

Molecular basis of olfactory receptors: in silico modeling and in vitro validation

Adrien Acquistapace, Véronique Martin, Jean-Francois Gibrat, Guenhaël Sanz, Gwenaëlle André-Leroux

► To cite this version:

Adrien Acquistapace, Véronique Martin, Jean-Francois Gibrat, Guenhaël Sanz, Gwenaëlle André-Leroux. Molecular basis of olfactory receptors: in silico modeling and in vitro validation. 19. Congrès GGMM (Groupe de graphisme et modélisation moléculaire), Groupe de Graphisme et Modélisation Moléculaire., May 2015, Sète, France. hal-02164158

HAL Id: hal-02164158 https://hal.science/hal-02164158

Submitted on 5 Jun2020

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.

Molecular basis of Olfactory Receptors activation In silico/in vitro studies of Prostate Specific G-coupled Receptor

Adrien Aquistapace[#], Véronique Martin^{*}, Jean-François Gibrat^{*}, Guenhaël Sanz[#] & Gwenaëlle André-Leroux^{*}

Gwenaelle.andre-Leroux@jouy.inra.fr

Inra, Unité de NeuroBiologie de l'Olfaction UR1197, Domaine de Vilvert, 78352 Jouy-en-Josas, France.* Inra, Unité de Mathématique et Informatique Appliquées du Génome à l'Environnement UR1404, Domaine de Vilvert, 78352 Jouy-en-Josas, France.

Ifactory receptors (ORs) belong to the highly versatile family of G Protein-Coupled Receptor family (GPCRs)1-2 and account for as much as 3% of the human genome. They have been first identified in the neurons of the olfactory epithelium where they detect and signal a large panel of odorants molecules.

In addition, ORs detect steroid hormones in non-olfactory tissues, regulate serotonin secretion in the gut or participate in cell migration & muscle adhesion 16. Noteworthy, some OR are overexpressed in tumors, where they promote in vitro cell invasiveness 17-11. Notably, LNCaP prostate cancer cells endogenously express an OR named PSGR, the Prostate Specific G protein-coupled Receptor¹². Sanz et al showed that stimulation of ORs expressed by LNCaP cancer cells increased the metastasis occurrence in vivo, in agreement with the increased cell invasiveness observed in vitro. All these data suggest that ORs can be stimulated by molecules in the microenvironment and could be involved in tumors. Thus, ORs could be novel therapeutic targets. As such, it is critical to identify their ligands and to unravel their specificity and activation mechanisms. Deciphering how OR functional behaviour reconciles with structural properties is a challenge, because there is no experimental structure available

for ORs and because most of them are orphans, *i.e* their physiological agonists are unknown. Despite the lack of available structures for PSGR and generally for OR, the increasing number of solved GPCR structures prompted us to investigate the molecular mechanism of PSGR activation upon agonist recognition, using molecular mechanics. We investigate in silico the molecular basis of folding and ligand binding in ORs. Activity tested on PSGR wild-type and mutants with α and β -ionone validated our modeling approach, including homology modeling, mutagenesis and docking.

