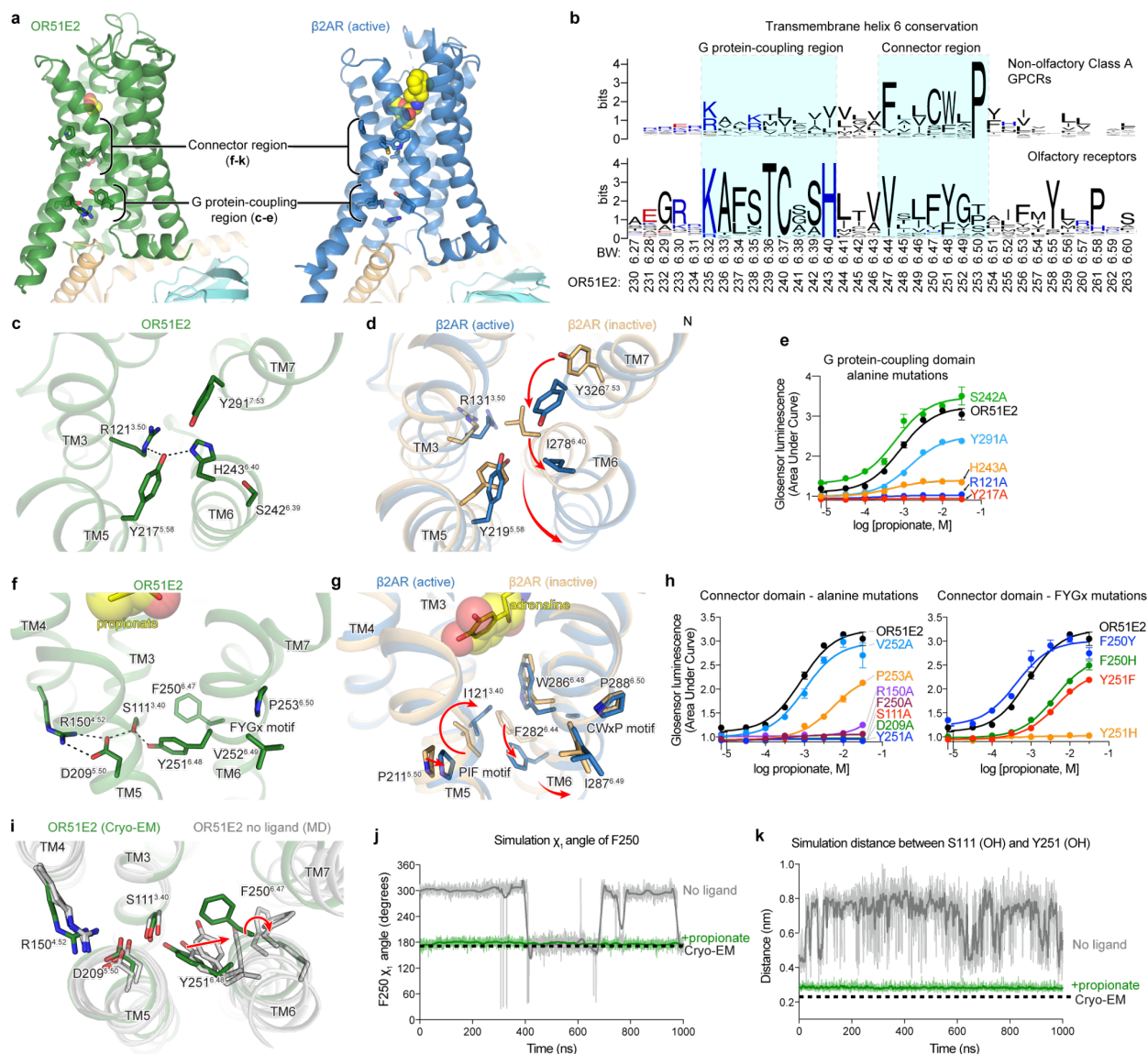
368

369 **Figure 2. Odorant binding pocket in OR51E2.** Comparison of propionate binding site in
 370 OR51E2 (**a**) to two other prototypical Class A GPCRs, the β 2-adrenergic receptor (β 2AR) bound
 371 to adrenaline (PDB 4LDO)²⁸ (**b**) and rhodopsin bound to all-trans retinal (PDB 6FUF)²⁹ (**c**).
 372 Propionate primarily contacts TM4, TM5, TM6 and ECL2. By contrast adrenaline and all-trans
 373 retinal make more extensive contacts with other GPCR transmembrane helices. **d**) The binding
 374 site of propionate in active OR51E2 is occluded from extracellular solvent. **e**) Close-up view of
 375 propionate binding site in OR51E2. **f**) Molecular dynamics simulations snapshots of OR51E2
 376 bound to propionate are shown as transparent sticks and overlaid on the cryo-EM structure.
 377 R262^{6.59} makes a persistent contact with propionate over 1000 ns of simulation time (see
 378 Supplementary Fig. 8 for simulation replicates). The minimum distance between any of R262^{6.59}
 379 sidechain nitrogens and propionate oxygens is shown. **g**) Heatmap of contact frequencies of
 380 interaction between OR51E2 binding site residues and propionate atoms (as labeled in (f))
 381 obtained from five independent molecular dynamics simulations each 1 μ s long (total time 5 μ s).
 382 Contact frequency cutoff between receptor residue and ligand atoms were set at 40%. **h**)
 383 Alanine mutagenesis analysis of propionate-contacting residues in OR51E2 using a real-time
 384 monitoring of cAMP concentration assay. Data points are mean \pm standard deviation from $n = 3$
 385 experiments.



386

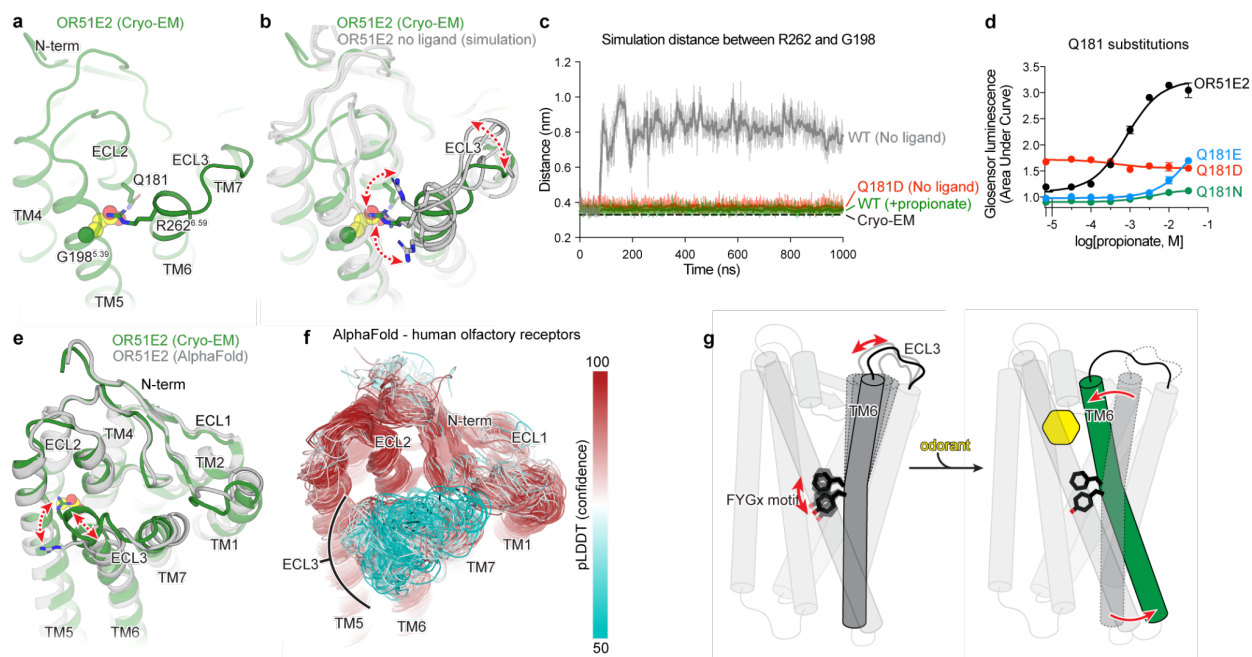
387 **Figure 3. Tuning OR51E2 odorant selectivity.** **a,b)** OR51E2 responds selectively to the short
388 chain fatty acids acetate and propionate as measured by a cAMP production assay. **c)** Close up
389 view of the OR51E2 binding pocket with the binding pocket cavity shown as gray surface.
390 Replacement of F155^{4,57} and L158^{4,60} with alanine is predicted to yield a binding pocket with
391 increased volume capable of accommodating a longer chain fatty acid. Docked poses of
392 octanoate (C8) and hexanoate (C6) are shown in the F155A and L158A mutants of OR51E2,
393 respectively. **d)** The F155A and L158A mutations in OR51E2 lead to increased sensitivity to
394 long chain fatty acids. Conversely, the potency for acetate and propionate is reduced for these
395 two mutants. Data points in b and d are mean \pm standard deviation from $n = 4$ experiments.



