

Supporting Information

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

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.

* Correspondence:

a.dipizio.leibniz-lsb@tum.de; Tel.: +49 8161716516

d.krautwurst.leibniz-lsb@tum.de; Tel.: +49 8161712634

Figure S1. Multiple Sequence Alignment of OR5K1 (residues 20-292) to β 2-adrenergic receptor (PDB ID: 6MXT), A2A receptor (PDB ID: 2YDV), Rhodopsin receptor (PDB ID: 4X1H), and CXCR4 receptor (PDB ID: 3ODU).

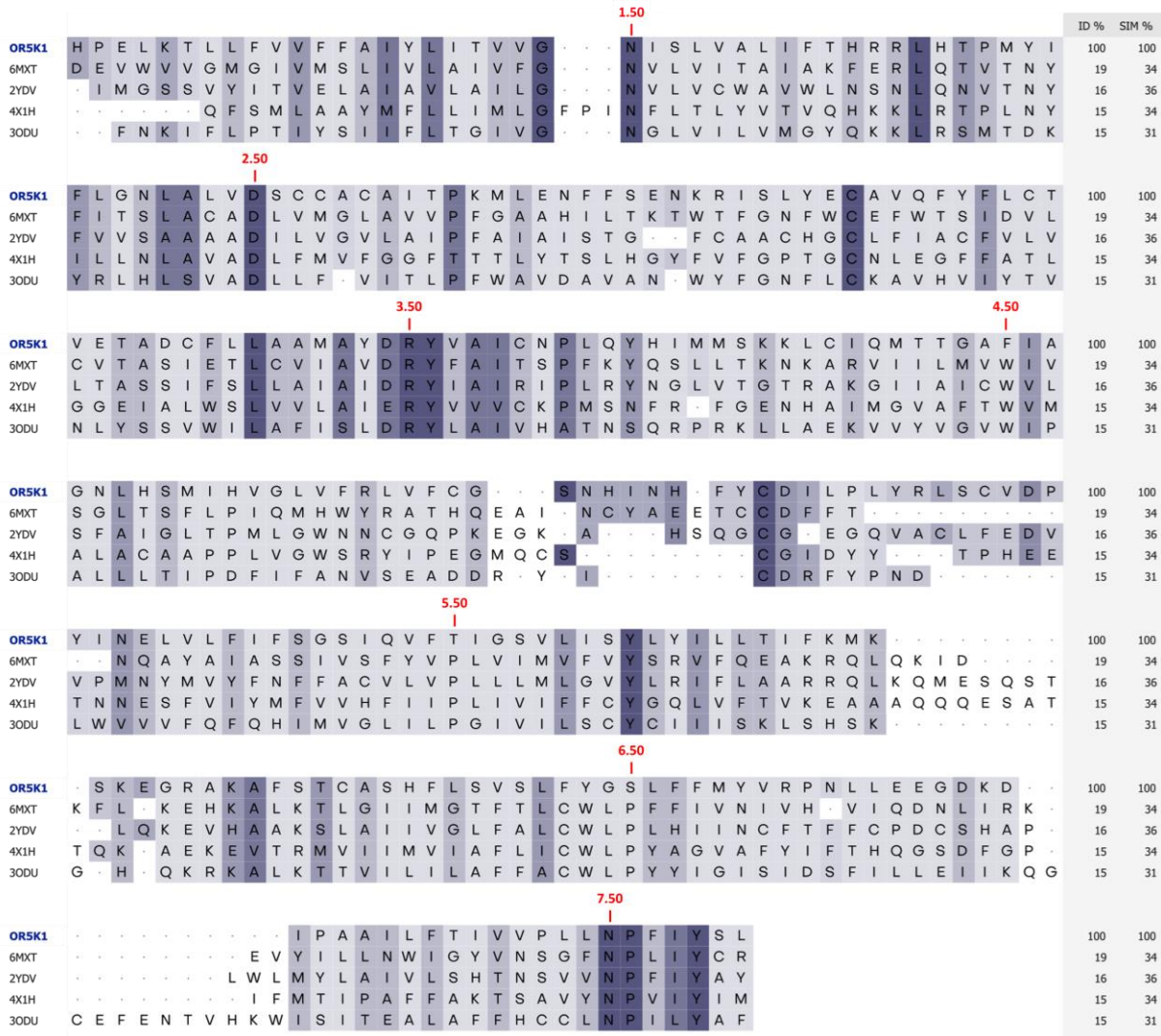


Figure S2. (a) Length of ECL2 for OR5K1 and experimental class A GPCRs. **(b)** Sequence Alignment of OR5K1 ECL2 region (S157^{4.57}-L188^{5.37}) to templates used for its modeling: NPY2 (PDB ID: 7DDZ) and CCK1 (PDB ID: 7MBY) before-Cys^{45.50} segment and Apelin (PDB ID: 6KNM) after-Cys^{45.50} segment.