	TM1 N 1.50	TM2 D 2.50	TM3 R 3.50	TM4 W 4.50	TM5 P 5.50	TM6 P 6.50	TM7 P 7.50
40	50 🤜 70	80 🔶 90	100 110 120 130 140 🕂 150	160 170 🦊 180 🖣 190	200 210 220 230 240 - 256 260 270	280 290 - 310 320	330 340 🕂 350 360
PSGR G - L E KAH F WV G F P EQLEEQK PLF - G S SSW - L T L F V PS V PLG - L KV T I VG L SPA - I PVI I T AV SLA - L AI AI T AL SMV - T AI T I MAL ET L - NKYVVI I A ANF - NKI F L PT I E PW - QF SMLAAY KEN - SIKL T SV V - IM - G S SVYI T V - QW - EA - GMS L L - W - VV - GMG I V - W - VV F I AF L	LL SMY VVAMF GNCIVVFIVRTERSLHAPM FLFMYLVTVAGNLLIILVIITDTOLHTPM YT GVFVVSLPLNIMAIVVFILKMKVKKPA YLAVCVGGLLGNCLVMYVILRHTKMKTAT YSVVFVVGLVGNSLVMFVIIRYTKMKTAT YSVVFVVGLGNSLVMFVIIRYTKMKTAT YSIVCVVGLFGNFLVMYVIVRYTKMKTAT YALVFLLSLLGNSLVMLVILYSRVGRSVT YSIIFLTGIVGNGLVILVMGYQKKLRSMT MFLLIMLGFPINFLTLYVTVQHKKLRTPL FILICCFIILENIFVLLTIWKTKKFHRPM ELAIAVLAILGNVLVCWAVWLNSNLQNVT YCALILAIVFGNGLVCWAVWLNSNLQNVT MALVVLLIVAGNVLVIAAIGSTORLOTLT MALVVLLIVAGNVLVIAAIGSTORLOTLT MALVVLLIVAGNVLVIAAIGSTORLOTLT MALVVLLIVAGNVLVIAAIGSTORLOTLT	AYLFLCMLAAIDLALSTSTMPKIL AYFFLANLSLADACFVSTTVPKML AVVYMLHLATADVLFV-SVLP INIYIFNLALADTLVL-LTLP INIYIFNLALADALVT-TTMP INIYIFNLALADALAT-STLP INYIFNLALADALAT-STLP INYILNLALADALAT-STLP INYILLNLAVADLLFA-LTLP INYILLNLAVADLFV-ITLP INYFIGNLALSDLLAGVAYTANLL INYFVVSLAAADIAVGVLAIP INYFITSLACADLVAGLVVP INYFITSLACADLVVGLVVP INYFLFSLACADLVGAVVMP INYFLFSLACADLIGVFSMN INYFLLSLACADLIIGVISMN	ALFWFDSREISFEACLTOMFFIHALSAIESTILLAMAFDRYVAICHPLF ANIQIOSQAISYSGCLLQLYFFMLFVMLEAFLLAVMAYDCYVAICHPLF FKISYYFSGSDWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMC FQGTDILL-GFWPFGNALCKTVIAIDYYNMFTSTFTLTAMSVDRYIAVCHPVK FQSTVYLM-NSWPFGDVLCKIVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPVK FQSAKYLM-ETWPFGELLCKAVLSIDYYNMFTSIFTLCTMSVDRYIAVCHPVK FQSAKYLM-GTWPFGNILCKIVISIDYYNMFTSIFTLCTMSVDRYIAVCHPVK FQSAKYLM-GTWPFGNILCKAVLSIDYYNMFTSIFTLCTMSVDRYLAVCHPVK FQSAKYLM-GTWPFGNILCKAVLSIDYYNMFTSIFTLCTMSVDRYLAVCHPVK FWAVDA-V-ANWYFGNFLCKAVHVIYTVNLYSSVWILAFISLDRYLAIVHATF FWAVDA-V-ANWYFGNFLCKAVHVIYTVNLYSSVWILAFISLDRYLAIVHATM TTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVCKPMS LSGATTYKLTPAQWFLREGSMFVALSASVFSLLAIAIDRYLAIRIPLF WVVYLEVTGGVWNFSRICCDVFVTLDVMMCTASIWNLCAISIDRYTAVVMPVF FGAH-ILMKMWTFGNFWCEFWTSIDVLCVTASIETLCVIAVDRYFAITSPFF FGATL-VVRGTWLWGSFLCELWTSLDVLCVTASIETLCVIAIDRYLAITSPFF MNILY-LLMSKWSLGRPLCLFWLSMDYVASTASIFSVFILCIDRYRSVQQPLF LYTLY-TVIGYWPLGPVVCDLWLALDYVVSNASVMNLLIISFDRYFCVTKPLT	RHAAVLN NTVTAQIGIVAVVRGSLFFFPLPLLIKRLAFC YILIMS PGLCIFLVSASWIMNALHSLLHTLLMNSLSFC SLSWRT LGRASFTCLAIWALAIAGVVPLLLKEQT - IQV ALDVRT SSKAQAVNVAIWALASVVGVPVAIMGSA - QVE (ALDFRT PLKAKIINICIWLLSSVGVSAIVLGGT - KVR (ALDFRT PRNAKINICIWLLSSSVGISAIVLGGT - KVR (ALDFRT