396
 397 **Figure 4. Activation mechanism of OR51E2.** **a)** Ribbon diagram comparing structures of
 398 propionate-bound OR51E2-miniGs complex (green) to BI-167107 bound β 2AR- G_s complex
 399 (blue, PDB 3SN6). For both receptors, the connector region couples conformational changes at
 400 the ligand binding site with the G protein-coupling region. **b)** Weblogo depicting conservation of
 401 transmembrane helix 6 amino acids in either human olfactory receptors or human non-olfactory
 402 Class A GPCRs. Amino acid numbering for OR51E2 and Ballosteros-Weinsten (BW) are
 403 indicated. Close-up view of the G protein-coupling domain in active OR51E2 **(c)** and both active
 404 and inactive β 2AR **(d)**. Activation of β 2AR is associated with an inward movement of TM7 and a
 405 contact between Y219^{5.58} and Y326^{7.53}. In OR51E2, by contrast, H243^{6.40} interacts with Y217^{5.58}
 406 in the active state. **e)** Alanine mutagenesis analysis of G protein-coupling domain residues in
 407 OR51E2 using a real-time monitoring of cAMP concentration assay. Close-up views of the
 408 connector region in active OR51E2 **(f)** and both active and inactive β 2AR **(g)**. **h)** Mutagenesis
 409 analysis of connector region residues in OR51E2 using a real-time monitoring of cAMP
 410 concentration assay. **i)** Snapshots from molecular dynamics (MD) simulations of OR51E2 with

411 propionate removed. Simulations show increased flexibility of TM6 in the connector region
412 residues. Molecular dynamics trajectories for representative simulations showing rotation of side
413 chain rotamer angle of F250^{6.47} (**j**) and distance between S111^{3.40} and Y251^{6.48} hydroxyl groups
414 (**k**) performed with or without propionate over the course of 1000 ns MD simulation (see
415 Supplementary Fig. 8 for simulation replicates). Data points in e and h mean \pm standard
416 deviation from $n = 4$ experiments.

417



418

419

420 **Figure 5: Structural dynamics of ECL3 in OR function. a)** Residue R262^{6,59} in ECL3 makes a

421 critical contact with propionate. Residue Q181^{ECL2} in ECL2 is highlighted. **b)** Molecular dynamics

422 snapshots of OR51E2 simulated in the absence of propionate shows increased flexibility of

423 R262^{6,59}. **c)** In simulations of wild-type (WT) OR51E2 bound to propionate, the distance

424 between R262^{6,59} and G198^{5,39} is stable and similar to the cryo-EM structure. Simulations of WT

425 OR51E2 without propionate (no ligand) show increased distances between R262^{6,59} and

426 G198^{5,39}. In simulations of Q181D mutant without propionate, the distance between R262^{6,59}

427 and G198^{5,39} is similar to WT OR51E2 bound to propionate. Distance was measured between

428 R262^{6,59} sidechain atoms and G198^{5,39} main chain atoms (excluding the hydrogens) over the

429 course of 1000 ns MD simulation (see Supplementary Fig. 8 for simulation replicates). **d)**

430 Conservative mutagenesis of Q181^{ECL2} shows that the Q181D mutant is constitutively active,

431 potentially because it substitutes a carboxylic acid in the OR51E2 binding pocket. **e)**

432 Comparison of cryo-EM structure of OR51E2 with the AlphaFold2 predicted structure shows

433 high similarity in the extracellular domain with the exception of the ECL3 region. The AlphaFold2

434 model shows an outward displacement of R262^{6,59} and ECL3 similar to simulations of apo

435 OR51E2. **f)** AlphaFold2 predictions for all human olfactory receptors show low confidence in the

436 ECL3 region and high confidence in other extracellular loops. **g)** A model for ECL3 as a key site

437 for olfactory receptor activation.

437 **METHODS**

438

439 **Expression and purification of OR51E2-miniG_s protein**

440 Human *OR51E2* was cloned into pCDNA-Zeo-TetO, a custom pcDNA3.1 vector containing a
441 tetracycline inducible gene-expression cassette⁴¹. The construct included an N-terminal
442 influenza hemagglutinin signal sequence and the FLAG (DYKDDDK) epitope tag. The construct
443 further included the miniG_{s399} protein⁵, which was fused to the C-terminus of OR51E2 with a
444 human rhinovirus 3C (HRV 3C) protease cleavage sequence flanked by Gly-Ser linkers.

445

446 The resulting construct (OR51E2-miniG_{s399}) was transfected into 1 L of inducible Expi293F-TetR
447 cells (unauthenticated and untested for mycoplasma contamination, Thermo Fisher) using the
448 ExpiFectamine 293 Transfection Kit (Thermo Fisher) as per manufacturer's instructions. After
449 16 hours, protein expression was induced with 1 µg/mL doxycycline hyclate (Sigma Aldrich),
450 and the culture was placed in a shaking incubator maintaining 37°C and a 5% CO₂ atmosphere.
451 After 36 hours cells were harvested by centrifugation and stored at -80°C.

452 For receptor purification, cells were thawed and hypotonically lysed in 50 mM HEPES, pH 7.50,
453 1 mM EDTA, 30 mM sodium propionate (Sigma Aldrich), 100 µM tris(2-carboxyethyl)phosphine
454 (TCEP, Fischer Scientific), 160 µg/mL benzamidine, 2 µg/mL leupeptin for 15 minutes at 4°C.