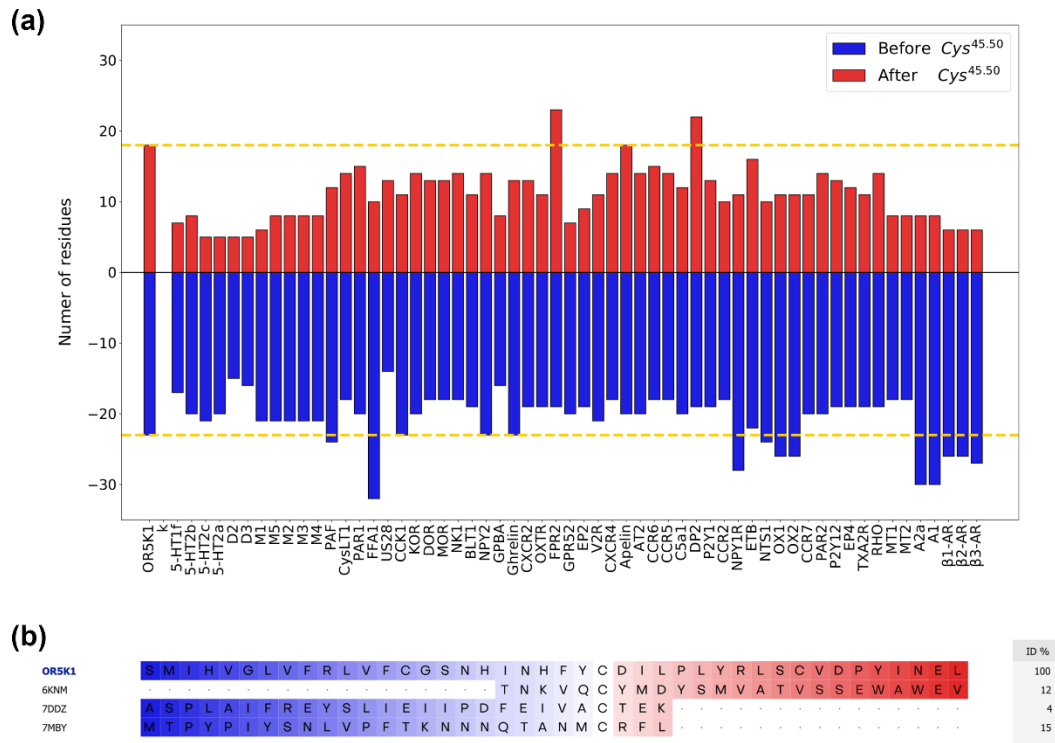


Figure S3. OR5K1 models built with AlphaFold 2 (a) and homology modeling (b). Alphafold2 model (AF2) is colored by pLDDT confidence score, while OR5K1 model predicted with homology modeling (HM) is colored by templates coverage.