PRNAKINICIWLLSSSVGISAIVLGGT - KVR (ALDFRT PRNAKIVNVCNWILSSAIGLPVMFMATT - KYR (ALDFRT RAKAKINICIWVLASGVGVPVIMATT - KYR (ALDFRT CTRAKGIIAICWVLSFAIGLTPM - LGWNNCGQ YNGLVT CTRAKGIIAICWVLSFAIGLTPM - LGWNNCGQ YQSLMT KTRASATILGAWFLSFL - WVIPI - LGWN - HFM YQSLMT TKRAGMMIAAAWVLSFILWAPAI - LFWQ - FIV YRAKRT TKRAGVMIGLAWVISFVLWAPAI - LFWQ - YFV	HSNVL SHSYCVHQDVMKLAYADTLPNVV- YGLTAILLVMGVDVMFISLSYFLIIRTVLQLPS- ANHEIPHFFCDINPLLSLSCTDPFTNEL- VIFITGGLTGLICVLCLIISYTNVFSTILKIPS- G LGITTCHDVLSETLLEGYYAYYFS-AFSAVFFFVPLIISTVCYVSIIRCLSSS- C VDVIECSLQFPDDDVS-WWDLFMKICVFIFAFVPLIISTVCYVSIIRCLSSS- C VDVIECSLQFPDDDYS-WWDLFMKICVFIFAFVPLIISTVCYVSIIRCLSSS- C VDVIECSLQFPDDDYS-WWDLFMKICVFIFAFVPLIITVCYGLMLRLKSVRLLS- C GAVVCMLQFPSPSW-YWDTVTKICVFIFAFVPLIITVCYGLMLRLKSV- O GAVVCMLQFPSPSW-YWDTVTKICVFIFAFVPLIITVCYGLMLRLKSV- O GSIDCTLTFSHPTW-YWENLLKICVFIFAFVPLIITVCYGLMLRLKSV- O GSIDCTLTFSHPTW-YWENLKICVFIFAFVPLIITVCYGLMLRLKSV- O SPVCYEVLGNDTA-KWRMVLRILPHTFGFIVPLFVMLFCYGLTISKLSHS- C SPVCYEVLGNDTA-KWRMVLRILPHTFGFIVPLFVMLFCYGLVFTKEAAAQQQ ALSS CSTVLPL- ALSS SVLVPLVPLV ALSS SVLVPLVPL QAL CSQCGEGQVACLFEDV- QAL SVLVPLCDFTN QAL SVLVPLVPLV QAL SVLVPLV QAL SVLVPLV QCCGEGQVACLFEDV SVLVPLVYFSVVFVLVVNVVVYARIYVLKQRR QCCGEGQVACLFEDV VPMNYMYFNFFACVLVPLLMLGVVLVYARIYVLKQRR QAL SVLVPLVV		 PIVRVVMGDIYLLLPPVINPIIYGAKTKQIRTRVLAMFKISI AQKDTVASVMYTVVTPMLNPFIYSLRNQEIKSSLRKLIWVR TEAAYFAYLLCVCVSSISCCIDPLIYYASSEC TAVAILRFCTALGYVNSCLNPILYAFLDENFKACFRDFCFP DPLVVAALHLCIALGYTNSSLNPVLYAFLDENFKRCFRDFCFP DPLVVAALHLCIALGYTNSSLNPVLYAFLDENFKRCFREFCIPT RNNIGRALDATEILGFLHSCLNPILYAFLGQNFRHGFLKILAMH ENTVHKWISITEALAFFHCCLNPILYAFLGAKFKTSAQHALTSG G-P-IFMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVTTLCCG D-ILFRAEYFLVLAVLNSGTNPIIYTTFNIEFRKAFIRINGRP A-PLWLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSH SPELYSATTWLGYVNSGFNPLIYCRS PDWLFVAFNWLGYNSGFNPLIYCRS NEHLHMFTIWLGYINSTLNPLIYPLCNENFKKTFKRILHI PNTVWTIGYWLCYINSTNPACYALCNATFKKTFKHLLM
- 00 00005043	2 4 4 8 4 7 4 5 7 4 3 * 4 8 5 8 4 5 5 2 2 2 3 2 6 6 4 4 5	5 3 4 7 4 4 3 * 85 5 * 5 6 6 5 - 4 4 5 6	12 42 1 - 00 - 002 05 400 12 000 14 4 12 5 4 4 35 4 5 35 8 4 4 9 7 6 9 5 9 3 6 8 4 2 6 3 2	2 0 0 0 0 2 0 0 2 0 3 2 4 5 4 5 4 5 6 6 3 7 4 5 0 4 6 4 0 3 - 4 1 2 2 - 0 0 0	0000002101231005314100426483444544464264225122	0000022743554447557476958427414411200211	2 0 1 1 2 1 5 3 2 5 6 5 2 4 2 6 5 5 8 * 6 8 * 3 2 2 0 0 1 1 3 0 0 1 0 0
++WE - +KVF I +AV	YSLVCLVGLLGN+LV++VI+R++KLKTVT	INY+I+NLALADLLVGVLT+PK-L	A-FQ+TYYL+MGTWPFGNFLCKLVLSIDYVNM+ASIFTLLVISIDRYIAICHPLF	RYLDFRT-S-P+RAKVIILA+WVLS+L+G+PPILLGW+RYF+	+++SHS-C+C+GGI+CCDFFPNP++YWDNVFKICTFI+AFYVP+LIITVCYGRIIR+LKSV++LS	GSREKD++ERKATRTVLIVVG+FILCWLPYFIVVLVNTFCDTICS+-	+++PL+LYVLCIWLGY+NSCLNPIIYAFLNENF+RCFRKILC++