455 Lysed cells were harvested by centrifugation at 16,000 x g for 15 min, and immediately dounce-
456 homogenized in ice-cold solubilization buffer comprising 50 mM HEPES, pH 7.50, 300 mM
457 NaCl, 1% (w/v) Lauryl Maltose Neopentyl Glycol (L-MNG, Anatrace), 0.1% (w/v) cholesteryl
458 hemisuccinate (CHS, Steraloids), 30 mM sodium propionate, 5 mM adenosine 5'-triphosphate
459 (ATP, Fischer Scientific), 2 mM MgCl₂, 100 µM TCEP, 160 µg/mL benzamidine, and 2 µg/mL
460 leupeptin. The sample was stirred for 2 hours at 4°C, and the detergent-solubilized fraction was

461 clarified by centrifugation at 20,000 x g for 30 min. The detergent-solubilized sample was
462 supplemented with 4 mM CaCl₂ and incubated in batch with homemade M1-FLAG-antibody
463 conjugated CNBr-sepharose under slow rotation for 1.5 hours at 4°C. The sepharose resin was

464 transferred to a glass column and washed with 20 column volumes ice-cold buffer comprising
465 50 mM HEPES, pH 7.50, 300 mM NaCl, 0.05% (w/v) L-MNG, 0.005% (w/v) CHS, 30 mM
466 sodium propionate, 2.5 mM ATP, 4 mM CaCl₂, 2 mM MgCl₂, and 100 µM TCEP. This was

467 followed by 10 column volumes of ice-cold 50 mM HEPES, pH 7.50, 150 mM NaCl, 0.0075%
468 (w/v) L-MNG, 0.0025% glyco-diosgenin (GDN, Anatrace), 0.001% (w/v) CHS, 30 mM sodium
469 propionate, 4 mM CaCl₂, and 100 µM TCEP. Receptor containing fractions were eluted with ice-
470 cold 50 mM HEPES, pH 7.50, 150 mM NaCl, 0.0075% (w/v) L-MNG, 0.0025% (w/v) GDN,

471 0.001% (w/v) CHS, 30 mM sodium propionate, 5 mM EDTA, 100 μ M TCEP, and 0.2 mg/mL
472 FLAG peptide. Fractions containing OR51E2-miniG_{s399} fusion protein were concentrated in a 50
473 kDa MWCO spin filter (Amicon) and further purified over a Superdex 200 Increase 10/300 GL
474 (Cytiva) size-exclusion chromatography (SEC) column, which was equilibrated with 20 mM
475 HEPES, pH 7.50, 150 mM NaCl, 0.005% (w/v) GDN, and 0.0005% CHS, 30 mM sodium
476 propionate, and 100 μ M TCEP. Fractions containing monodisperse OR51E2-miniG_{s399} were
477 combined and concentrated in a 50 kDa MWCO spin filter prior to complexing with G β ₁ γ ₂ and
478 Nb35.

479

480 **Expression and purification of G β ₁ γ ₂**

481 A baculovirus was generated with the pVLDual expression vector encoding both the human G β ₁
482 subunit with a HRV 3C cleavable N-terminal 6x His-tag and the untagged human G γ ₂ subunit, in
483 *Spodoptera frugiperda* Sf9 insect cells (unauthenticated and untested for mycoplasma
484 contamination, Expression Systems). For expression, *Trichoplusia ni* Hi5 insect cells
485 (unauthenticated and untested for mycoplasma contamination, Expression Systems) were
486 infected at a density of 3.0 x 10⁶ cells/mL with high titer G β ₁ γ ₂-baculovirus, and grown at 27 °C
487 with 130 rpm shaking. After 48 hours, cells were harvested and resuspended in lysis buffer
488 comprising 20 mM HEPES, pH 8.00, 5 mM β -mercaptoethanol (β -ME), 20 μ g/mL leupeptin, and
489 160 μ g/mL benzamidine. Lysed cells were pelleted at 20,000 x g for 15 min, and solubilized with
490 20 mM HEPES pH 8.0, 100 mM sodium chloride, 1% (w/v) sodium cholate (Sigma Aldrich),
491 0.05% (w/v) n-dodecyl- β -D-maltopyranoside (DM, Anatrace), and 5 mM β -mercaptoethanol (β -
492 ME). Solubilized G β ₁ γ ₂ was clarified by centrifugation at 20,000 x g for 30 min, and was then
493 incubated in batch with HisPur Ni-NTA resin (Thermo Scientific). Resin-bound G β ₁ γ ₂ was
494 washed extensively, before detergent was slowly exchanged on-column to 0.1% (w/v) L-MNG,
495 and 0.01% (w/v) CHS. G β ₁ γ ₂ was eluted with 20 mM HEPES pH 7.50, 100 mM NaCl, 0.1% (w/v)
496 L-MNG, 0.01% (w/v) CHS, 300 mM imidazole, 1 mM DL-Dithiothreitol (DTT), 20 μ g/mL leupeptin,
497 and 160 μ g/mL benzamidine. Fractions containing G β ₁ γ ₂ were pooled and supplemented with
498 homemade 3C protease before overnight dialysis into buffer comprised of 20 mM HEPES pH
499 7.50, 100 mM NaCl, 0.02% (w/v) L-MNG, 0.002% (w/v) CHS, 1 mM DTT, and 10 mM imidazole.
500 Uncleaved G β ₁ γ ₂ was removed by batch incubation with Ni-NTA resin, before the unbound
501 fraction containing cleaved G β ₁ γ ₂ was dephosphorylated by treatment with lambda phosphatase
502 (New England Biolabs), calf intestinal phosphatase (New England Biolabs), and antarctic
503 phosphatase (New England Biolabs) for 1 hour at 4°C. Geranylgeranylated G β ₁ γ ₂ heterodimer
504 was isolated by anion exchange chromatography using a MonoQ 4.6/100 PE (Cytiva) column,

505 before overnight dialysis in 20 mM HEPES, pH 7.50, 100 mM NaCl, 0.02% (w/v) L-MNG, and
506 100 μ M TCEP. Final sample was concentrated on a 3 kDa MWCO spin filter (Amicon), and 20%
507 (v/v) glycerol was added before flash freezing in liquid N₂ for storage at -80°C.

508

509 **Expression and purification of Nb35**

510 DNA encoding Nb35 (described by Rasmussen *et al.*⁶) was cloned into a modified pET-26b
511 expression vector harboring a C-terminal His-tag followed by a Protein C (EDQVDPRLIDGK)
512 affinity tag. The resulting DNA was transformed into competent Rosetta2 (DE3) pLysS
513 *Escherichia coli* (UC Berkeley QB3 MacroLab) and inoculated into 100 ml Luria Broth
514 supplemented with 50 μ g/mL kanamycin, which was cultured overnight with 220 rpm shaking at
515 37°C. The following day, the starter culture was inoculated into 8 x 1 L of Terrific Broth
516 supplemented with 0.1% (w/v) dextrose, 2 mM MgCl₂, and 50 μ g/mL kanamycin which were
517 further cultured at 37°C with shaking. Nb35 expression was induced at OD₆₀₀ = 0.6, by addition
518 of 400 μ M Isopropyl β -D-1-thiogalactopyranoside (IPTG, GoldBio) and lowering the incubator
519 temperature to 20°C. After 21 hours of expression, cells were harvested by centrifugation and
520 were resuspended in SET Buffer comprising 200 mM tris(hydroxymethyl)aminomethane (Tris,
521 Sigma Aldrich), pH 8.00, 500 mM sucrose, 0.5 mM EDTA, 20 μ g/mL leupeptin, 160 μ g/mL
522 benzamidine, and 1 U benzonase. After 30 minutes of stirring at RT, hypotonic lysis was initiated
523 by a 3-fold dilution with deionized water. Following 30 minutes of stirring at RT, ionic strength
524 was adjusted to 150 mM NaCl, 2 mM CaCl₂, and 2 mM MgCl₂ and the lysate was cleared by
525 centrifugation at 20,000 x g for 30 min. The cleared lysate was incubated in batch with
526 homemade anti-Protein C antibody coupled CNBr-sepharose under slow rotation. The resin was
527 extensively washed with buffer comprising 20 mM HEPES, pH 7.50, 300 mM NaCl, and 2 mM
528 CaCl₂, and Nb35 was eluted with 20 mM HEPES pH 7.50, 100 mM NaCl, 0.2 mg/mL Protein C
529 peptide, and 5 mM EDTA. Nb35 containing fractions were concentrated in a 10 kDa MWCO spin
530 filter (Amicon) and further purified over a Superdex S75 Increase 10/300 GL column (Cytiva)
531 SEC column equilibrated with 20 mM HEPES, pH 7.50, and 100 mM NaCl. Fractions containing
532 monodisperse Nb35 were concentrated and supplemented with 20% glycerol prior to flash
533 freezing in liquid N₂ for storage at -80°C.