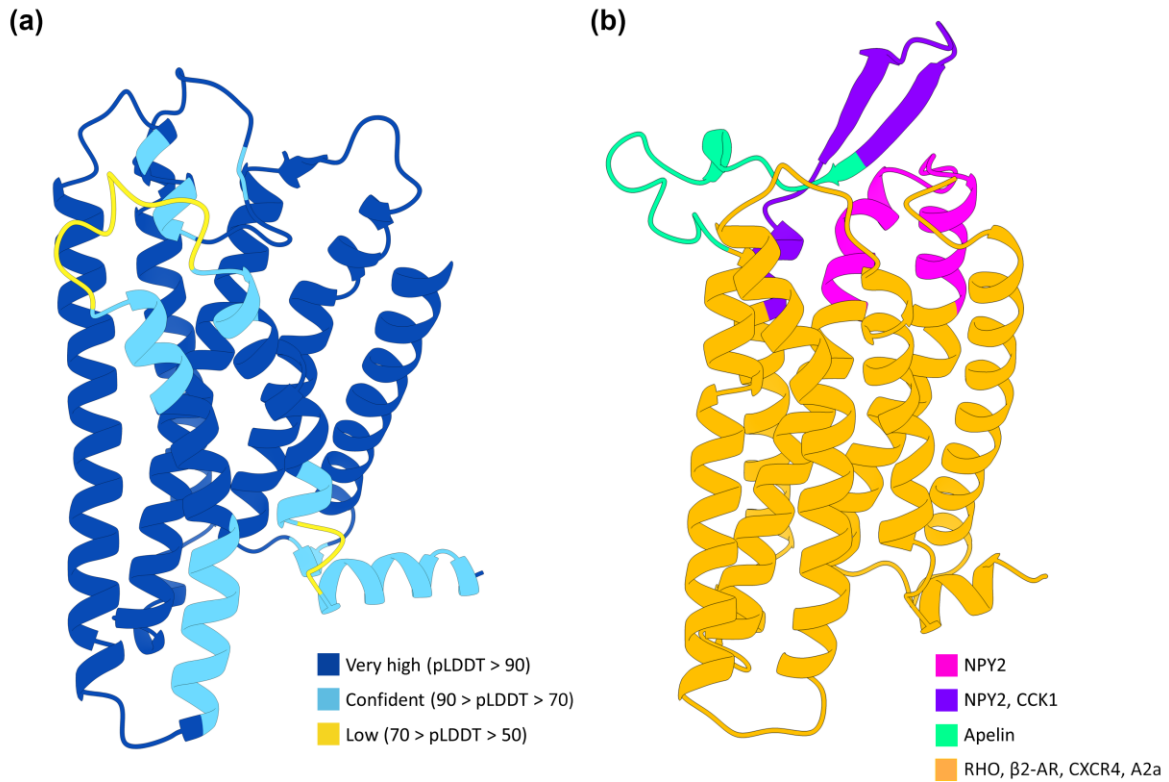


Figure S4. ROC analysis of the starting OR5K1 AF2 and HM models. ROC curves plot the rate of true positives (TPR) against false positives (FPR). ROC curves are colored in blue for the AF2 model and orange for HM.

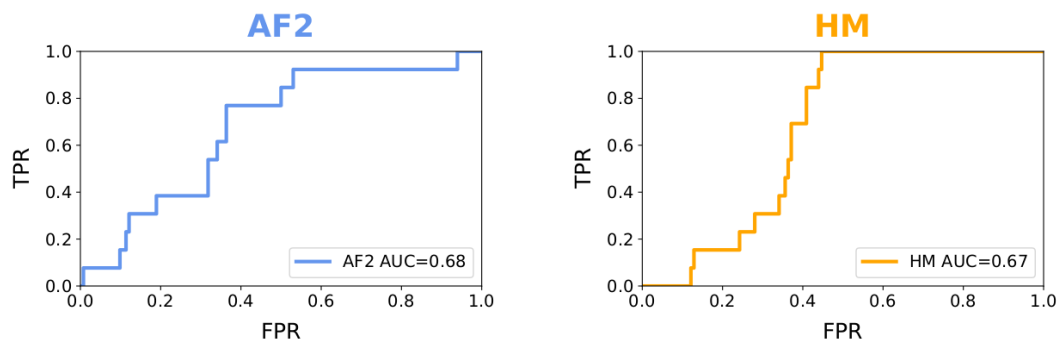


Figure S5. (a) ROC analysis of the OR5K1 AF2 and HM models after the first IFD simulation round. **(b)** Predicted binding modes of compound **1** within the OR5K1 binding site refined after the first IFD simulation round from AF2 and HM models.