PSGR sequence aligned with solved structures of GPCRs

Sequence analysis of ORs	Docking of ligands		
lfactory proteome non redundant & checked => 437 ORs.	3.32 3.33 3.36 TM3 Central hub		

Clustering with nrdb 90 (alignment at 90% d'id min) => 377 OKS. Alignment with Clustalw and analysis using HMM profiling of TM segments

Conservation of Ballesteros-Weinstein^[14]:

 $\bigcirc 0$

INIT 1.50	N de GN	N $de \ll GN \gg (-1/3/7)$ TM6 P kink ext	
TM2 2.50	D	D/E (- 8/377)	
TM3 3.50	R de D R YX ₅ P	R de DRY/FX ₅ P (- 46/377) $^{\text{TM5 P kink}}$	
TM4 4.50	W	W/Y/hydrophobe	
TM5 5.50	ΡΧ ₇ Υ	P/D/C/S/TX ₆ SY	
TM6 6.50	P de K/RX ₁₇ P	P/S/T de $\mathbf{K}X_{17}\mathbf{P}$ (X_{17} conserved)	*
TM7 7.50	P de NP	P de NP	4

TM4-TM6 less conserved in GPCR but highly conserved among ORs Specificity of ORs.

Homology modeling

• For each solved structure of GPCR that defines a template, 100 homology models of PSGR were computed using Modeller ^[14], in respect with the multiple alignement, described above.

Active and inactive conformations were modelled.

• For each series of PSGR structure, 5 models were selected, on the GABE and DOPE scores, and analyzed with Molprobity. They were then minimized with Charmm forcefield^[15] 5,000 it, 0,01 RMS and Smart Algorithm from DiscoveryStudio[©].



Position 6.48 is part of the transmission switch [16]

« Upon agonist binding, a rearrangement of a cluster of conserved hydrophobic and aromatic residues occurs deeper in the receptor core. »



 $\circ \alpha$ -ionone, β -ionone, testosterone and other ligands were alternatively docked inside the PSGR binding cradle that defines the receptor site. Multiple positions were tested and ranked with Charmm FF, during when the PSGR was kept rigid. The complexes, ranked as best, were minimized and their total potential energy plus interaction energy were computed. Data with $\alpha \& \beta$ ionones are below.



By analogy with the solved GPCR^[16], the binding cradle was proposed to be composed of I103, H104 & S107 of TM3, Y251 & L254 of TM6 and D277 of TM7 and subjected to alanine site-directed mutagenesis.

W e designed homology models of active & inactive conformations of PSGR, in respect with the available template structures of GPCR.

Six residues that putatively define the agonist binding site were mutated into alanine residues. Notably, four of them result in PSGR functional impairment, thus validating our computational analysis, including sequences alignement, homology modeling and docking. We now step further to identify the molecular determinants that discriminate agonist from antagonist binding in the PSGR. Ligand specificity in ORs will then be more largely addressed. Eventually, this work is dedicated to develop a more accurate version of our GPCR automodel webservice^[17].

Bibliography

[1] Buck L et al. (1991) Cell 65: 175-187 [2] Niimura Y et al. (2007) PLoS One 2: e708 pourquoiAcy-tually [3] Neuhaus EM *et al.* (2009) J Biol Chem 284: 16218-16225. [4] Braun T et al. (2007) Gastroenterology 132: 1890-1901

[5] Kidd M et al. (2008) Am J Physiol Gastrointest Liver Physiol 295: G260-272 [6] Griffin CA et al. (2009) Dev Cell 17: 649-661 [7] Weng J et al. (2005) Int J Cancer 113: 811-818 [8] Leja J et al. (2009) Mod Pathol 22: 261-272 [9] Muranen TA et al. (2011) Breast Cancer Res 13: R90

[10] Cui T et al. (2013) Eur J Endocrinol 168: 253-261 [11] Giandomenico Vet al. (2013) J Mol Endocrinol PMID: 23969981 [12] Sanz G et al. (2014) PLoS One 9: e85110 [13] Ballesteros & Weinstein (1995) *Methods Neurosci* [14] Sali et Blundell (1999) J Mol Biol 234: 779-815

[15] Brooks B.R et al. (2009) J Comput Chem 30: 1545-1614 [16] Venkatakrishnan A.J et al. (2013) Nature 494: 185-194 [17] Launay G et al. (2012) PEDS 25: 377-386