534

535 **Preparation of the active-state OR51E2-G_s complex**

536 To prepare the OR51E2-G_s complex, a 2-fold molar excess of purified G β ₁ γ ₂ and Nb35 was
537 added to the SEC purified OR51E2-miniG_{s399} followed by overnight incubation on ice. The
538 sample was concentrated on a 50 kDa MWCO spin filter (Amicon), and injected onto a

539 Superdex 200 Increase 10/300 GL SEC column, equilibrated with 20 mM HEPES, pH 7.50, 150
540 mM NaCl, 0.0075% (w/v) L-MNG, 0.0025% (w/v) GDN, 0.001% (w/v) CHS, and 30 mM sodium
541 propionate. Fractions containing the monomeric OR51E2-G_s complex were concentrated on a
542 100 kDa MWCO spin filter immediately prior to cryo-EM grid preparation.

543

544 **Cryo-EM vitrification, data collection, and processing**

545 2.75 μ L of purified OR51E2-G_s complex was applied to glow discharged 300 mesh R1.2/1.3
546 UltraAuFoil Holey gold support films (Quantifoil). Support films were plunge-frozen in liquid
547 ethane using a Vitrobot Mark IV (Thermo Fisher) with a 10 s hold period, blot force of 0, and
548 blotting time varying between 1-5 s while maintaining 100% humidity and 4°C. Vitrified grids
549 were clipped with Autogrid sample carrier assemblies (Thermo Fisher) immediately prior to
550 imaging. Movies of OR51E2-G_s embedded in ice were recorded using a Titan Krios Gi3
551 (Thermo Fisher) with a BioQuantum Energy Filter (Gatan) and a K3 Direct Electron Detector
552 (Gatan). Data were collected using Serial EM⁴² running a 3 x 3 image shift pattern at 0° stage
553 tilt. A nominal magnification of 105,000 x with a 100 μ m objective was used in super-resolution
554 mode with a physical pixel size of 0.81 \AA pixel⁻¹. Movies were recorded using dose fractionated
555 illumination with a total exposure of 50 e⁻ \AA^{-2} over 60 frames yielding 0.833 e⁻ \AA^{-2} frames⁻¹.

556

557 16,113 super-resolution movies were motion-corrected and Fourier cropped to physical pixel
558 size using UCSF MotionCor2⁴³. Dose-weighted micrographs were imported into cryoSPARC
559 v3.2 (Structura Biotechnology), and contrast transfer functions were calculated using the Patch
560 CTF Estimation tool. A threshold of CTF fit resolution > 5 \AA was used to exclude low quality
561 micrographs. Particles were template picked using a 20 \AA low-pass filtered model that was
562 generated *ab initio* from data collected during an earlier 200 kV screening session. 8,884,130
563 particles were extracted with a box size of 288 pixels binned to 72 pixels and sorted with the
564 Heterogeneous Refinement tool, which served as 3D classification with alignment. Template
565 volumes for each of the four classes were low-pass filtered to 20 \AA and comprised an initial
566 OR51E2-G_s volume as well as three scrambled volumes obtained by terminating the Ab-Initio
567 Reconstruction tool before the first iteration. The resulting 1,445,818 particles were re-extracted
568 with a box size of 288 pixels binned to 144 pixels and sorted by an additional round of
569 Heterogeneous Refinement using two identical initial models and two scrambled models.
570 776,527 particles from the highest resolution reconstruction were extracted with an unbinned
571 box size of 288 pixels, and were subjected to Homogeneous Refinement followed by Non-
572 Uniform Refinement. Particles were exported using `csparc2star.py` from the `pyem v0.5` script

573 package³⁹, and an inclusion mask covering the 7TM domain of OR51E2 was generated using
574 the Segger tool in UCSF ChimeraX⁴⁴ and the mask.py tool in pyem v0.5. Particles and mask
575 were imported into Relion v3.0⁴⁵ and sorted by several rounds of 3D classification without image
576 alignment, where the number of classes and tau factor were allowed to vary. The resulting
577 204,438 particles were brought back into cryoSPARC and subjected to Non-Uniform
578 Refinement. Finally, Local Refinement using an inclusion mask covering the 7TM domain was
579 performed, using poses/shift gaussian priors with S.D. of rotational and shift magnitudes limited
580 to 3° and 2 Å respectively.

581

582 **Model building and refinement**

583 Model building and refinement were carried out using an Alphafold2³⁸ predicted structure as a
584 starting model, which was fitted into the OR51E2-G_s map using UCSF ChimeraX. A draft model
585 was generated using ISOLDE⁴⁶ and was further refined by iterations of real space refinement in
586 Phenix⁴⁷ and manual refinement in Coot⁴⁸. The propionate model and rotamer library were
587 generated with the PRODRG server⁴⁹, docked using Coot, and refined in Phenix. Final map-
588 model validations were carried out using Molprobtity and EMRinger in Phenix.

589

590 **Site Directed Mutagenesis**

591 Generation of OR51E2 mutants was performed as described previously⁵⁰. Forward and reverse
592 primers coding for the mutation of interest were obtained from Integrated DNA Technologies.
593 Two successive rounds of PCR using Phusion polymerase (Thermo Fisher Scientific: F-549L)
594 were performed to amplify ORs with mutations. The first round of PCR generated two
595 fragments, one containing the 5' region upstream of the mutation site and the other the 3'
596 downstream region. The second PCR amplification joined these two fragments to produce a full
597 ORF of the olfactory receptor. PCR products with desired length were gel purified and cloned
598 into the MluI and NotI sites of mammalian expression vector pCI (Promega) that contains rho-
599 tag. Plasmids were purified using the Thomas Scientific (1158P42) miniprep kit with modified
600 protocol including phenol-chloroform extraction before column purification.

601

602 **cAMP signaling assays**

603 The GloSensor cAMP assay (Promega) was used to determine real-time cAMP levels
604 downstream of OR activation in HEK293T cells, as previously described⁵¹. HEK293T cells were
605 cultured in Minimum Essential Media (MEM, Corning) supplemented by 10 % Fetal Bovine
606 Serum (FBS, Gibco), 0.5 % Penicillin-Streptomycin (Gibco) and 0.5 % Amphotericin B (Gibco).