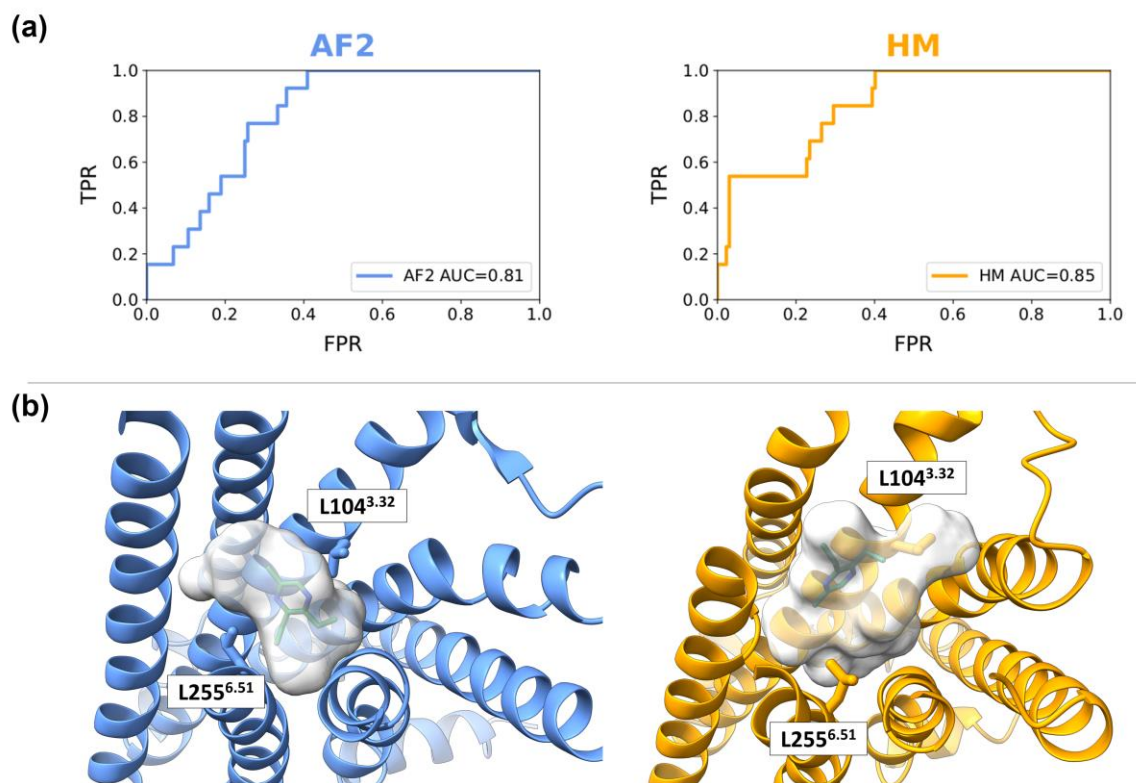


Figure S6. Distribution of the clusters binding poses of compound **1** in proximity to L104^{3.32} and L255^{6.51}.