607 Cultured HEK293T cells were plated the day before transfection at 1/10 of 100 % confluence
608 from a 100 mm plate into 96-well plates coated with poly D lysine (Corning). For each 96-well
609 plate, 10 µg pGloSensor-20F plasmid (Promega) and 75 µg of Rho-tagged OR in the pCI
610 mammalian expression vector (Promega) were transfected 18 to 24 h before odorant stimulation
611 using Lipofectamine 2000 (Invitrogen: 11668019) in MEM supplemented by 10% FBS. On
612 stimulation day, plates were injected with 25 µl of GloSensor substrate (Promega) and
613 incubated for 2 hours in the dark at room temperature and in a odor-free environment. Odorants
614 were diluted to the desired concentration in CD293 media (Gibco) supplemented with copper
615 (30 µM CuCl₂, Sigma-Aldrich) and 2 mM L-glutamine (Gibco) and pH adjusted to 7.0 with a 150
616 mM solution of sodium hydroxide (Sigma-Aldrich). After injecting 25 µl of odorants in CD293
617 media into each well, GloSensor luminescence was immediately recorded for 20 cycles of
618 monitoring over a total period of 30 minutes using a BMG Labtech POLARStar Optima plate
619 reader. The resulting luminescence activity was normalized to a vector control lacking any OR,
620 and the OR response was obtained by summing the response from all 20 cycles to determine
621 an area under the curve (AUC). Dose-dependent responses of ORs were analyzed by fitting a
622 least squares function to the data using GraphPrism 9.

623

624 **Evaluating Cell Surface Expression**

625 Flow-cytometry was used to evaluate cell surface expression of olfactory receptors as described
626 previously⁵². HEK293T cells were seeded onto 35-mm plates (Greiner Bio-One) with
627 approximately 3.5×10^5 cells (25 % confluency). The cells were cultured overnight. After 18 to
628 24 hours, 1200 ng of ORs tagged with the first 20 amino acids of human rhodopsin (rho-tag) at
629 the N-terminal ends⁵³ in pCI (Promega) and 30 ng eGFP were transfected using Lipofectamine
630 2000 (Invitrogen: 11668019). 18 to 24 hours after transfection, the cells were detached and
631 resuspended using Cell stripper (Corning) and then transferred into 5 mL round bottom
632 polystyrene (PS) tubes (Falcon) on ice. The cells were spun down at 4°C and resuspended in
633 phosphate buffered saline (PBS, Gibco) containing 15 mM NaN₃ (Sigma-Aldrich) and 2% FBS
634 (Gibco). They were stained with 1/400 (v/v) of primary antibody mouse anti rhodopsin clone 4D2
635 (Sigma-Aldrich: MABN15) and allowed to incubate for 30 minutes then washed with PBS
636 containing 15 mM NaN₃ and 2% FBS. The cells were spun again and then stained with 1/200
637 (v/v) of the phycoerythrin (PE)-conjugated donkey anti-mouse F(ab')₂ fragment antibody
638 (Jackson Immunologicals: 715-116-150) and allowed to incubate for 30 minutes in the dark. To
639 label dead cells, 1/500 (v/v) of 7-Amino-actinomycin D (Calbiochem: 129935) was added. The
640 cells were then immediately analyzed using a BD FACSCanto II flow cytometer with gating

641 allowing for GFP positive, single, spherical, viable cells and the measured PE fluorescence
642 intensities were analyzed and visualized using Flowjo v10.8.1. Normalizing the cell surface
643 expression levels of the OR51E2 mutants was performed using wild-type OR51E2 which
644 showed robust cell surface expression and empty plasmid pCI which demonstrated no
645 detectable cell surface expression.

646

647 **Molecular dynamics simulations**

648 All MD simulations were performed using the GROMACS package⁵⁴ (version 2021) with the
649 CHARMM36m forcefield⁵⁵ starting from the OR51E2 EM structure with and without propionate.
650 The G protein was removed in all these simulations. The GPCR structures were prepared by
651 Maestro “protein preparation wizard” module⁵⁶. The missing side chains and hydrogen atoms
652 were added. Furthermore, protein chain termini were capped with neutral acetyl and
653 methylamide groups, and histidine protonated states were assigned, after which minimization
654 was performed. The simulation box was created using CHARMM-GUI⁵⁷. We used the PPM 2.0
655 function of OPM (Orientation of proteins in membranes) structure of OR51E2 for alignment of
656 the transmembrane helices of protein structure and inserted into a 75% palmitoyl-oleoyl-
657 phosphatidylcholine (POPC) / 25% Cholesteryl Hemi Succinate deprotonated (CHSD) bilayer.
658 The CHSDs were placed around the GPCR structure. TIP3P water model was used for
659 solvation and 0.15 M potassium chloride ions were added for neutralization. The final system
660 dimensions were about 85 × 85 × 115 Å. The system was minimized with position restraints (10
661 kcal/mol/Å²) on all heavy atoms of GPCR and ligand, followed by a 1 ns heating step which
662 raise the temperature from 0K to 310K in NVT ensemble with Nosé-Hoover thermostat⁵⁸. Then
663 we performed a single long equilibration for lipid and solvent (1000 ns) in NPT ensemble. During
664 the heating step and the long equilibration, position restraints were placed of 10 kcal/mol-Å²
665 applied on the receptor, propionate and POPC/CHSD for the first 1 ns. Later, the restraint on
666 lipids was reduced from 5 kcal/mol-Å² to 0 kcal/mol-Å² in steps of 1 kcal with 5 ns of simulations
667 per step. Then the POPC/CHSD were allowed to freely move during the rest of the long
668 equilibration and the final snapshot was used as the initial conformation for equilibrating the
669 protein and ligand. The position restraints were applied on the protein (backbone and side
670 chain) and ligand starting at 5 kcal/mol-Å² reducing to 0 kcal/mol-Å² in steps of 1 kcal/mol-Å²
671 with 5 ns of simulation per step. The last snapshot of the equilibration step was used as initial
672 conformation for five production runs with random seeds. This snapshot was also used as
673 reference conformation for all the RMSD in coordinates. The pressure was controlled using
674 Parrinello-Rahman method⁵⁹ and the simulation system was coupled to 1 bar pressure bath. In

675 all simulations LINCS algorithm is applied on all bonds and angles of waters with 2 fs time step
676 used for integration. We used a cut-off of 12 Å for non-bond interaction and particle mesh Ewald
677 method⁶⁰ to treat long range L-J interaction. The MD snapshots were stored at every 20 ps
678 interval. Trajectories were visualized with VMD and PyMOL (Molecular Graphics System,
679 Version 2.5 Schrödinger) and analyzed using the GROMACS package (version 2016/2019). All
680 MD analysis was done on the aggregated trajectories from the 5 runs (total $5 \times 1 \mu\text{s} = 5 \mu\text{s}$).
681 Heatmaps and other MD related plots were generated with Graphpad Prism 9, whereas
682 structural figures were generated using PyMOL.

683

684 **Molecular dynamics analysis**

685 *Ligand-receptor and intramolecular interactions*

686 Contact frequencies were calculated using the “get_contacts” module
687 (<https://getcontacts.github.io/>). The following interaction types were calculated between ligand
688 and receptor: hydrogen bonds, hydrophobic and van der Waals interactions.

689

690 *Calculation of Residue Distances*

691 For the distance between two residues, we used *gmx mindist* (GROMACS package 2016/2019),
692 which calculates the minimal distance between two atoms (e.g., sidechain, Ca, oxygens,
693 nitrogens) of one of each residue over time. Distance analysis on the static structures were
694 done using the measurement tool in PyMOL. Chosen atoms for distance calculations are
695 described in each legend.

696

697 *Rotamer Analysis of F250*

698 For the rotamer analysis of residues of interest, we used the VMD tcl script
699 “Calculate_dihedrals” (https://github.com/ajasja/calculate_dihedrals).

700

701 *Conformational Clustering*

702 To select representative snapshots from MD simulations that are shown in Fig. 4i and Fig. 5b,
703 we clustered (*gmx cluster*, GROMACS package 2016/2019) the aggregated trajectories using
704 transmembrane helix backbone atoms. An RMSD cutoff for clustering was set at 0.08 nm for
705 propionate-bound simulations, 0.085 nm for no-ligand WT simulations and 0.085 nm for no-
706 ligand Q181D simulations. For propionate clustering (Fig. 2f), we used an RMSD cut-off of 0.01
707 nm for the ligand.

708

709 *Root-Mean-Square Deviation (RMSD)*

710 The *gmx rms* (GROMACS package 2016/2019) function was used to determine whether
711 simulations were stable. For this we used the transmembrane backbone of OR51E2 by selecting
712 the following residues: 23-50 (TM1), 57-86 (TM2), 93-126 (TM3), 137-164 (TM4), 191-226
713 (TM5), 230-264 (TM6), and 269-294 (TM7). As reference, we used the equilibrated MD
714 structure of propionate bound, apo and Q181D OR51E2. In order to assess the stability of the
715 ligand in the binding pocket over time, the RMSD of propionate was calculated using the
716 equilibrated MD structure of propionate-bound as a reference.

717

718 *Volume of the ligand binding pocket*

719 The volume and surface area of the propionate binding pocket in OR51E2 was calculated using
720 the Maestro SiteMap module^{61,62}. Three structures were used for the volume calculation: 1) the
721 OR51E2 cryo-EM structure bound to propionate, 2) the OR51E2-L158A model bound to
722 hexanoate, 3) the OR51E2-F155A model bound to octanoate. To prepare the L158A and F155A
723 models we used the Maestro mutation function to introduce the substitutions onto the cryo-EM
724 structure of OR51E2; these models were then energy minimized using the
725 ProteinPreparationWizard module using default parameters⁵⁶. We then used Maestro Glide
726 Docking⁶³⁻⁶⁵ to dock hexanoate and octanoate into the resulting models of OR51E2-L158A and
727 OR51E2-F155A, respectively. We prepared the docking grid box for both OR51E2-L158A and
728 OR51E2-F155A by defining a box centered at propionate, with a box length of 2.5 nm. Glide
729 ligand docking was performed using XP precision and default parameters to yield a model for
730 OR51E2-L158A bound to hexanoate and OR51E2-F155A bound to octanoate. To calculate
731 ligand binding site volumes using the SiteMap module, we defined the ligand binding pocket as
732 a maximum of 0.6 nm around selected ligand (propionate/hexanoate/octanoate) with at least 15
733 site points (probes) per reported site. The grid size for the probes was set to 0.035 nm. Using
734 this approach, the calculated volumes for wild-type OR51E2, OR51E2-L158A, and OR51E2-
735 F155A were 31 Å³, 68 Å³, and 90 Å³, respectively.

736

737 **Phylogenetic tree**

738 A phylogenetic tree of human Class A GPCRs was made by analyzing 677 sequences. Of
739 these, 390 sequences were from olfactory receptors (56 Class I ORs and 334 Class II ORs),
740 while 287 were from non-olfactory Class A GPCRs. Sequences were aligned with Clustal⁶⁶ on
741 Jalview 2.11.2.5⁶⁷. On R studio 202.07.01, alignment reading and matrix of distance between
742 sequences (by sequence identity) calculation were performed with the Biostrings⁶⁸ and seqinr⁶⁹

743 packages. Neighbor-Joining tree and tree visualization were realized with packages ape⁷⁰ and
744 ggtree⁷¹ and the tree is plotted unrooted with the daylight method.

745

746 **Data Availability**

747 Coordinates for propionate OR51E2-G_s have been deposited in the RCSB PDB under
748 accession code 8F76 . EM density maps for OR51E2-G_s and the 7TM domain of OR51E2 have
749 been deposited in the Electron Microscopy Data Bank under accession codes EMD-28896, and
750 EMD-28900, respectively. The MD simulation trajectories have been deposited in the GPCRMD
751 database.

752

753 **Acknowledgements**

754 We thank Dan Toso at Cal-Cryo at QB3-Berkeley for help in microscope operation and data
755 collection. H.M., C.A.D.M. and J.T. thank Mengjue J. Ni and Hsiu-Yi Lu for their technical
756 support. This work was supported by the National Institutes of Health (NIH) grant
757 R01DC020353 (H.M., N.V., and A.M.) and K99DC018333 (C.A.D.M.). Cryo-EM equipment at
758 UCSF is partially supported by NIH grants S10OD020054 and S10OD021741. This project was
759 funded by the UCSF Program for Breakthrough Biomedical Research, funded in part by the
760 Sandler Foundation. A.M. acknowledges support from the Edward Mallinckrodt, Jr. Foundation
761 and the Vallee Foundation.

762

763 **Contributions**

764 C.B.B., C.A.D.M., W.J.C.v.d.V., N.V., H.M., and A.M. designed the study. C.B.B. cloned
765 constructs, prepared baculoviruses, expressed and purified G protein complexing reagents, and
766 optimized large scale production of OR51E2. C.B.B. worked out conditions to biochemically
767 purify and stabilize the propionate-bound OR51E2-G_s complex, and identified optimal cryo-EM
768 grid preparation procedures following screening, collection, and processing of 200 kV cryo-EM
769 data. B.F. and A.M. performed 300 kV cryo-EM data collection. C.B.B. determined high-
770 resolution cryo-EM maps by extensive image processing with input from A.M. A.M. and C.B.B.
771 built, refined models of propionate-bound OR51E2 in complex with G_s and Nb35. C.B.B. and
772 A.M. analyzed cryo-EM data and models and prepared figures and tables. C.A.D.M. and J.T.
773 analyzed OR models and sequences to design and clone OR mutants, performed Glosensor
774 signaling experiments for OR functional activity, and generated OR cell surface expression data
775 by flow cytometry with input from H.M. C.A.D.M and J.T. analyzed and prepared figures and
776 tables for signaling and flow cytometry data. C.A.D.M. built the phylogenetic tree of ORs and

777 non-olfactory Class A GPCRs. N.M. set up molecular dynamics simulations, ligand docking, and
778 performed binding pocket volume calculations. W.J.C.v.d.V. analyzed simulation trajectories
779 and prepared figures describing simulation data. W.J.C.v.d.V., N.M. and N.V. provided
780 mechanistic insight from simulation data. C.L.D.T. performed bioinformatic analysis of OR and
781 non-olfactory Class A GPCR conservation. L.L. and C.B.B. performed pilot GloSensor signaling
782 studies in suspension cells. C.B.B., C.A.D.M., and A.M. wrote an initial draft of the manuscript
783 and generated figures with contributions from all authors. Further edits to the manuscript were
784 provided by W.J.C.v.d.V., N.M., V.N., and H.M. The overall project was supervised and funded
785 by N.V., H.M., and A.M.