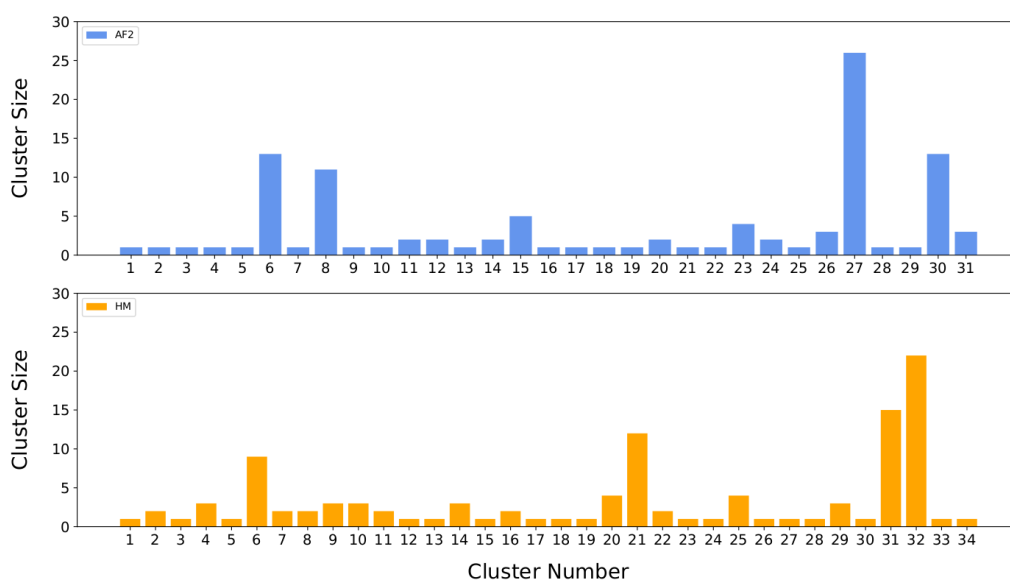
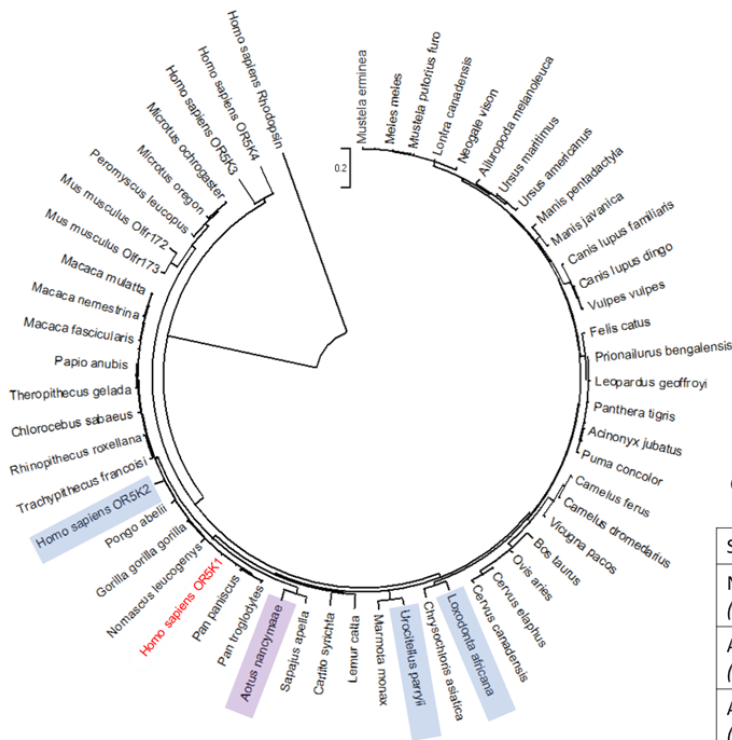


Figure S7. Leucine residues L104^{3.32} and L255^{6.51} are highly conserved in OR5K1 homologs. Circular, phylogenetic relationship of OR5K1 homologs with human rhodopsin as outgroup. Marked in purple is the species with an amino acid exchange at both conserved leucines. Highlighted in blue are the species with an amino acid exchange at L255^{6.51}. *Homo sapiens* OR5K1 is highlighted in red.



Conservative changes at pyrazine-binding leucins in OR5K1 homologs

Species	L104 ^{3.32}	L255 ^{6.51}
Nancy Ma's night monkey (<i>Aotus nancymae</i>)	Ile	Ile
African elephant (<i>Loxodonta africana</i>)		Ile
Arctic ground squirrel (<i>Urocitellus parryii</i>)		Ile
OR5K2 (<i>Homo sapiens</i>)		Ile

Figure S8. Heat maps representing the pairwise all atom RMSD matrices for the orthosteric binding site of the best performing models obtained after clustering IFD3 models. Matrices on the left refer to binding site residues, while matrices on the right include ligand coordinates in the selection of atoms for RMSD calculation. **(a)** Models from AF2. Cells are colored from dark orange to white according to increased RMSD values. Residues defining the binding site are: 104, 105, 108, 159, 199, 202, 206, 255, 256, 276, 279, 280. **(b)** Models from HM. Cells are colored from dark blue to white according to increased RMSD values. Residues defining the binding site are: 101, 104, 105, 108, 178, 180, 181, 199, 255, 258, 259, 275, 278, 279.