786

787 **Competing Interests**

788 H.M. has received royalties from Chemcom, research grants from Givaudan, and consultant
789 fees from Kao.

790 References

- 791 1 Buck L, Axel R. A novel multigene family may encode odorant receptors: a molecular basis
792 for odor recognition. *Cell* 1991; **65**: 175–187.
- 793 2 Malnic B, Hirono J, Sato T, Buck LB. Combinatorial receptor codes for odors. *Cell* 1999; **96**:
794 713–723.
- 795 3 Zhao H, Ivic L, Otaki JM, Hashimoto M, Mikoshiba K, Firestein S. Functional expression of
796 a mammalian odorant receptor. *Science* 1998; **279**: 237–242.
- 797 4 Mayhew EJ, Arayata CJ, Gerkin RC, Lee BK, Magill JM, Snyder LL *et al.* Transport features
798 predict if a molecule is odorous. *Proc Natl Acad Sci U S A* 2022; **119**: e2116576119.
- 799 5 Niimura Y, Matsui A, Touhara K. Extreme expansion of the olfactory receptor gene
800 repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13
801 placental mammals. *Genome Res* 2014; **24**: 1485–1496.
- 802 6 Malnic B, Godfrey PA, Buck LB. The human olfactory receptor gene family. *Proc Natl Acad*
803 *Sci U S A* 2004; **101**: 2584–2589.
- 804 7 Glusman G, Yanai I, Rubin I, Lancet D. The complete human olfactory subgenome.
805 *Genome Res* 2001; **11**: 685–702.
- 806 8 Jones DT, Reed RR. Golf: an olfactory neuron specific-G protein involved in odorant signal
807 transduction. *Science* 1989; **244**: 790–795.
- 808 9 Pourmorady A, Lomvardas S. Olfactory receptor choice: a case study for gene regulation in
809 a multi-enhancer system. *Curr Opin Genet Dev* 2022; **72**: 101–109.
- 810 10 Butterwick JA, Del Marmol J, Kim KH, Kahlson MA, Rogow JA, Walz T *et al.* Cryo-EM
811 structure of the insect olfactory receptor Orco. *Nature* 2018; **560**: 447–452.
- 812 11 Del Marmol J, Yedlin MA, Ruta V. The structural basis of odorant recognition in insect
813 olfactory receptors. *Nature* 2021; **597**: 126–131.
- 814 12 Ikegami K, de March CA, Nagai MH, Ghosh S, Do M, Sharma R *et al.* Structural instability
815 and divergence from conserved residues underlie intracellular retention of mammalian
816 odorant receptors. *Proc Natl Acad Sci U S A* 2020; **117**: 2957–2967.
- 817 13 Saito H, Kubota M, Roberts RW, Chi Q, Matsunami H. RTP family members induce
818 functional expression of mammalian odorant receptors. *Cell* 2004; **119**: 679–691.
- 819 14 Cook BL, Steuerwald D, Kaiser L, Graveland-Bikker J, Vanberghem M, Berke AP *et al.*
820 Large-scale production and study of a synthetic G protein-coupled receptor: human
821 olfactory receptor 17-4. *Proc Natl Acad Sci U S A* 2009; **106**: 11925–11930.
- 822 15 Katada S, Tanaka M, Touhara K. Structural determinants for membrane trafficking and G
823 protein selectivity of a mouse olfactory receptor. *J Neurochem* 2004; **90**: 1453–1463.
- 824 16 Lee S-J, Depoortere I, Hatt H. Therapeutic potential of ectopic olfactory and taste
825 receptors. *Nat Rev Drug Discov* 2019; **18**: 116–138.

- 826 17 Freitag J, Ludwig G, Andreini I, Rössler P, Breer H. Olfactory receptors in aquatic and
827 terrestrial vertebrates. *J Comp Physiol A* 1998; **183**: 635–650.
- 828 18 Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD. Odor coding by a Mammalian
829 receptor repertoire. *Sci Signal* 2009; **2**: ra9.
- 830 19 Xu LL, Stackhouse BG, Florence K, Zhang W, Shanmugam N, Sesterhenn IA *et al.* PSGR,
831 a novel prostate-specific gene with homology to a G protein-coupled receptor, is
832 overexpressed in prostate cancer. *Cancer Res* 2000; **60**: 6568–6572.
- 833 20 Gelis L, Jovancevic N, Veitinger S, Mandal B, Arndt H-D, Neuhaus EM *et al.* Functional
834 Characterization of the Odorant Receptor 51E2 in Human Melanocytes. *J Biol Chem* 2016;
835 **291**: 17772–17786.
- 836 21 Kotlo K, Anbazhagan AN, Priyamvada S, Jayawardena D, Kumar A, Chen Y *et al.* The
837 olfactory G protein-coupled receptor (Olf-78/OR51E2) modulates the intestinal response to
838 colitis. *Am J Physiol Cell Physiol* 2020; **318**: C502–C513.
- 839 22 Vadevoo SMP, Gunassekaran GR, Lee C, Lee N, Lee J, Chae S *et al.* The macrophage
840 odorant receptor Olf78 mediates the lactate-induced M2 phenotype of tumor-associated
841 macrophages. *Proc Natl Acad Sci U S A* 2021; **118**. doi:10.1073/pnas.2102434118.
- 842 23 Pluznick JL, Protzko RJ, Gevorgyan H, Peterlin Z, Sipos A, Han J *et al.* Olfactory receptor
843 responding to gut microbiota-derived signals plays a role in renin secretion and blood
844 pressure regulation. *Proc Natl Acad Sci U S A* 2013; **110**: 4410–4415.
- 845 24 Flegel C, Manteniots S, Osthold S, Hatt H, Gisselmann G. Expression profile of ectopic
846 olfactory receptors determined by deep sequencing. *PLoS One* 2013; **8**: e55368.
- 847 25 Nakashima A, Takeuchi H, Imai T, Saito H, Kiyonari H, Abe T *et al.* Agonist-independent
848 GPCR activity regulates anterior-posterior targeting of olfactory sensory neurons. *Cell*
849 2013; **154**: 1314–1325.
- 850 26 Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS *et al.* Crystal
851 structure of the β 2 adrenergic receptor-Gs protein complex. *Nature* 2011; **477**: 549–555.
- 852 27 Nehmé R, Carpenter B, Singhal A, Strege A, Edwards PC, White CF *et al.* Mini-G proteins:
853 Novel tools for studying GPCRs in their active conformation. *PLoS One* 2017; **12**:
854 e0175642.
- 855 28 Ring AM, Manglik A, Kruse AC, Enos MD, Weis WI, Garcia KC *et al.* Adrenaline-activated
856 structure of β 2-adrenoceptor stabilized by an engineered nanobody. *Nature* 2013; **502**:
857 575–579.
- 858 29 Tsai C-J, Pamula F, Nehmé R, Mühle J, Weinert T, Flock T *et al.* Crystal structure of
859 rhodopsin in complex with a mini-Go sheds light on the principles of G protein selectivity.
860 *Sci Adv* 2018; **4**: eaat7052.
- 861 30 Ballesteros JA, Weinstein H. [19] Integrated methods for the construction of three-
862 dimensional models and computational probing of structure-function relations in G protein-
863 coupled receptors. In: Sealfon SC (ed). *Methods in Neurosciences*. Academic Press, 1995,
864 pp 366–428.