Models from IFD3 are available at https://github.com/dipizio/OR5K1_binding_site

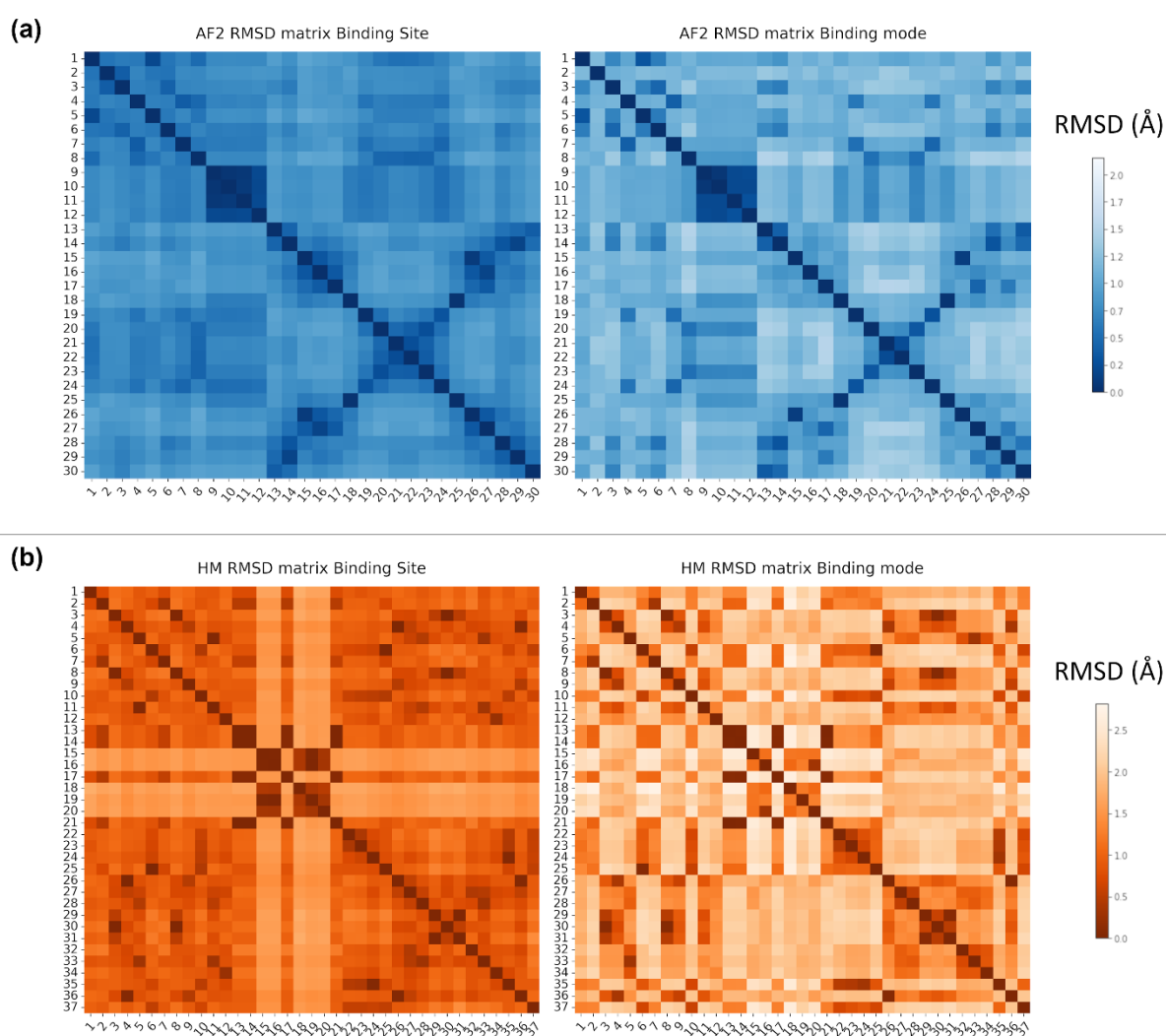


Table S1. Models from IFD1 and IFD2 with $d < 0.4$, AUC > 0.8 . Model names indicate if the model is resulting from the refinement of HM or AF2 models and if it is resulting from the first (IFD1) or second (IFD2) simulation round. We report AUC and EF_{15%} values, as well as information about the clustering and predicted van der Waals (vdW, kcal/mol) contribution of L104 and L255 to the binding of compound 1.