- 865 31 de March CA, Kim S-K, Antonczak S, Goddard WA 3rd, Golebiowski J. G protein-coupled
866 odorant receptors: From sequence to structure. *Protein Sci* 2015; **24**: 1543–1548.
- 867 32 Cichy A, Shah A, Dewan A, Kaye S, Bozza T. Genetic Depletion of Class I Odorant
868 Receptors Impacts Perception of Carboxylic Acids. *Curr Biol* 2019; **29**: 2687–2697.e4.
- 869 33 Pronin A, Slepak V. Ectopically expressed olfactory receptors OR51E1 and OR51E2
870 suppress proliferation and promote cell death in a prostate cancer cell line. *J Biol Chem*
871 2021; **296**: 100475.
- 872 34 Manglik A, Kruse AC. Structural Basis for G Protein-Coupled Receptor Activation.
873 *Biochemistry* 2017; **56**: 5628–5634.
- 874 35 de March CA, Yu Y, Ni MJ, Adipietro KA, Matsunami H, Ma M *et al.* Conserved Residues
875 Control Activation of Mammalian G Protein-Coupled Odorant Receptors. *J Am Chem Soc*
876 2015; **137**: 8611–8616.
- 877 36 Bushdid C, de March CA, Topin J, Do M, Matsunami H, Golebiowski J. Mammalian class I
878 odorant receptors exhibit a conserved vestibular-binding pocket. *Cell Mol Life Sci* 2019; **76**:
879 995–1004.
- 880 37 Shim T, Pacalon J, Kim W-C, Cong X, Topin J, Golebiowski J *et al.* The Third Extracellular
881 Loop of Mammalian Odorant Receptors Is Involved in Ligand Binding. *Int J Mol Sci* 2022;
882 **23**. doi:10.3390/ijms232012501.
- 883 38 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O *et al.* Highly accurate
884 protein structure prediction with AlphaFold. *Nature* 2021; **596**: 583–589.
- 885 39 Asarnow D, Palovcak E, Cheng Y. *asarnow/pyem: UCSF pyem v0.5*. 2019
886 doi:10.5281/zenodo.3576630.
- 887 40 Dang S, Feng S, Tien J, Peters CJ, Bulkley D, Lolicato M *et al.* Cryo-EM structures of the
888 TMEM16A calcium-activated chloride channel. *Nature* 2017; **552**: 426–429.
- 889 41 Staus DP, Wingler LM, Choi M, Pani B, Manglik A, Kruse AC *et al.* Sortase ligation enables
890 homogeneous GPCR phosphorylation to reveal diversity in β -arrestin coupling. *Proc Natl*
891 *Acad Sci U S A* 2018; **115**: 3834–3839.
- 892 42 Mastronarde DN. SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai
893 Microscopes Using Prediction of Specimen Position. *Microsc Microanal* 2003; **9**: 1182–
894 1183.
- 895 43 Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA. MotionCor2:
896 anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat*
897 *Methods* 2017; **14**: 331–332.
- 898 44 Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI *et al.* UCSF
899 ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*
900 2021; **30**: 70–82.
- 901 45 Scheres SHW. RELION: implementation of a Bayesian approach to cryo-EM structure
902 determination. *J Struct Biol* 2012; **180**: 519–530.

- 903 46 Croll TI. ISOLDE: a physically realistic environment for model building into low-resolution
904 electron-density maps. *Acta Crystallogr D Struct Biol* 2018; **74**: 519–530.
- 905 47 Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N *et al.* PHENIX: a
906 comprehensive Python-based system for macromolecular structure solution. *Acta*
907 *Crystallogr D Biol Crystallogr* 2010; **66**: 213–221.
- 908 48 Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D*
909 *Biol Crystallogr* 2004; **60**: 2126–2132.
- 910 49 Schüttelkopf AW, van Aalten DMF. PRODRG: a tool for high-throughput crystallography of
911 protein–ligand complexes. *Acta Crystallogr D Biol Crystallogr* 2004; **60**: 1355–1363.
- 912 50 Bushdid C, de March CA, Matsunami H, Golebiowski J. Numerical Models and In Vitro
913 Assays to Study Odorant Receptors. *Methods Mol Biol* 2018; **1820**: 77–93.
- 914 51 Zhang Y, Pan Y, Matsunami H, Zhuang H. Live-cell Measurement of Odorant Receptor
915 Activation Using a Real-time cAMP Assay. *J Vis Exp* 2017. doi:10.3791/55831.
- 916 52 Zhuang H, Matsunami H. Evaluating cell-surface expression and measuring activation of
917 mammalian odorant receptors in heterologous cells. *Nat Protoc* 2008; **3**: 1402–1413.
- 918 53 Krautwurst D, Yau KW, Reed RR. Identification of ligands for olfactory receptors by
919 functional expression of a receptor library. *Cell* 1998; **95**: 917–926.
- 920 54 Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel
921 molecular dynamics implementation. *Comput Phys Commun* 1995; **91**: 43–56.
- 922 55 Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL *et al.* CHARMM36m: an
923 improved force field for folded and intrinsically disordered proteins. *Nat Methods* 2017; **14**:
924 71–73.
- 925 56 Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand
926 preparation: parameters, protocols, and influence on virtual screening enrichments. *J*
927 *Comput Aided Mol Des* 2013; **27**: 221–234.
- 928 57 Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for
929 CHARMM. *J Comput Chem* 2008; **29**: 1859–1865.
- 930 58 Evans DJ, Holian BL. The Nose–Hoover thermostat. *J Chem Phys* 1985; **83**: 4069–4074.
- 931 59 Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular
932 dynamics method. *J Appl Phys* 1981; **52**: 7182–7190.
- 933 60 Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums
934 in large systems. *J Chem Phys* 1993; **98**: 10089–10092.
- 935 61 Halgren T. New method for fast and accurate binding-site identification and analysis. *Chem*
936 *Biol Drug Des* 2007; **69**: 146–148.
- 937 62 Halgren TA. Identifying and characterizing binding sites and assessing druggability. *J Chem*
938 *Inf Model* 2009; **49**: 377–389.

- 939 63 Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA *et al.* Extra
940 precision glide: docking and scoring incorporating a model of hydrophobic enclosure for
941 protein-ligand complexes. *J Med Chem* 2006; **49**: 6177–6196.
- 942 64 Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Thomas Pollard W *et al.* Glide: A
943 New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in
944 Database Screening. *Journal of Medicinal Chemistry*. 2004; **47**: 1750–1759.
- 945 65 Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT *et al.* Glide: A New
946 Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking
947 Accuracy. *Journal of Medicinal Chemistry*. 2004; **47**: 1739–1749.
- 948 66 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al.*
949 Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; **23**: 2947–2948.
- 950 67 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a
951 multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009; **25**:
952 1189–1191.
- 953 68 Pagès H, Aboyou P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of
954 biological 370 strings. 10.18129/B9.bioc.2022.
- 955 69 Charif D, Lobry JR. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical
956 Computing Devoted to Biological Sequences Retrieval and Analysis. In: Bastolla U, Porto
957 M, Roman HE, Vendruscolo M (eds). *Structural Approaches to Sequence Evolution:
958 Molecules, Networks, Populations*. Springer Berlin Heidelberg: Berlin, Heidelberg, 2007, pp
959 207–232.
- 960 70 Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary
961 analyses in R. *Bioinformatics* 2019; **35**: 526–528.
- 962 71 Xu S, Li L, Luo X, Chen M, Tang W, Zhan L *et al.* Ggtree : A serialized data object for
963 visualization of a phylogenetic tree and annotation data. *iMeta* 2022. doi:10.1002/imt2.56.
- 964