Model	AUC	EF _{15%}	d (nm)	Cluster	Cluster Size	vdW L104 ^{3,32}	vdW L255 ^{6,51}
HM_IFD2_36	0.901	0.615	0.278	4	3	-2.801	-1.379
HM_IFD2_1	0.872	0.400	0.273	5	1	-2.477	-0.697
HM_IFD2_5	0.810	0.313	0.197	6	9	-3.376	-0.955
HM_IFD2_56	0.834	0.313	0.281	8	2	-2.295	-0.768
HM_IFD2_21	0.869	0.500	0.38	11	2	-1.58	-0.657
HM_IFD2_20	0.921	0.909	0.31	20	4	-0.329	-2.36
HM_IFD2_18	0.902	0.615	0.283	21	12	-1.698	-1.562
HM_IFD2_23	0.864	0.400	0.391	23	1	-1.903	-0.385
HM_IFD1_14	0.914	0.615	0.345	31	15	-0.273	-1.439
HM_IFD1_3	0.864	0.615	0.33	32	22	-0.221	-1.734
HM_IFD1_11	0.837	0.211	0.26	33	1	-0.183	-2.274
HM_IFD1_13	0.855	0.615	0.399	34	1	-0.125	-1.503
AF2_IFD2_120	0.833	0.235	0.285	8	11	-2.477	-0.697
AF2_IFD2_105	0.883	0.538	0.356	15	5	-1.036	-1.246
AF2_IFD2_55	0.808	0.400	0.265	21	1	-2.690	-0.937
AF2_IFD2_79	0.855	0.400	0.153	27	26	-2.258	-2.669
AF2_IFD2_70	0.864	0.500	0.149	30	13	-2.815	-1.221
AF2_IFD2_98	0.841	0.313	0.216	31	3	-2.675	-1.284

Table S2. Oligonucleotides for molecular cloning of OR5K1.

Gene	Oligo-nucleotide	Restriction Site	TM (°C)		Sequence 5' → 3'
OR5K1	al-359	EcoRI	58	fw	CTGT <i>GAATTC</i> ATG GCT GAA GAA AAT CAT ACC ATG AAA AAT GAG TTT ATC
	al-360	NotI	59	rv	CTGC <i>GCGGCCGC</i> GTA ATT TCA CAT GGA AGT TTT TGC TGC ATT TAA ATC

TM = melting temperature, fw = forward, rv = reverse. *Italic letters highlight the restriction sites.*
Start and Stop codons are printed in bold.

Table S3. Vector internal oligonucleotides.

Vector	Oligonucleotide	TM (°C)		Sequence 5' → 3'
pFN210A	520	60	fw	GTG GAC ATC GGC CCG GGT C
	550	52	rv	CAC AAA TAA AGC ATT TTT TTC ACT GC

TM = melting temperature, fw = forward, rv = reverse

Table S4. Oligonucleotides for Homo sapiens OR5K1 site directed mutagenesis.

Gene	Oligonucleotide	TM (°C)		Sequence 5' → 3'
OR5K1 L104A	pm-275	57	fw	<i>GTACAGTTTTATTTTGCTTGCACTGTGG</i>
	pm-276	57	rv	<i>GTTTCCACAGTGCAAGCAAAATAAACTG</i>
OR5K1 L255A	pm-281	58	fw	<i>CATTATTCTATGGATCTGCTTTCTTCATGTAC</i>
	pm-282	58	rv	<i>GTACATGAAGAAAGCAGATCCATAGAATAATG</i>

TM = melting temperature, fw = forward, rv = reverse

Table S5. NCBI reference sequences of olfactory receptor genes investigated.

Gene			NCBI Reference Sequence
Description	Species	Common Species Name	(Accession-number)
OR5K1	<i>Acinonyx jubatus</i>	Cheetah	XP_014936391.1
OR5K1	<i>Ailuropoda melanoleuca</i>	Giant panda	XP_011231001.2
OR5K1	<i>Aotus nancymaae</i>	Nancy Ma's night monkey	XP_012332612.1
OR5K1	<i>Bos taurus</i>	Cattle	NP_001377368.1
OR5K1	<i>Camelus dromedarius</i>	Dromedary	XP_010976859.1
OR5K1	<i>Camelus ferus</i>	Wild Bactrian camel	XP_006181138.2
OR5K1	<i>Canis lupus dingo</i>	Australian dingo	XP_025272970.1
OR5K1	<i>Canis lupus familiaris</i>	Domestic dog	NP_001376061.1
OR5K1	<i>Carlito syrichta</i>	Philippine tarsier	XP_008064628.1
OR5K1	<i>Cervus canadensis</i>	Elk	XP_043304847.1
OR5K1	<i>Cervus elaphus</i>	Red deer	XP_043748867.1

OR5K1	<i>Chlorocebus sabaeus</i>	Green monkey	XP_007984292.2
OR5K1-like	<i>Chrysochloris asiatica</i>	Cape golden mole	XP_006868660.1
OR5K1	<i>Felis catus</i>	Domestic cat	XP_003991608.3
OR5K1	<i>Gorilla gorilla gorilla</i>	Western lowland gorilla	XP_004036007.1
OR5K1	<i>Homo sapiens</i>	Human	NP_001004736.2
OR5K2	<i>Homo sapiens</i>	Human	NP_001004737.1
OR5K3	<i>Homo sapiens</i>	Human	NP_001005516.1
OR5K4	<i>Homo sapiens</i>	Human	NP_001005517.1
OR5K1	<i>Lemur catta</i>	Ring-tailed lemur	XP_045396378.1
OR5K1	<i>Leopardus geoffroyi</i>	Geoffroy's cat	XP_045293055.1
OR5K1	<i>Lontra canadensis</i>	North American river otter	XP_032708517.1
OR5K1	<i>Loxodonta africana</i>	African bush elephant	XP_003418985.1
OR5K1	<i>Macaca fascicularis</i>	Crab-eating macaque	XP_045241289.1
OR5K1	<i>Macaca mulatta</i>	Rhesus macaque	NP_001180719.3
OR5K1	<i>Macaca nemestrina</i>	Southern pig-tailed macaque	XP_011732248.1
OR5K1	<i>Manis javanica</i>	Sunda pangolin	XP_036864239.1
OR5K1	<i>Manis pentadactyla</i>	Chinese pangolin	XP_036772454.1
OR5K1	<i>Marmota monax</i>	Groundhog	XP_046311290.1
OR5K1	<i>Meles meles</i>	European badger	XP_045859641.1
OR5K1	<i>Microtus ochrogaster</i>	Prairie vole	XP_005344986.1
OR5K1	<i>Microtus oregon</i>	Stoat	XP_041528659.1
Olfr172	<i>Mus musculus</i>	Mouse	NP_667212.2
Olfr173	<i>Mus musculus</i>	Mouse	NP_667211.2
OR5K1	<i>Mustela erminea</i>	Ermine	XP_032200493.1
OR5K1	<i>Mustela putoriusfuro</i>	Ferret	XP_004772048.1
OR5K1	<i>Neogale vison</i>	American mink	XP_044110727.1
OR5K1	<i>Nomascus leucogenys</i>	Northern white-cheeked gibbon	XP_003261757.2
OR5K1	<i>Ovis aries</i>	Domestic sheep	XP_004002914.1
OR5K1	<i>Pan paniscus</i>	Bonobo	XP_003821943.1
OR5K1	<i>Panthera tigris</i>	Tiger	XP_007096003.2
OR5K1	<i>Pan troglodytes</i>	Common chimpanzee	XP_526253.2
OR5K1	<i>Papio anubis</i>	Olive baboon	XP_031518815.1
OR5K1	<i>Peromyscus leucopus</i>	White-footed mouse	XP_028723391.1
OR5K1	<i>Pongo abelii</i>	Sumatran orangutan	XP_002813420.2
OR5K1	<i>Prionailurus bengalensis</i>	Leopard cat	XP_043451993.1
OR5K1	<i>Puma concolor</i>	Cougar	XP_025769507.1
OR5K1	<i>Rhinopithecus roxellana</i>	Golden snub-nosed monkey	XP_010352038.2
OR5K1	<i>Sapajus apella</i>	Tufted capuchin	XP_032154801.1
OR5K1	<i>Theropithecus gelada</i>	Gelada	XP_025233483.1
OR5K1	<i>Trachypithecus francoisi</i>	François' langur	XP_033067425.1
OR5K1	<i>Urocitellus parryii</i>	Arctic ground squirrel	XP_026258216.1
OR5K1	<i>Ursus americanus</i>	American black bear	XP_045641565.1
OR5K1	<i>Ursus maritimus</i>	Polar bear	XP_008700260.1
OR5K1	<i>Vicugna pacos</i>	Alpaca	XP_006208125.1
OR5K1	<i>Vulpes vulpes</i>	Redfox	XP_025846410.1